Nasir Faruk
Y. A. Sha'aban
I. J. Umoh
E. A. Adedokun
Bamidele Oluwade (Eds.)

The Role of Computing in the Evolution and Development of Emergent and Alternative Technologies

2nd International Conference of the IEEE Nigeria Computer Chapter: IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019
Proceedings

Computer Engineering
ABU Zaria

IEEE
Advancing Technology for Humanity

Proceedings of:

2nd International Conference of the IEEE Nigeria Computer Chapter

Theme: The Role of Computing in the Evolution and Development of Emergent and Alternative Technologies

IEEEnigComputConf'19

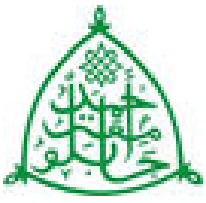
In Collaboration with the:
Department of Computer Engineering, Faculty of Engineering

Ahmadu Bello University, Zaria, Nigeria

October 14-17, 2019

IEEE Catalog Number: CFP19G60-PRT

Online ISBN: 978-1-7281-0713-4
Print ISBN: 978-1-7281-0712-7



Proceedings of the 2nd International Conference of the IEEE Nigeria Computer Chapter: IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

Editors

- Prof. Bamidele Oluwade, Chair, IEEE Nigeria Computer Chapter
- Dr. Nasir Faruk, University of Ilorin
- Dr. Y. A. Sha'aban, Ahmadu Bello University, Zaria
- Dr. I. J. Umoh, Ahmadu Bello University, Zaria
- Dr. E. A. Adedokun, Ahmadu Bello University, Zaria

2019 2nd International Conference of the IEEE Nigeria Computer Chapter: IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

ISBN 978-1-7281-0713-4

Preface

It is our great pleasure to introduce the proceedings of the Second Edition of the International Conference on of the IEEE Nigeria Computer Chapter, held during October 14–17, 2019, at the Ahmadu Bello University, ABU, Zaria, Kaduna State, Nigeria. This conference drew together researchers and developers from both academia and industry - especially in the domains of computing, networking and communications engineering.

The theme of IEEEnigComputConf'19 was "The Role of Computing in the Evolution and Development of Emergent and Alternative Technologies" and was organised by the Department of Computer Engineering, Faculty of Engineering, Ahmadu Bello University, Zaria and technically co-sponsored by the Computer Science Chapter of the Institute of Electrical and Electronics Engineers (IEEE), the Nigeria Section.

The technical program of IEEEnigComputConf'19 consisted of 90 full papers in oral presentation sessions in the main conference tracks and 2 keynotes presentations.
The major conference tracks were:

- Track 1 - Control and Optimizations
- Track 2 - Algorithms and Machine Learning
- Track 3 - IoT, Image Processing, Cryptography and Networking
- Track 4 - RF, Power, Proof-of-Concept

Apart from the high-quality technical paper presentations, the technical program also featured two keynote speeches and one invited talk. The two keynote speakers were Professor Rasheed Gbenga JIMOH, Professor of Computer Science and Dean of Faculty of Communication and Information Sciences, University of Ilorin, Nigeria and Professor Marco A. Wiering, Department of Artificial Intelligence, University of Groningen, Groningen, The Netherlands. It was also a great pleasure to work with such an excellent organising committee, who put in very hard work in organising and supporting the conference. In addition, the work of the Technical Programme Committee is also greatly appreciated: they completed the peer-review process of technical papers and compiled a high-quality technical programme.

We strongly believe that the IEEEnigComputConf'19 provided a good forum for all researchers, developers and practitioners to discuss recent advancements in computing, networking and communications engineering. We also expect that the future IEEEnigComput conferences will be as successful and stimulating, as indicated by the contributions presented in this proceeding.

October 2019

Nasir Faruk
Y. A. Sha'aban
I. J. Umoh
E. A. Adedokun
Bamidele Oluwade

Committees

Program Committee Chairs

- Professor Bamidele Oluwade, Chair, IEEE Nigeria Computer Chapter
- Dr. Nasir Faruk, University of Ilorin
- Dr. Y. A. Sha'aban, Ahmadu Bello University, Zaria
- Dr. I. J. Umoh, Ahmadu Bello University, Zaria
- Dr. E. A. Adedokun, Ahmadu Bello University, Zaria
- Engr. Tunde Salihu, Shaybis Nigeria Limited

Organizing committee

- Professor Bamidele Oluwade, Chair, IEEE Nigeria Computer Chapter
- Engr. Tunde Salihu, Immediate Past Chair, IEEE Nigeria Computer Chapter
- Professor M. B. Muazu, Ahmadu Bello University, Zaria
- Engr. Raphael Onoshakpor, Chair, IEEE Nigeria Section
- Dr. Y. A. Sha'aban, Ahmadu Bello University, Zaria
- Professor Olumide Longe, Caleb University, Imota
- Professor A. A. Ayeni, University of Ilorin
- Dr. Kingsley Ukaoha, University of Benin
- Dr. O. A. Adeyemo, University of Ibadan
- Dr. A. O. Babatunde, University of Ilorin
- Dr. Abel Ajibesin, American University of Nigeria, Yola
- Engr. Abdulateef Aliyu, Membership Development Officer (MDO), IEEE Nigeria Section
- Dr. Vincent Akpan, Federal University of Technology, Akure
- Dr. E. A. Adedokun, Ahmadu Bello University, Zaria
- Prince Ibe, Prince Ibe LLC, USA
- Dr. (Mrs.) Tinuke Oladele, University of Ilorin

- Dr. (Mrs.) Victoria Yemi-Peters, Kogi State University, Anyigba
- Dr. Emmanuel Ogala, Kogi State University, Anyigba
- Ugochukwu Onwudebelu, Federal University, Ndufu Alike-Ikwo, Ebonyi State
- Engr. Soji Lawal, IEEE Nigeria Computer Chapter
- David Okereafor, Webmaster, IEEE Nigeria Section
- Dr. Kennedy C. Okafor, Federal University of Technology, Owerri
- Dr. J. F. Opadiji, University of Ilorin
- Dr. A. T. Ajiboye, University of Ilorin
- Dr. (Mrs.) T. O. Oladele, University of Ilorin
- Dr. (Mrs.) O. C. Abikoye, University of Ilorin
- Dr. Nasir Faruk, University of Ilorin
- Dr. (Mrs.) Rukayat A. Koleoso, University of Lagos
- Morufu Amusa, Asia e-University
- Professor J. O. Fatokun, Anchor University
- Dr. Julius Odili, Anchor University
- Professor J. O. Afolayan, Anchor University
- Dr. R. G. Jimoh, University of Ilorin
- Dr. (Mrs.) M. Mabayoje, University of Ilorin
- Dr. A.O. Bajeh, University of Ilorin
- Dr. D. R. Aremu, University of Ilorin
- Dr. (Mrs.) Ifeyinwa E. Achumba, Federal University of Technology, Owerri
- Mrs. Oluwakemi S. Ayodele, Kogi State Polytechnic, Lokoja
- Joshua Agbogun, Kogi State University, Anyigba

Local Organizing Committee

- Prof Mu'azu Muhammed Bashir
- Dr Abdulrazaq Muhammad Bashir
- Dr Sha'aban Yusuf Abubakar
- Dr Adedokun Emmanuel Adewale
- Mr Ajayi Ore-Ofe
- Dr Ibrahim Yusuf
- Mr Umar Abubakar
- Mrs Adebiyi Risikat Folashade
- Mr. Yusuf Shehu Mohammed
- Dr. Salawudeen Ahmed Tijani
- Dr Bello Salau Habeeb
- Dr. Okafor Emmanuel
- Mr. Adekale Abdulfatai Dare
- Dr. Yahaya Basira
- Dr. Sadiq Bashir Olaniyi
- Dr. Umoh Ime Jarlath
- Mr Yusuf Ibrahim Salihu
- Mr Haruna Zaharuddeen
- Mrs. Usman Shehu Nafisa

Conference Organization

Chairman, Central Planning Committee (CPC)/Technical Program Committee (TPC) & Chair, IEEE Nigeria Computer Chapter

Professor Bamidele Oluwade, Ph.D., SMIEEE, CITP, FNCS

Co-Chairman, CPC & Immediate Past Chair, IEEE Nigeria Computer Chapter

Engr. Tunde Salihu, SMIEEE, MNSE, Shaybis Nigeria Limited

Head, Department of Computer Engineering, Ahmadu Bello University, Zaria & Member, CPC

Professor M. B. Muazu, Ph.D. Department of Computer Engineering, ABU, Zaria

Chair, LOC

Dr. Y. A. Sha'aban, Ph.D
Department of Computer Engineering
Ahmadu Bello University, Zaria

Chair, Conference Publication

Dr. Nasir Faruk, Ph.D
Department of Telecommunication Science
University of Ilorin

Chair Conference Venue, Accommodation & Logistics Committee

Dr. E. A. Adedokun, Ph.D
Department of Computer Engineering
Ahmadu Bello University, Zaria

Co-Chair, (LOC)

Dr. I. J. Umoh, Ph.D.
Department of Computer Engineering
Ahmadu Bello University, Zaria

Secretary, Central Planning Committee

Dr. Kingsley Ukaoha, Ph.D.
Department of Computer Science
University of Benin

Reviewers

1. Mohammed Abdullahi
2. Muhammad Bashir Abdulrazaq
3. Fasola Abifarin
4. Quadri Adebowale
5. Emmanuel Adewale Adedokun
6. Barroon Isma'Eel Ahmad
7. Salawudeen Ahmed Tijani
8. Ore-Ofe Ajayi
9. Taiwo Ajayi
10. Razak Alli-Oke
11. Habeeb Bello Salau
12. Haruna Chiroma
13. Nasir Faruk
14. Zaharuddeen Haruna
15. Yusuf Ibrahim
16. Donfack A. F. Kana
17. Abdullahi Bala Kunya
18. Ameer Mohammed
19. Muhammed Bashir Muazu
20. Okechukwu Ochia
21. Seyi Oderanti
22. Emmanuel Okafor
23. Adejoke Olamiti
24. Bamidele luwade
25. Oluwaseyitanfunmi Osunade
26. Bashir Olaniyi Sadiq
27. Abdulazeez Femi Salami
28. Hamza Salami
29. Yusuf Sha'Aban
30. Kingsley Chiwuike Ukaoha
31. Abubakar Umar
32. Ime Jarlath Umoh
33. Kaisan Usman Muhammad
34. Nancy Woods
35. Ibrahim Yusuf
36. Abubakar Abdulkarim
37. Abudulrahman Tosho
38. Adebowale Ramon
39. Adewole Kayode

Table of Contents

Big Data Analytics: Emerging Strategy for the Ubiquitous Computing Era **17**
Rasheed Gbenga JIMOH

Reinforcement Learning for Playing Games **18**
Marco A. Wiering

Design and Implementation of Electricity On-line Billing Payment System **19**
Amanda Andrew, Patrick Ubeh Okorie and Ibrahim A. Abdu

Genetic Algorithm Based Dual Input Power System Stabilizer **25**
Aliyu I. Usman and Nuradeen Magaji

Machine Learning Classification Algorithms for Phishing Detection: A Comparative Appraisal and Analysis **32**
Noah Ndakotsu Gana and Shafi'I Muhammad Abdulhamid

Development of a True Random Number Generator based on Computer Mouse Wait time and Datetime **43**
M.A Idakwo, M.B Muazu, E.A Adedokun and Y.A Sha'aban

Implementation of Timing and Synchronization in Digital Clocks: A Wireless Communication Design **50**
Basira Yahaya, Suleiman Garba, Bashir Olaniyi Sadiq, Daniel Udekwe and Zainab Abubakar Mukhtar

Development and Testing of Q-Basic Computer Software for Newton-Raphson Power Flow Studies **55**
Lambe Mutalub Adesina, James Katende and Ganiyu Adetayo Ajenikoko

Smart Energy System: The Panacea to Nigeria Ecliptic Power Supply **62**
Sylvester Ajah

Polynomial Reduction of TSP to Freely Open-loop TSP **70**
Muhammad Bashir Abdulrazaq, Yusuf Suleiman Tahir, Suleiman Sha'Aban and Muhammed Sani Jibia

Design, Construction and Testing of an Electronic Hospital Queuing Solution System **74**
Ajekiigbe Enoch Seun, Usman Bature Isyaku, Kabiru Ibrahim Jahun, Nasir Abukar Yakubu, Abbas Musa Hassan and Usman Saleh Toro

IoT-Based Home Automation, Energy Management and Smart Security System **80**
Shadrach Adamu, Usman Bature Isyaku, Nasir Abubakar Yakubu, Abbas Musa Hassan, Kabiru Ibrahim Jahun and Usman Saleh Toro

A Data Visualization Approach for Enhancing Room Management in University Buildings **86**
Fortune Chuku, Charles Nche and Sandip Rakshit

Temporal Variation of Duty Cycle in the GSM Bands **95**
Ganiyu A. Ishola, Balogun V. Segun, Nasir Faruk, Kayode S. Adewole, Abdulkarim A. Oloyede, Lukman A. Olawoyin and Rasheed.G. Jimoh

Evaluating Ensemble Techniques for Predicting Students' Grades Based on their Interaction with Learning Management System **100**

Umar Isa Usman, Salisu Aliyu, Baroon Isma'il Ahmad and Yusuf Abubakar

Conceptual Framework for Detection of Learning Style from Facial Expressions using Convolutional Neural Network **109**
Farouk Lawan Gambo, Gregory Maksha Wajiga and Etemi Joshua Garba

A New Concept of an Intelligent Protection System Based on a Discrete Wavelet Transform and Neural Network Method for Smart Grids **114**
Ali Hadi Abdulwahid

Automatic Generation of Design Class Model from Use Case Specifications **120**
Muhammad Shaheed Abdullahi, Talal Shaikh, John Kanyaru and Muhammad Aminu Umar

Control of a Double Pendulum Crane System Using PSO-Tuned LQR **125**
Mustapha Muhammad, Ibrahim Abdulhamid and Amina Khaleel

Problems and Prospects of Internet of Things to the Automobile Industry in Nigeria **134**
Aliyu Mustapha, Abubakar Mohammed Idris, Mohammed Abdulkadir, Isah Abubakar Ndakara, Abdullahi Abubakar Kutiriko, Haruna Ahmed Dokoro and Abdullahi Raji Egigogo

Intrusion Detection System Based on Support Vector Machine Optimised with Cat Swarm Optimization Algorithm **138**
Suleiman Idris, Ishaq Oyefolahan O and Juliana Ndunagu N.

Evaluating GrabCut Algorithm for Cervical Cancer Lesion Segmentation **146**
Suleiman Mustafa and Mohammed Dauda

Fetal Heart Rate Estimation: Adaptive Filtering Approach vs Time-Frequency Analysis **150**
Ashraf A. Ahmad, Dominic S. Nyitamen, Sagir Lawan and Chiroma L. Wamdeo

A Resonant Fault Current Limiting Prediction Technique based on Auditory Machine Intelligence **157**
Biobele A. Wokoma and Emmanuel N. Osegi

Development of an Improved Intelligent Hybrid Expert System for Diagnosis of Lassa Fever **162**
Samuel Ekene Nnebe, Okoh Nora Augusta Ozemoya, Adetokunbo Macgregor John-Otumu and Emmanuel Osaze Oshoiribhor

Hybrid Approach for Selection of Academic Major Using Naïve Bayes **173**
Farouk Gambo, Adamu Yahaya, Khalid Haruna, Abdullah Abubakar and Aminu Abdullahi

Adoption of E-commerce in Nigeria Challenges and Future Prospects **178**
Abbas Muhammad Rabiu, Abdulhamid Ibrahim, Aisha Muhammad Abdullahi, Mua'Zu Dauda, Maryam Ibrahim Muktar and Abdul'Azeez Yusuf

Disturbance Events Analysis and Partial System Collapse Hunting on the Nigerian Power System Using Frequency Disturbance Recorder (FDR) **185**

Saidu Kumo Mohammed, Usman Aliyu, Ganiyu Bakare and Yau Haruna

Effect of Youtube-Videoembedded Instruction On Students' Academic Achievement In Automotive Technology Education In Tertiary Institutions Of North-Eastern Nigeria 189
Isma'Ila Y. Shehu, Umar Abubakar and Aminu M. Kawu

honeyCAPTCHA: An Enhanced Intrusion Detection Model 193
M. Abdullahi, S. Aliyu, S. B. Junaidu

Blockchain: Panacea for Corrupt Practices in Developing Countries 198
Abdurrashid Ibrahim Sanka and Ray C.C. Cheung

Robust H-Infinity Control for Magnetic Levitation System 205
Kabiru Ibrahim, Ahmad Abubakar, Sulaiman Sulaiman and Abdullahi Kunya

Web-Based Cataract Detection System Using Deep Convolutional Neural Network 212
Musa Yusuf, Samuel Theophilous, Annah Hassan and Jadesola Emmanuel

Regulation of Induced Movement using Back Stepping Control 220
Mohammed Ahmed, Mohammed Saiful Huq and Babul Salam Ksm Kader Ibrahim

Steamblockly: A Block-Oriented Programming Environment to Enhance and Engage Interest in Software Engineering 228
Jerry D. Emmanuel, Afolayan A. Obiniyi and Donfack A. F. Kana

An Improved Digital Imagery Compression Approach Using Huffman And Chinese Remainder Theorem Schemes 233
Mohammed Ibrahim and Kazeem Gbolagade

Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant 241
Sani Isah Abba, Rabiu Aliyu Abdulkadir, Muhammad Gaya Sani, Mubarak Auwalu Saleh, Parvaneh Esmaili and Jazuli Abdullahi

Learning from Imbalanced Data 248
Danung Simtong, Musa Yusuf and Peter Ogedebe

Fraud Detection in Credit Card and Application of VAT Clustering Algorithm: A Review 254
Abdulkabiru Abdulrazaq, Muhammad Bashir Abdulrazaq, Adedokun Adewale and Umoh Ime

Investigation of Frequency Dynamics on the Nigerian Power System Using Frequency Disturbance Recorder (DFR) 261
Saidu Kumo Mohammed and Yau Haruna

Industrialization of The Nigerian Economy: The Imperatives Of Imbibing Artificial Intelligence And Robotics For National Growth And Development 265
Victor Akang, Matthew Afolayan, Msuega Iorpenda and Janet Akang

Improved Trajectory Planning for FES-Induced STS Movement Restoration in Paraplegia 281

Mohammed Ahmed, Mohammed Saiful Huq and Babul Salam Ksm Kader Ibrahim

Detection and Sentiment Analysis of Hate Speech on Twitter in Nigerian Politics **285**
Garba Aliyu, Fatima Rilwan and Sandip Rakshit

A linguistic Approach to Relational Databases Using Fuzzy Logic **289**
Ridwan Salahudeen, Garba Aliyu and Armand Donfack Kana

Implementation of a Data Transmission System using Li-Fi Technology **294**
Ifada E, Nazmat Surajudeen-Bakinde, Nasir Faruk, Abubakar Abdulkareem, Mohammed O O and Otuoze A O

Analysis of A GSM Network Quality of Service Using Call Drop Rate And Call Setup Success Rate As Performance Indicators **300**
Abdoulie Momodu Sunkary Tekanyi, Hassan Abdulkareem and Abduljalal Yusha'U Kassim

Analysis of GSM Network Quality of Service Using Call Setup Failure Rate and Handover Failure Rate Indices **306**
Abdoulie Momodu Sunkary Tekanyi, Hassan Abdulkareem and Ziyaulhaq Zakariyya Muhammad

Evaluating the State of the Art Antivirus Evasion Tools on Windows and Android Platform **311**
Faisal Ali Garba, Khadijah Musa Muhammad, Nasir Nuhu Wali, Abubakar Balarabe Isa, Kabiru Ibrahim Kunya and Shazali Ibrahim

Real Time Analysis of Telephone Conversation to Detect and Prevent Attacks in Nigeria **314**
Garba Aliyu, Ridwan Salahudeen and Sandip Rakshit

Design and Implementation of a Solar Powered Security/Garden Light **317**
Sunday Iyere, J. B. Erua, Joseph K. Yeboah and Adetokunbo Macgregor John-Otumu

Big Data Analytic: Strategic Tool For Counter Insurgency **326**
Sagir Lawan

Combined Cycle Power Plant Technology: Prospects in Boosting Nigeria's Electricity Production **331**
Fatima Sule Alhassan, Sadiq Thomas, Rabiu Aminu, Zainab Jibril Barau, Oluseun Oyeleke and Rukaiya Abdullahi

Energy Harvesting: Battery Less Life **336**
Abdullahi Oluwadamilola Adegoke, Armagan Dursun, Oluseun Oyeleke, Sadiq Thomas and Dominic Nyitamen

Challenges to Blue Screen Composition Detection in Digital Videos **339**
Kasim Shafii, Mustapha Aminu Bagiwa, Nura Sulaiman and Abubakar Usman Muhammed

A Secured Automated Electronic Bimodal Biometric Voting System **342**
Kennedy Okokpujie, Etinosa Noma-Osaghae, Donald Ehikioya Ufuah, Martins Nwagu, Samuel Ndueso John and Charles Ndujiuba

Development of Facial Trait Recognition System **346**
Kennedy Okokpujie, Etinosa Noma-Osaghae, John Abubakar, Imhade P. Okokpujie, Samuel Ndueso John and Charles Ndujiuba

A High Gain Patch Antenna Array for 5G Communication **350**
Zakariyya Olayinka Sikiru and Bashir Sadiq

Crone Controller Based Speed Control of Permanent Magnet Direct Current Motor **356**
Idris Garba, Yusuf Abubakar Sha'Aban, Muhammed Bashir Mu'Azu and Zaharaddeen Haruna

Validation of Aerodynamic Coefficients for Flight Control System of a Medium Altitude Long Endurance Unmanned Aerial Vehic **366**
Mudashir Kadiri, Ameer Mohammed and Sunday Sanusi

Flight Performance of Solar-Powered Unmanned Aerial Vehicles: A Review **370**
Aminu Hamza, Ameer Mohammed and Abubakar Isah

Electric Power-Assisted Steering: A Review **375**
Abubakar Isah, Ameer Mohammed and Aminu Hamza

Towards morphing wing technology in aircraft for improved flight performance **381**
Emmanuel Zebedee, Ameer Mohammed and Muhammad Sharrif Lawal

Implementing Flash Event Discrimination Policy in Internet Protocol Traceback using Shark Smell Optimization Algorithm **389**
Omoniyi Wale Salami, Ime Jarlath Umoh, Emmanuel Adewale Adedokun, Muhammed Bashir Mu'Azu and Lukman Adewale Ajao

Modification of bacterial foraging optimization algorithm using elite opposition strategy **398**
Danlami Maliki, Mohammed Bashir Muazu and Olaniyi Olayemi Mikail

Energy efficient learning automata based QL-RACH (EELA-RACH) access scheme for cellular M2M communications **406**
Nasir Aliyu Shinkafi, Lawal Mohammed Bello, Dahiru Shu'Aibu and Ibrahim Saidu

Hauwe: hausa words embedding for natural language processing **412**
Idris Abdulmumin and Bashir Shehu Galadanci

Comparison of various instantaneous power methods for the estimation of time-parameters of poly phase coded radar signal **418**
Ashraf Adamu Ahmad, Muhammad Ajiya, Lawal Mohammed Bello and Zainab Yunusa Yusuf

Cultured artificial fish swarm algorithm: an experimental evaluation **425**
Imam Maryam Lami, Gbenga Olarinoye, Busayo Adebiyi, Muyideen Omuya Momoh and Habib Bello-Salau

Speed control of a three phase induction motor using a pi controller **431**
Gbenga Olarinoye and Charles Akinropo

Usability and accessibility evaluation of Nigerian mobile network operators' websites **438**
Solomon Adepoju, Ayobami Ekundayo, Oluwaseun Adeniyi Ojerinde and Ahmed Rofiat

LWT-CLAHE based color image enhancement technique: an improved design **445**
Oluwakemi Marufat Gazal, Adewale Adedokun, Ime Jarlath Umoh and Muyideen Omuya Momoh

Impact of household construction materials on wi-fi (2.4 Ghz) signal **457**
A. Musa, M. Ahmed and R. Buhari

Cultural evolution from nomadic pastoralism to nomadic pastoralist optimization algorithm (NPOA): the mathematical models **459**
Ibrahim Abdullahi, Muhammed Mu'Azu, Olayemi Olaniyi and James Agajo

Sinkhole attack detection in a wireless sensor networks using enhanced ant colony optimization to improve detection rate **466**
Kenneth E. Nwankwo and Shafi'I Muhammad Abdulhamid

Application of filter bank multicarrier offset quadrature amplitude modulation to large MIMO systems: a review **472**
Mahmud Jaafar, Suleiman Muhammad Sani, Abdulmalik Shehu Yaro and Muazu Jibrin Musa

Modified handover decision algorithm in long term evolution advanced network **480**
Abbas Sada, Z.M Abdullahi and A.D Usman

Multi-tenancy cloud-enabled small cell security **485**
Babangida Albaba Abubakar and Haralambos Mouratidis

Improving energy efficiency of low density parity code system **491**
Olawoyin Lukman Abiodun, Nasir Faruk, Olaniran Kayode Olawale, Abdulkarim Oloyede, Olumuyiwa Ayeni, Sakariyah Adewole and Ismaeel Sikiru

Tools for analytical and simulation investigation of named data networking **496**
Muktar Hussaini, Habib Bello-Salau, Hiba Jasim Hadi, Shahrudin Awang Nor, Amran Ahmad and Ibrahim Muhammad Hassan

Mobility support challenges for the integration of 5g and iot in named data networking **505**
Muktar Hussaini, Shahrudin Awang Nor, Habeeb Bello-Salau, Hiba Jasim Hadi, Aminu Abbas Gumel and Kabiru Abdullahi Jahun

Privacy token technique for protecting user's pii in a federated identity management system for the cloud environment **513**
Maria Abur, Sahalu Junaidu, Saleh Abdullahi and Afolayan Obiniyi

Delay aware power saving scheme (dapss) base on traffic load in ieee 802.16e networks **523**
Daniel Dauda Wisdom, Ahmed Yusuf Tambuwal, Aminu Mohammed, Ahmed Audu, Micheal Bamidele Soroyewun and Samson Isaac

An approach for journal summarization using clustering based microsummary generation 531
Hammed Adeleye MOJEED, Shakirat Aderonke SALIHU, Ummu SANOH

A PC-Based Multifunctional Virtual Oscilloscope 537
Abdussamad U Jibiak, Ndubuisi D. Robinson

Machine Learning for Strategic Urban Planning 545
S. N. Odaudu, I. J. Umoh and M. B. Mu'azu

Buffer Occupancy Threshold Selection Strategy for Optimal Traffic Distribution Across a Hierarchical Network Onchip Structure 553
Umar I. Mohammed, Umoh J. Imeh, Abdulrazaq M.

PMU optimal placement in smart grid using wlav power system estimator 560

Mohammad Shoaib Shahriar, Yusuf Sha'Aban, Ibrahim Omar Habiballah and Farhan Ammar Ahmad

A Direct Optimal Control of the Jebba Hydroelectric Power Station 565
Olalekan Ogunbiyi, Cornelius Thomas and Busayo Adebiyi

Big Data Analytics: Emerging Strategy for the Ubiquitous Computing Era

Rasheed Gbenga JIMOH, Ph.D.
Department of Computer Science
University of Ilorin, Nigeria
Jimoh_rasheed@unilorin.edu.ng

ABSTRACT

Due to the fact that individuals and organizations (corporate/public) are dealing with data in high volume, velocity and varieties. This is applicable to the capturing, storing and processing/analysis of such data. Currently, over 2 billion people worldwide are connected to the Internet, and over 5 billion individuals own mobile phones. By 2020, 50 billion devices are expected to be connected to the Internet. As information is transferred and shared at light speed on optic fiber and wireless networks, the volume of data and the speed of market growth increases. However, the fast growth rate of such large data generates numerous challenges, such as the rapid growth of data, transfer speed, diverse data, and security. Nature has therefore forced it on us the approach of big data analytics. Hence, this paper discusses big data analytics as a strategy for sustaining the ubiquitous computing era.

Keywords: Big data analytics, Strategy, Ubiquitous, Computing, Era

Reinforcement Learning for Playing Games

Marco A. Wiering

Department of Artificial Intelligence
University of Groningen
Groningen, The Netherlands
m.a.wiering@rug.nl

ABSTRACT

Reinforcement learning (RL) algorithms enable an agent to learn to play games from its interaction with a specific game simulator and the obtained rewards during game-play. This approach has been successfully used to create computer players which are better than human experts in many types of games such as Go and StarCraft II. Most RL algorithms make use of value functions that are trained using the experiences of the agent to predict the future cumulative reward given that an action is selected in a state, which allows for selecting the most promising action. Because games usually contain a huge amount of possible states, RL algorithms are generally combined with function approximation techniques such as (deep) neural networks to approximate the value function. In this presentation, we will first look at the advantages and disadvantages of using human engineered features and hallow neural networks compared to using raw game data such as pixels and deep neural network architectures. Then, our novel deep RL algorithms are described and their competitive results on a benchmark of Atari 2600 games are shown.

DESIGN AND IMPLEMENTATION OF AN ELECTRICITY ON-LINE BILLING PAYMENT SYSTEM

A. Andrew[1]
Nigerian Electricity Regulatory Commission, Phase 2, 274 10th Street, Central Business District
Abuja –Nigeria.
nokamanda@gmail.com

P. U. Okorie[2]
Department of Electrical Engineering, Faulty of Engineering, Ahmadu Bello University, Zaria.
Zaria - Nigeria
patrickubeokorie@yahoo.com

A.I. Abdu[3]
Department of Computer Engineering, Faulty of Engineering, Ahmadu Bello Univesity, Zaria
Zaria - Nigeria
aiabdu@abu.edu.ng

Abstract — ***Electricity consumers are often faced with the problem of inaccuracy and delay in monthly billing due to the drawback in reading pattern and human errors. Thus, it is essential to have an efficient system for such purposes via electronic platform with consideration to proximity. The proposed system automates the conventional process of paying electricity bill by visiting the Electricity Board which is tiresome and time consuming. It is also designed to automate the electricity bill calculation and payment for user convenience. The system is developed with Microsoft Visual Studio using C# as the base programming language which can be used to develop websites, web applications and web services. The Microsoft Structured Query Language (SQL) server is also used for creating back-end database. The system would be having two logins: the administrative and user login. The administrator can view the user's account details and can add or update the customer's information of consuming units of energy of the current month in their account. The Admin has to feed the system with the electricity usage data into respective users account. The system then calculates the electricity bill for every user and updates the information into their account every month. Users can then view their electricity bill and pay before the month end.***

Keywords —Billing system, bill payment, C#, e-payment, internet billing, IOT, PBS, SQL

I. INTRODUCTION

The Electricity Online Bill Payment web application is provided to all the users who want to pay their electricity bill. This web application can reduce the pressure of the user to stand in queues for bill payment and charges. This limitation has led to the development of a small model that enables the user to perform all the payment of bills and charges sitting at the comfort of his/her own desk. The importance of an Electricity Online Bill System cannot be over emphasized because its calculation reflects the exact power consumption for prospective consumers, and in monitoring the billing details of electricity consumers [1]. It provides an environment to maintain the consumer details starting from receiving bill, making bill payments and so on [2]. Consumers can lodge complaint and make their bill payments just by logging into the system.

system to capture data related to the consumer's profile in order to assign an identification code with transactions relating to power billing. It constitutes various modules, among which the administrator and consumer module are integral. The consumer is granted access only through the username and password created from the first visit to the online system, or while registering at the web portal. The administrator module is handled by an authorized Electricity Board employee, in order to grant request relating to validating every transaction online and to order or confirm payment via the electronic system.

The Microsoft Visual Studio and Microsoft Structured Query Language (SQL) server will be used as front end and back end databases respectively for developing the project. Microsoft.

Visual Studio is primarily a visual design environment which will be used for creating the graphical user interface. The Microsoft SQL server is a structured query language-based client/server relational database. It can be used to efficiently create and manipulate database systems. Figure 1 shows a simple block diagram of the proposed online billing system:

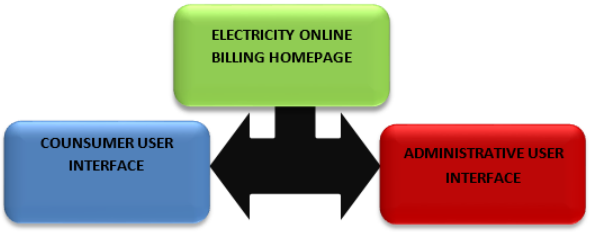


FIGURE I. PROPOSED ON-LINE BILLING SYSTEM

II. LITERATURE REVIEW

Various empirical and theoretical studies have been undertaken at the national and international level to analyze various online bill payment systems. The studies mainly focus on online bill payment systems such as: an online power billing system, an e-payment system, a mobile based billing system and so on. These services have not only improved the satisfaction level of customers, but it has also helped in reduction of processing time and transaction time. The survey of literature review covers some major works that have been carried out on online electricity billing. These include but are not limited to:

In [3] this paper presented the design and implementation of a web-based application with online capability called Power Billing System (PBS). PBS is a solution system developed with Microsoft Visual Web Development and Microsoft Access with Structured Query Language (SQL) for back-end database. In [4] the paper, proposed an online payment scheme which uses the traditional e-payment infrastructure but which

reveals no payment information to the seller. In [5] this paper presented the design of an advanced centralized billing system using Internet of Things (IOT). In [6] this paper, the author first identified some vulnerability in the mobile billing system. The proposed system can achieve authentication, non-repudiation, and fairness, which are desirable requirements for an undeniable mobile billing system. The billing system design based on internet environment was proposed by [7]. Concerning the performance of four network scenarios for billing purposes was presented and discussed by [8] and the result shows that the environments that use differentiated services are both convenient for customers and service providers. The authors of [9] also trying to improve the e-payment system with a smart card. In this paper, the third party can link a payment to a corresponding withdrawal to prevent money laundering and blackmailing.

III. PROPOSED METHOD

This section gives the detailed design and operation of the proposed system. It also explains in details the operation and functions of each module in the proposed system.

A. Proposed System

Considering the anomalies into the existing system, a computerized system is built using Asp.Nnet with C# as a base language. The system enhances and upgrades the old existing system by increasing its efficiency and effectiveness. The software improves the working methods by replacing the existing manual system with the computer-based system. The proposed system automates each and every activity of the manual system and increases its throughput. Thus, the response time of the system is less and works very fast. The system uses a quick response with very accurate information regarding the user's electricity bill information. The proposed system has a very user-friendly interface; thus the user will feel very easy to work on it. The software provides accuracy along with a pleasant interface. The transactions reports of the system can be retried as and when required. Thus, there is no delay in the availability of any information, as whatever will be needed, can be captured quickly and easily.
The Microsoft Visual Studio will be used as a front end and The Microsoft Structured Query Language (SQL) as back end for developing the project. Visual studio is primarily a visual design environment. This design environment will be used to create text boxes, buttons and adding support codes in the respective modules (the administrator and user modules). The Microsoft SQL server is a powerful database application with which the user can efficiently create and manipulate database systems.

B. System Modules and their Description

The system comprises of two major modules, which are further divided into sub- modules. They are as follows:

i. Administrative Login: In the administrative login, the administrator has the authority of the system. The administrator can add, delete notifications and update the system. The administrator performs the following functions on the system:

- Consumers Registration: The Admin will register the consumer by entering the consumer's basic registration details such as: Name, Contact Address, Residential Address, Consumer type, Password and Email Address.
- View Registered Users: The Administrator can also view all the registered users in the system, make necessary updates and update the bill status of the consumers.

ii. Consumer Login: Consumers can login into the system and perform the following tasks:

- View Bill: Consumers can view their balance electricity bill amount of each month.
- Consumption Calculation: Consumers can calculate the total amount of units they consume in a month using a consumption calculator.
- Bill Calculation: Consumers can calculate the total number total amount that he/she will pay based on the units consumed, using a bill calculator.

C. Database Design and Tables

Many physical database design decisions are implicit or eliminated when the database management technology is chosen to use with the information system that is being designed. Since many organizations have a standard for operating systems, database management system, and data access languages, only choices that are not implicit are dealt with in the given technologies. The primary goal of a database design is data processing efficiency.

There are four different tables used for storing different records in the system. They include: Consumer Table, Administrator Table, Bill Charge Table, and Bill Calculation Table. Tables 1 to 4 show the consumer table, bill charge table, administrator table and bill calculation table respectively.

TABLE I: CONSUMER TABLE

Field Name	Data Type	Size	Constraint
User id	Integer	Nil	Nil
Consumer id	Integer	Nil	Nil
Firat Name	Varchar	50	Nil
Last Name	Varchar	50	Nil
Username	Varchar	50	Nil
Password	Varchar	50	Nil
Email Address	Varchar	50	Nil

Contact Address	Varchar	50	Nil
City id	Int	-	Nil

TABLE II: BILL CHARGE TABLE

Field Name	Data Type	Size	Constraint
Bill id	Integer	Nil	Primary Key
Bill Number	Integer	Nil	Nil
Fix Charge	Integer	Nil	Nil
Maintenance Charge	Integer	Nil	Nil
Total Charge	Integer	Nil	Nil
Total Unit	Integer	Nil	Nil

TABLE III: ADMINISTRATOR TABLE

Field Name	Data Type	Size	Constraint
Customer id	Integer	Nil	Primary Key
City	Varchar	50	Nil
State id	Integer	Nil	Nil

TABLE IV: BILL CALCULATION TABLE

Field Name	Data Type	Size	Constraint
Bill Number	Integer	Nil	Nil
Bill Date	Integer	Nil	Nil
Amount	Integer	Nil	Nil
City	Integer	Nil	Nil

These are the tables which are used in the development for storing different records. Every table has a primary key for storing a unique record.

Consumers are classified based on their mode of consumption and by tariff classification. Electricity tariff is defined as the rate at which energy is selling to the consumers (Othman, *et al* 2015). Usually electricity tariff are fixed by Government. In Nigeria Electricity consumers are divided into five categories namely: Residential, Commercial, Industrial, Street Lighting and Special Tariff respectively.

D. Formal Model of Proposed System

The formal model of the proposed system is represented in flowchart diagrams. All these models will give the conceptual view of and provide the graphical analysis of the user's requirements. As a major modeling tool, entity relationship diagrams helps in the organizing of the functional elements of the system into entries and also define the relationships between the entities. This process will enable the analyst to understand the database structure so that data can be stored and retrieved in the most efficient manner. The flowchart shows the flow of data from external entities into the system. It also shows how data moves from one process to another as well as its logical storage. Figures 2 and 3 shows the flowchart activity of the administrator and consumer respectively.

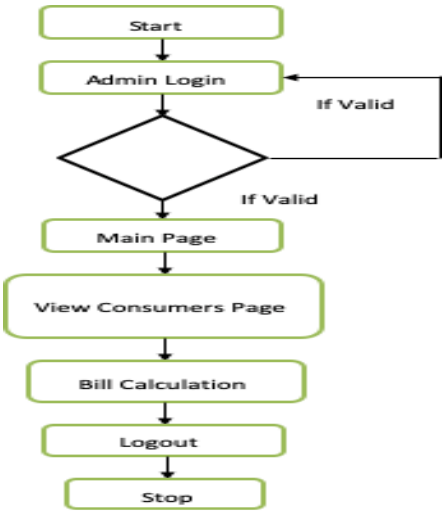


FIGURE II. ACTIVITY DIAGRAM OF THE ADMINISTRATOR

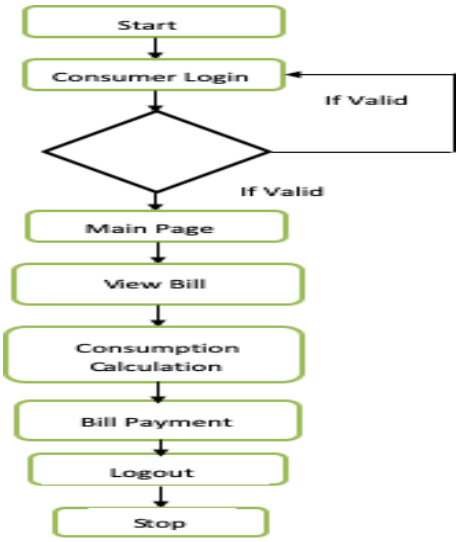


FIGURE III. ACTIVITY DIAGRAM OF THE CONSUMER

E. System Architecture

The system architecture gives the overview of the organizational system that shows the system boundaries, external entities that interact with the system, and the major information that flows between the entities and the system. Figure 4 shows the system architecture of the system

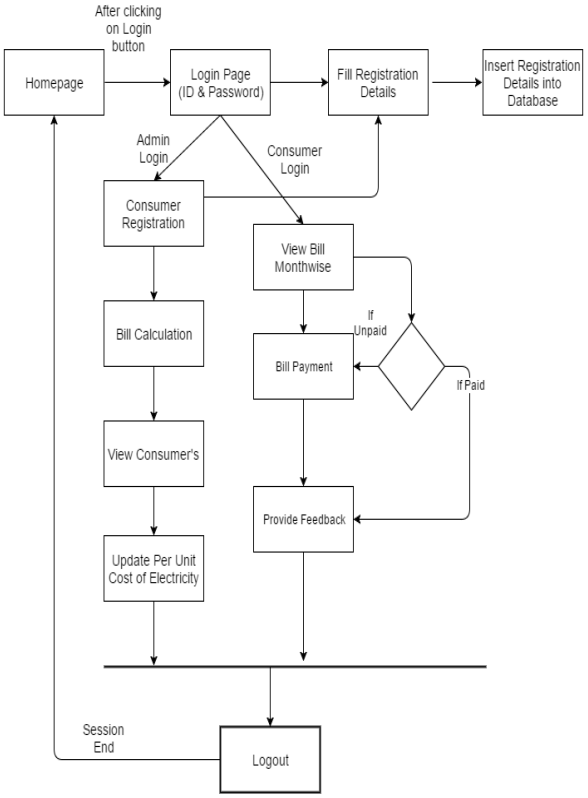


FIGURE IV. SYSTEM ARCHITECTURE OF THE SYSTEM

IV. SYSTEM IMPLEMENTATION AND OPERATION

The implementation was done using Microsoft Visual Studio and SQL server for backend application. The system captures information related to energy demand, payments, and so on, from various level of organization with the aim of capturing it from as close to the source as possible.

The system consists of two login pages namely: the administrator page and the consumer page. The system operation is described as follows:

- After connecting online, the application starts by displaying the homepage. There after the username and password are requested for the specified status (the consumer or administrator). If it is a new user, the user will have to register by filling all the required registration details.
- If the password and username details are valid, the consumer then gets access to the main page. On the main page, the consumer can:

1. View his/her bill status
2. Use the consumption calculator to calculate the total amount of units he/she will want to consume. Figure 5 shows the consumption calculator page.

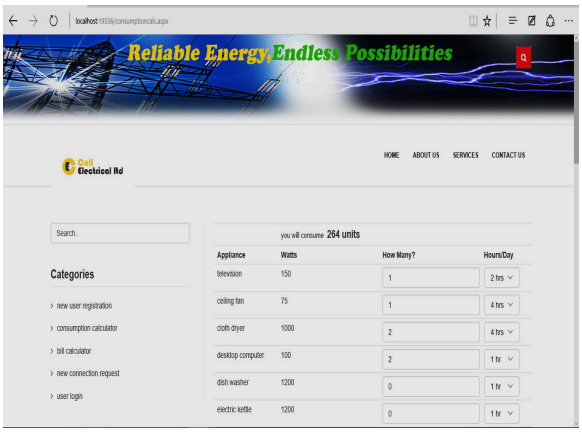


FIGURE V. CONSUMPTION CALCULATION PAGE

The user starts by selecting the appliance he/she uses on daily basics. The wattage of each of the appliance is provided. If a particular appliance is not listed, the user can select the 'others' option and insert the wattage of that particular appliance. The calculator now calculates the total amount of wattage of all the appliances based on the selections made by the user. The user now selects the total number of the appliances that he/she has and the total time per day he/she uses the appliances. The consumption calculator now uses this information to calculate the total units consumed in kWh. The consumption calculator uses eqn. (1) to calculate the total unit consumed in kilo Watt hour (kWh) per month:

$$\frac{kWh}{m} = \frac{ABC}{1000} x\,30 \qquad (1)$$

Where:

Kwh/m is the total kWh per month

A is the total wattage of the appliances;

B is the number of appliances;

C is the number of hours used daily.

From Figure 5 the total units consumed using eqn. (1), is calculated as follows:

For the Television:

Total kWh consumed = [(150 × 1 ×2) ÷ 1000] × 30

Total kWh consumed = 9 kWh per month.

For the Ceiling fan:

Total kWh consumed = [(75 × 1 × 4) ÷ 1000] × 30

Total kWh consumed = 9kWh per month

For the Cloth dryer:

Total kWh consumed = [(1000 × 2 × 4) ÷ 1000] × 30

Total kWh consumed = 240 kWh

For Desktop computer

Total kWh consumed = [(100 × 2× 1) ÷ 1000] × 30

Total kWh consumed = 6 kWh

The total kWh consumed for the month = 9kWh + 9 kWh + 240 kWh + 6 kWh

Total kWh consumed for the month = 264 kWh per month

3. Use the bill calculator to calculate the total amount he/she will pay based on the units consumed as shown Figure 5. The total billing charge can be computed by using (eqn. 2).

$$\gamma = \varphi\omega \quad (2)$$

Where:

γ *is the total bill charge in Naira*

φ *is the tariff (unit cost),* *and* ω is the total units consumed (kWh/m)?

From Figure 5, the total tariff is ₦10.00 therefore the total bill in Naira for 264 kWh per month using eqn. 2 is given as:

Total bill in Naira = ₦10.00 × 264 kWh

Total bill in Naira = ₦2640

4. After the total bill is calculated, the consumer then clicks on the 'Pay Now' button. This will direct the user to the bill information page, where other charges such as the maintenance fee charge, fixed charge, and Value Added Tax (VAT) charge are added to the unit charge give the grand total bill the consumer will pay. Figure 6 shows the bill information page.

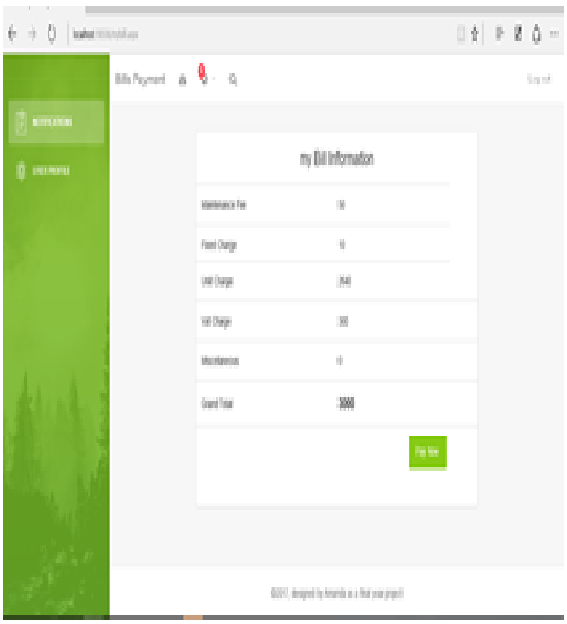


FIGURE VI. BILL INFORMATION PAGE

The grand total bill is calculated using eqn. 3

$$\beta = \alpha x \delta x \mu x \rho x \tau \quad (3)$$

Where:

β *is the grand total bill in Naira,*

α *is the maintenance fee,*

δ *is the fixed charge,*

μ *is the unit charge (kWh).*

ρ *is the VAT' charge, and* τ *is the miscellaneous* (2)

From Figure 6, the grand total bill calculated using eqn. (3) is given as:

Grand total bill in Naira = 50 + 10 + 2640 + 300 + 0

Grand total bill in Naira = ₦3000

5. After the grand total bill is calculated, the consumer then clicks the 'pay now' button. The consumer is then directed to the Internet Payment Gateway page where he/she makes payment for the electricity bill.

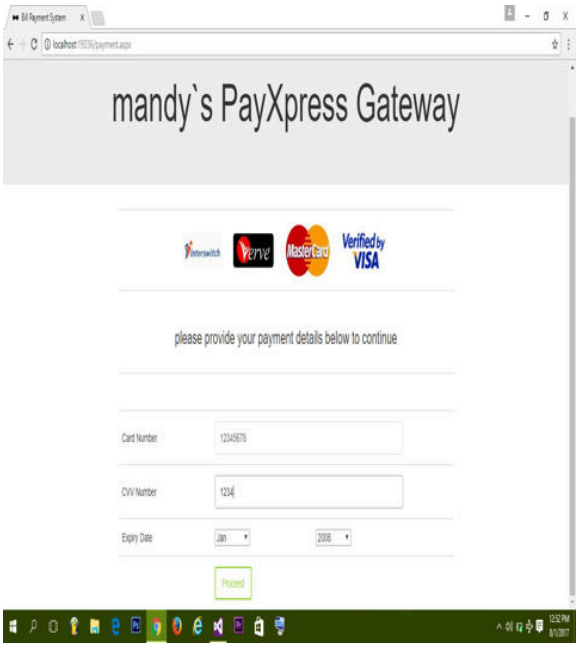


FIGURE VII. INTERNET PAYMENT GATEWAY PAGE

6. After the payment has been successfully made, the consumer then gets a 12 digit token code which will be sent to the notification bar in the profile page of the consumer. Which the consumer will enter into the power meter. The consumer can then successfully logout out of the system.

V. CONCLUSION

Usability testing was part of the post implementation review and performance evaluation for the Electricity Online Bill Payment System, in order to ensure that the intended users of the newly developed system can carry out the intended task effectively using real data so as to ascertain the acceptance of the system and operational efficiency. It caters for consumers' bills and also enables the administrator to

generate monthly reports. It is possible for the administrator to know the consumers have made payment in respect of their bills for the current month, thereby improving the billing accuracy, reduce the consumption and workload on the Electricity Board employees or designated staff., increase the velocity of electricity distribution, connection, tariff scheduling and eliminates variation in bills based on market demand. The conceptual framework allows necessary adjustments and enhancement maintenance to integrate future demands according to the technological or environmental changes with time. It manages the consumers' data and validates their input with immediate notification centralized in Electricity Board offices across the nation.

REFERENCES

[1]. Arimoro, T. A., Oyetunji, A. K., & Odugboye, O. E. (2019). Analysis of Electricity Billing System in Corporate Buildings in Lagos, Nigeria. *Studies*, *1*(6), 10-20.

[2]. Panthala, S., Islam, N., & Habib, S. A. (2015). Automated industrial load measurement system.

[3]. Adegboyega, A., Gabriel, A. A., Ademola, A. J., Victor, A. I., & Nigeri, K. (2013). Design and Implementation of an Enhanced Power Billing System for Electricity Consumers in Nigeria. *African Journal of Computing & ICT*, *6*(1).

[4]. Antoniou, G., Batten, L., Narayan, S., & Parampalli, U. (2009). A privacy preserving e-payment scheme. In *Intelligent Distributed Computing III* (pp. 197-202). Springer, Berlin, Heidelberg

[5]. Rahul K. Sai (2016) Advanced Centralized Electricity Billing System Using (IOT). International Journal of Advanced Research in Computer Science and Software Engineering icrosoft Developer Network (MSDN): http://msdn2.microsoft.com/en-us/default.aspx

[6]. Li, S., Wang, G., Zhou, J., & Chen, K. (2009). Fair and secure mobile billing systems. *Wireless personal communications*, *51*(1), 81-93.

[7]. Al-Ani, M. S., & Noory, R. (2012). Billing system design based on internet environment. *Editorial Preface*, *3*(9).

[8]. Barreto, P. S., Amvame-Nze, G., Silva, C.V., Oliveira, J. S. S., de Carvalho, H. P., Abdalla, H., ... & Puttini, R. (2005, April). A study of billing schemes in an experimental next generation network. In *International Conference on Networking* (pp. 66-74). Springer, Berlin, Heidelberg.

[9]. Bo, Y., Dongsu, L., & Yumin, W. (2002).An anonymity-revoking e-payment system with a smart card. *International Journal on Digital Libraries*, *3*(4), 291-2

Proceedings of the 2nd International Conference of the IEEE Nigeria Computer Chapter:
IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

Genetic Algorithm Based Dual Input Power System Stabilizer

Aliyu I. Usman
Department of Electrical Engineering
Bayero University
Kano, Nigeria
aliyubit@gmail.com

N. Magaji
Department of Electrical Engineering
Bayero University
Kano, Nigeria
nmagaji2000@gmail.com

***Abstract*— This study presents a new generator's Power System Stabilizer design methodology of a single machine infinite bus. A dual input power system stabilizer genetic algorithm based (GAPSS4B) is analyzed, it performance is improved as compared to the single input genetic Algorithm based Stabilizer (GAPSS) over different operating points. The optimal parameters of both GAPSS and GAPSS4B are obtained by minimizing the bode plot of the transient model of SMIB with help of genetic algorithm (GA). The proposed technique is evaluated against the single input genetic algorithm-based PSS (GAPSS) at a SMIB considering system parametric uncertainties. The simulation results show that the proposed dual input GAPSS provides improved performance as compared to single input GAPSS.**

Keywords-- Dual input power system stabilizer, genetic algorithms, low frequency oscillations, time domain, SMIB System

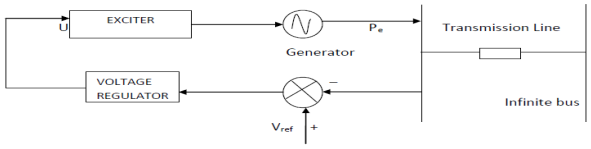
I. INTRODUCTION

To enhance dynamic stability of power systems, power system stabilizers (PSS) need to be Apply so as to damp out rotor oscillation within the range of low frequencies (Usually 0.2-2.5 Hz) due to disturbances are used for years. Excitation system that is incorporated in the Single input PSS of a generator, is one of various methods used to stabilize power system oscillations. One of the efficient way to enhance the power system stability is Excitation control also known as Automatic Voltage Regulator (AVR) [1]. Conventionally, automatic voltage regulator controls generator output regardless of the perturbations, as such the output voltage and reactive power can be preserved at a desired values [1]. The action of the Automatic Voltage Regulator in steady state, the generator's speed deviation of the power system is almost zero. Though, in the transient state, the rotor angle changes causing the rotor to swings therefor the entire oscillate, and affect terminal voltage of the generator. Rotor Oscillation in the power system generally can be addressed by employing Power System Stabilizer (PSS) that has AVR in the generator, as such the oscillation can be damp out Rotor speed is usually used as the pss input and its output is fed into the AVR input. PSS was found very effective tool for improving generator's oscillation [4]. Though, the effectiveness of the conventional PSS becomes un-optimal because of the variations in system parameters due to frequent change in load [1]. Delta-Omega (rotor speed deviation input) PSS is frequently used. Compensation technique is commonly used in optimizing the Delta-omega PSS [1], [2]. This is the most straightforward approach, easily understood and implemented in the field. However, the gain setting is obtained by trial and error approach [2]. The optimization technique adopted in this research work is genetic Algorithm to optimize pss parameters

The performances of the single and double input pss GAPSS and GAPSS4 respectively are analyzed and compared in order to demonstrate the effectiveness of GAPSS4B in contrast to GAPSS. Genetic algorithm is known to be very effective optimization tool, Optimal tuning of PSS parameter is carried out with the help of genetic algorithm. This study deals with a design methodology for the stability enhancement of a SMIB as a base system, using dual input PSS (PSS4B), where GA is used in tuning its parameters. To show the simulation results, single machine infinite bus (SMIB) is use as base to show the performance of the proposed technique.

II. SMIB SYSTEM UNDER INVESTIGATION

Single machine infinite bus is the most suitable and simple machine used in demonstrating the effectiveness of



power system stabilizer, it is made up of a generator sending power to the infinite bus connected by a transmission line, bellow is the schematic diagram given in fig. 1.

Fig. 1 Single Machine Infinite Bus

To study the electromechanical behavior of a generator and power system single machine infinite bus system will be the simple system to be used. The model developed by Heffron- Philips of a SMIB system was used to investigate the effect of frequency oscillation within the range of 0.2-2.5 Hz and to test the effectiveness of the propose pss tuning

technique, but the linear model will be used to tune the pss parameter[5].

III. DYNAMIC MODEL OF DOUBLE INPUT POWER SYSTEM STABILIZER

State space model of the system for small disturbances was derived from the transfer function of Fig. 2 bellow.

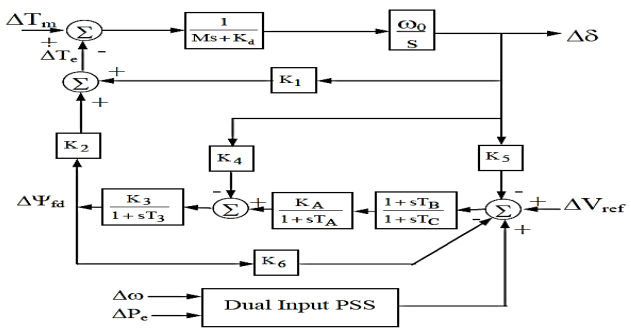


Fig.2 SMIB system and dual input pss

NOMENCLATURE

ω	= Synchronous speed
δ	= Synchronous angle
P_m	= Input Mechanical power
P_e	= Output electrical power
M	= Inertia
E_q	= Q axis voltage
E_{fd}	= Field voltage
E_q'	= Transient voltage of Q axis
T_{do}'	= Transient time constant of Q axis
K_a	= Excitation system gain
T_a	= Excitation system time constant
V_t	= Terminal voltage
V_{ref}	= Reference voltage of excitation system
T_m	= Mechanical torque
CPSS	= Conventional power system stabilizer
GAPSS	= Genetic Algorithm power system stabilizer
GAPSS4B	= Dual inputs Genetic Algorithm power system stabilizer
GA	= Genetic Algorithm
ITAE	= Integral of time multiplied Absolute value of the Error
PSS	= Power System Stabilizer

The non-linear model of the system can be linearized about any particular operating point. The equations bellow describes the steady state of the model

The system Nonlinear Dynamic Model is given as in (1)

$$\begin{aligned}\dot{\omega}(t) &= \frac{P_m - P_e - D\Delta\omega(t)}{M}\\ \dot{\delta}(t) &= \omega_0(t)\left(\omega(t)-1\right)\\ E_q &= \frac{\left(-E_q + E_{fd}\right)}{T_{do}}\\ \dot{E}_{fd} &= \frac{-E_{fd} + K_a\left(V_{ref} - V_t\right)}{T_a}\end{aligned} \tag{1}$$

To find the dynamic linear model of the system, the non-linear model has been linearized around a normal operating point. The linearized model given in (2), and the state space model is given as in (3).

$$\begin{aligned}
\Delta\dot{\delta} &= \omega_0 \Delta\omega \\
\Delta\dot{\omega} &= \frac{-\Delta P_e - D\Delta\omega}{M} \\
\Delta\dot{E}_q &= \left(-\Delta E_q + \Delta E_{fd}\right)/T_{do} \\
\Delta\dot{E}_{fd} &= -\left(\frac{1}{T_A}\right)\Delta E_{fd} - \left(\frac{K_A}{T_A}\right)\Delta v
\end{aligned} \tag{2}$$

$$\begin{bmatrix} \Delta\dot{\delta} \\ \Delta\dot{\omega} \\ \Delta\dot{E}_q \\ \Delta\dot{E}_{fd} \end{bmatrix} = \begin{bmatrix} 0 & \omega_0 & 0 & 0 \\ \frac{K_1}{M} & 0 & \frac{K_2}{M} & 0 \\ \frac{K_4}{T_{do}} & 0 & -\frac{K_3}{T_{do}} & \frac{1}{T_{do}} \\ \frac{K_A K_5}{T_A} & 0 & -\frac{K_A K_6}{T_A} & -\frac{1}{T_A} \end{bmatrix} \times \begin{bmatrix} \Delta\delta \\ \Delta\omega \\ \Delta E_q \\ \Delta E_{fd} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{M} & 0 \\ 0 & 0 \\ 0 & \frac{K_A}{T_A} \end{bmatrix} \times \begin{bmatrix} \Delta T_m \\ \Delta v_{ref} \end{bmatrix} \tag{3}$$

IV. POWER SYSTEM STABILIZER (PSS)

To enhance in damping out rotor oscillations of the a generator in the power system, Power System Stabilizer (PSS) is introduced. The power system stabilizer produces an electrical damping torque (ΔTe) in phase with the speed deviation (Δω) so that it amend the rotor swing [4]. In this study, PSS4B and PSS are developed.

A. Conventional power system stabilizer (CPSS)

The rotor oscillations of the synchronous generator can be damped using Conventional Power System Stabilizer by controlling its excitation [8]. To model the CPSS, the nonlinear system in Fig. 3 is used, the model comprises of a first filter (low pass), gain (k), second filter (washout high pass filter), third filter (lead-lag) and a limiter [9]. The amount of damping produced is determined by the general gain of the pss. Low frequencies present in the delta-omega signal is removed by washout high-pass filter. The lead-lag transfer functions is used for the compensation of phase lag generator.

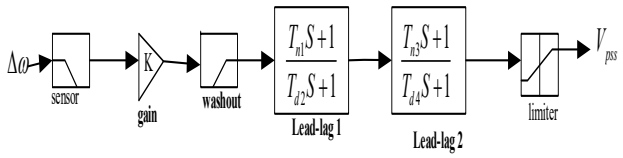


Fig. 3 Conventional power system stabilizer

B. Dual input power system stabilizer

Generator speed deviation and change in active power are commonly used as pss inputs, in both local and inter-area oscillation mode. When the former is used as input it has some advantages and short comings likewise the latter. In this work the combination of both is proposed. The parameter optimization should not be separate because unexpected instability may happen at the oscillation mode. We optimized the pss input signal together in this work.

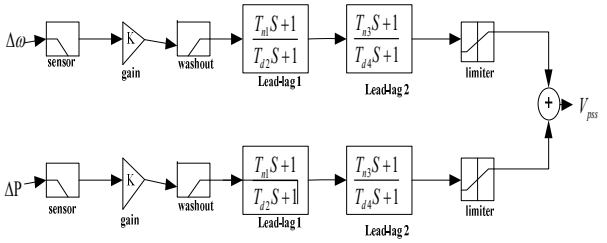


Fig. 4: Dual input power system stabilizer

V. GENETIC ALGORITHMS

Genetic algorithm is one of the optimization technique used in searching for a global optima. It mimic the operations of natural selection in genetics. It works on a population of individuals in which individual is selected at random and it's performance is sought, from the characteristics or objective function of the population, the fitness that is performance of each individual is calculated, From the individual performance a fitness value is obtained given by the objective function. The fittest individual has greater chance to survive and successively jump to the next generation continuously reducing the search focus to the region of best performance known as global optima or the fittest individual. Through genetic operators need to be apply in order to further modify the selected individuals. The genetic operators are selection, mutation and crossover, which are apply in order to ensure the successive generations. GA is good optimization tool, as it converges at the global optima unlike the conventional optimization tool which may converge at local optima.

A. GA operations

In nature, procreation and selection make individuals that adopt the environment to survive, and those that could not withstand the environment to pass a way. To use GA in solving a particular problem the following steps is followed: initialize the population (chromosomes) 'p' at random, each individual at the current populations is decoded and from it the fitness function, then selection which is done in accordance to their fitness value to form next generations, then new created individuals supersedes the current population using genetic operator, i.e. selection crossover and mutation.

1) Selection: From Darwin's theory of evolution only the best indibidual can survive and move to the next generation to creat the new offspring according to the theory of natural selection. The selection of best chromosome methods is employed in this theory, for example steady state The best chromosomes or the fittest individual has greter chance of survival.

2) Crossover: Crossover is also known as recombination. It is used to combine the chromosomes of two parent of the current genration at a particular point to produce two offspring, this done just to get new solutions

that will pass to the next generation. Crossover is demostrsted bellow between twoo parents A and B.

Parent A	0 0 0 0 0 1 0 1
Parent B	1 1 1 0 1 0 0 1
Child A	0 0 1 0 1 0 0 1
Child B	1 1 0 0 0 1 0 1

3) Mutation: This is a genetic operator used in maintaining diversity of a given generation to the next, by altering one or two gene of a single individual as demostrated for a child A.

Child A	0 1 0 1 1
New Child A	0 1 1 1 1

All the three genetic operators described above operates several times until the offspring found to be the best i.e. most fittest will put back the entire populations. If fresh solution of chromosome is developed, the new generation that totally replaces the parents are considered as the solution. It is important to know that genetic operator must be repeated until the population converges at the global optima known as optimum solution, depending on the complexity of the problem, the number increasing.

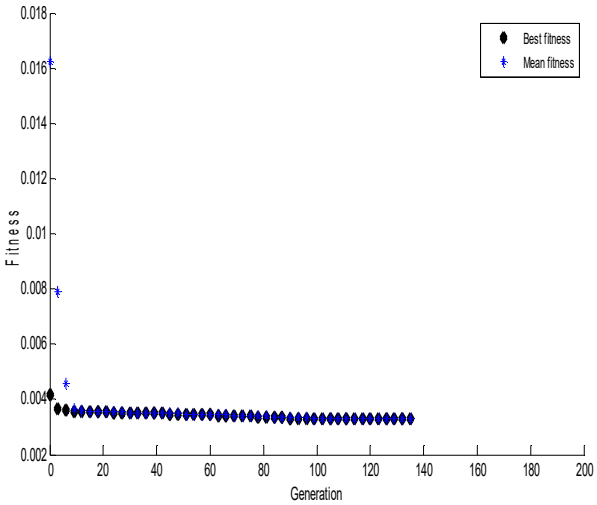
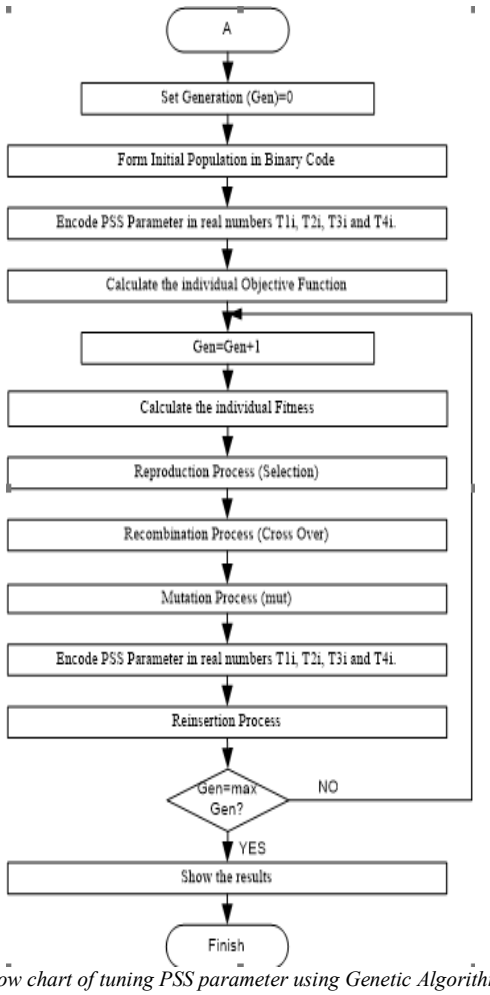


Fig. 5 Convergence characteristics of GA

VI. METHODOLOGY



Flow chart of tuning PSS parameter using Genetic Algorithm

The methods used in this paper for optimizing the time constants using GA is as follows:

- Set initial population of 50 individuals selection at a specified range.
- Derive the objective function value.

$$\mathrm{ITSE} = \int_0^t t\left[\Delta\omega(t)\right]^2 dt$$

- Run until when the erro is minimum, then keep time constants equal to that value gotten otherwise proceed to step 4.
- Regenerate the population of 50 individuals from parents using reproduction techniques. Go to step 2.
- The global optima is reached at 15 consecutive generations.

VII. LEAD LAG PSS DESIGN

Root locus and minimum phase are normally used in pss design. In the first approach the model is analyses and the eigenvalues is found and shifted until to left hand side of the s-plane. But in the second approach, the time constant of the pss are tuned to eliminate the lagging, using phase analysis. The latter is adopted in this work.

VIII. TRANSIENT MODEL OF A SMIB

For the purpose of modelling in the PSS tuning some Adjustment are made to the model of Fig. 2, just to make it suitable for the analysis and to know the amount of phase that the PSS should provide to the system so as to compensate the phase lag. To add damping to the rotor oscillations, the PSS has to produce a torque components in phase with the rotor speed deviations [2]. In order to achieve this, phase lag has to be compensated. Fig. 5 bellow shows the transient model of the systems.

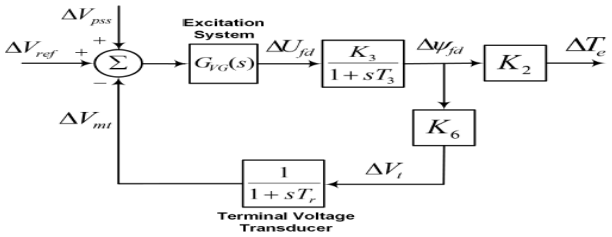


Fig. 5 Transient Model of SMIB [6]

The parameters of the phase compensation filter stage can be tuned base on (4).

$$H(s) = K * \frac{1}{1+sT_m} * \frac{sT_w}{1+sT_w}\left(\frac{1+sT_{lead}}{1+sT_{lag}}\right)^N \qquad (4)$$
$$= KH_1(s)$$

$$N = \frac{\theta_{pss}}{55^o} = \begin{Bmatrix} 1 \rightarrow \theta_{pss} \le 55^o \\ 2 \rightarrow \theta_{pss} \le 110^o \\ 3 \rightarrow \theta_{pss} \le 180^o \end{Bmatrix} \qquad (5)$$

$$\sigma = \frac{1 - \sin\frac{\theta_{pss}}{N}}{1 + \sin\frac{\theta_{pss}}{N}}$$

$$T_{nl} = \frac{1}{\omega_{osc}\sqrt{\sigma}}$$

$$T_{dp} = \sigma T_{nl}$$

To form the fitness function, there are two variables, phase margin θpss and corresponding frequency ωosc which can be derived from the bode plot of the transient model of the system, thus (5) can be minimize by GA, and the performance index is delibrates as in (6). Integral of the Time multiplied Square value of the Error (ITSE), is the performance index.

$$ITSE = \int_0^t t\left[\Delta\omega(t)\right]^2 dt \qquad (6)$$

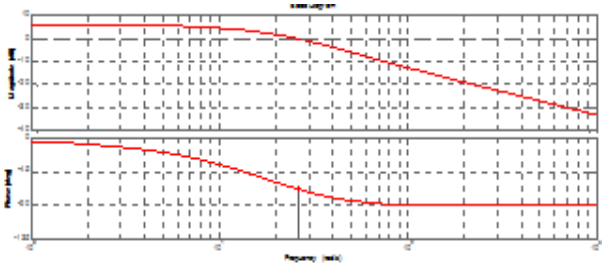


Figure 6 Bode Plot of the Transient Model

From the transient model bode Plot, the phase Margin θpss is and corresponding frequency ωosc is hence from the value of θpss N is two that is two number of filters in cascade are required. By adjusting the plot new θpss and its corresponding ωosc is tabulated as given in table 1. Therefor the value of the leading time constant and that of the lagging time constant can be calculated from (4). And listed in table 2.

A 0.1 step change in the reference mechanical torque (ΔTm) is assumed, in order compute the optimum parameter values, and (5) is minimize using GA. The optimum values of the parameters are listed in the Table 2.

Table 1: Obtained phase margin and its corresponding frequency

θ_{pss}	89.7	85	80	70	60
ω_{osc}	253	260	280	300	320

Table 2: Obtained parameters of CPSS and GAPSS

Parameter	T_{n1}	T_{d2}	T_{n3}	T_{d4}
CPSS	0.0035	0.0013	0.0035	0.0013
GAPSS	0.0333	0.0120	0.0333	0.0120

Table 3: The calculated Performance Index

Operating conditions	CPSS	CPSS4B	GAPSS	GAPSS4B
ITSE	6.945×10^{-4}	1.381×10^{-4}	4.083×10^{-4}	1.076×10^{-4}
ISE	8.198×10^{-4}	3.165×10^{-4}	5.736×10^{-4}	2.464×10^{-4}
ITAE	6.736×10^{-2}	1.305×10^{-2}	4.224×10^{-2}	1.171×10^{-2}
IAE	4.500×10^{-2}	1.910×10^{-2}	3.426×10^{-2}	1.697×10^{-2}

IX. RESULTS AND DISCUSSION

The response in Fig. 7 shows the result obtained when the system was simulated with AVR only. The result

show that AVR introduces low frequency oscillation when a step input of 0.1 was applied. But after the PSS is introduced the oscillation was damped as shown in fig. 8.

Figure 9 shows the simulation result when power is used as PSS input, and the rotor oscillations decayed slowly. But when speed is used as input, the oscillations is damped faster as shown in fig. 11, a significant improvement was experienced when both power and speed are used as input to the manually tuned PSS given in fig. 10. A look at fig. 12 reveals how fast the oscillation is damped, which is achieved by using the dual input PSS and tuned with GA.

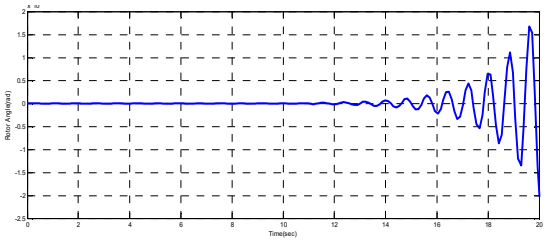


Figure 7 Rotor angle deviation (NO PSS)

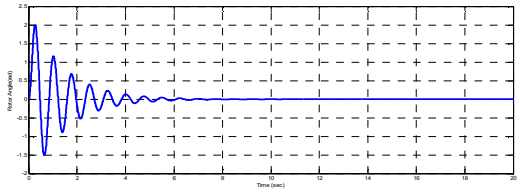


Figure 8 Rotor Speed (ω) as Input to (CPSS)

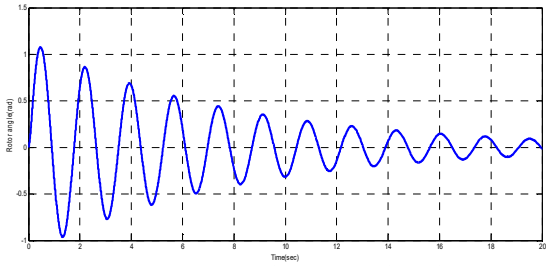


Figure 9 Power (P) as Input to (CPSS)

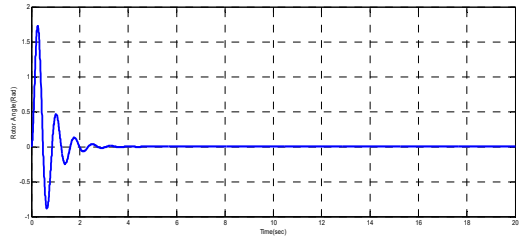


Figure 10 Dual Input P+ω (CPSS2B)

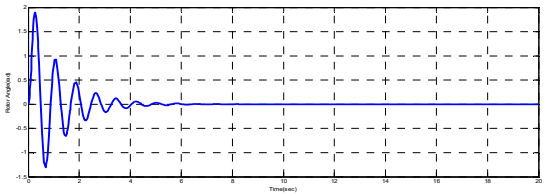


Figure 11 Rotor Speed (ω) as Input to (GAPSS)

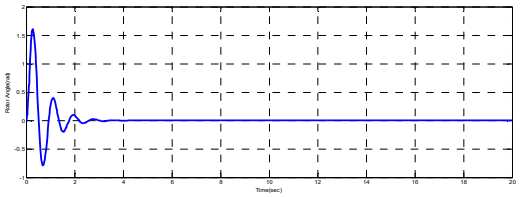


Figure 12 Dual Input P+ω (GAPSS2B)

The response in Figure 13 shows the result obtained when the system is simulated with single input PSS rotor speed feeds, for both Conventional and GA based PSS. The result shows that both the PSS were able to damped the oscillation as expected however, the GA based PSS performs much better as can be seen in the detail description of the response characteristic shown in Table 4 and table 3

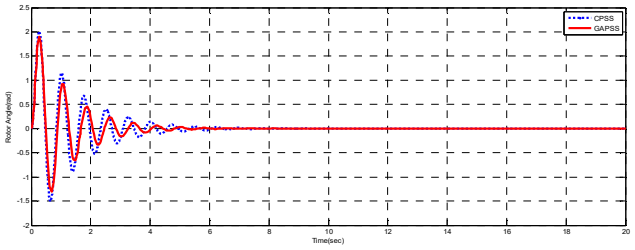


Fig. 13 Comparison between single input CPSS and Single input GAPSS

Table 4 Response with Angular Speed as the Input to PSS

PERFORMANCE INDEX	CPSS	GAPSS
SETTLING TIME	5.5718 sec	4.6733 sec
RISE TIME	0.0025 sec	0.0026 sec
OVERSHOOT%	151.3694	130.6646

The response in Figure 14 shows the result obtained when the system is simulated with GA optimized speed input PSS and GA optimized double input PSS, The result shows that both the two conditions are able to damped the oscillation as expected however, the Dual input GAPSS outperforms the Single input GAPSS, as can be seen in the detail description of the response characteristic shown in Table 5 and table 3.

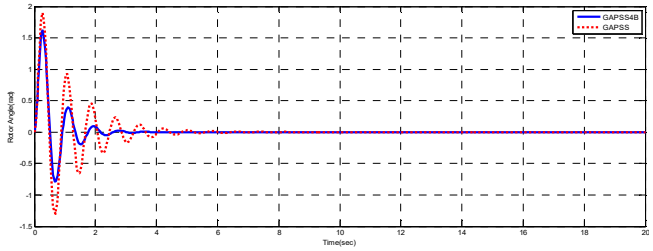


Figure 14 Comparison between dual and Single input GAPSS

Table 5 Comparison between GAPSS speed inputs and GAPSS4B

PERFORMANCE INDEX	GAPSS	GAPSS4B
SETTLING TIME	19.9656 sec	2.4677 sec
RISE TIME	16.7671 sec	0.0021 sec
OVERSHOOT	80.6543	79.3393

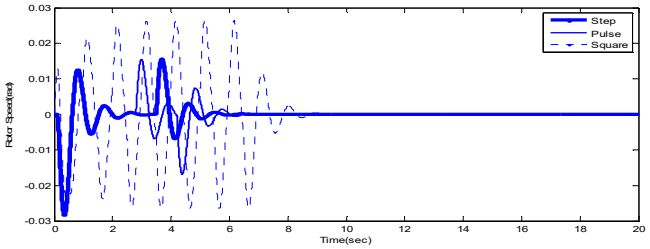


Fig. 15 GAPSS4B perturbed with Step, Pulse and Triangular Signals

Simulation results and the comparative analysis of the transient response of both GAPSS4B and GAPSS provide system model are given in fig. 15 and 16. To demonstrate the robustness of the proposed technique, the system is perturbed with different signals and in each case, the proposed technique was able to damp out the oscillations as given in figure 15. Figure 16 reveals that after the fault is introduced the PSS4B based result resumes from this disturbance with much lesser fluctuation in angular speed as compared to that of GAPSS based response. The result shows the robustness of the proposed technique GAPSS4B. Table 3 is the calculated performance index of four operating conditions where GAPSS4B has the smallest angular speed deviation in comparison to that of GAPSS, Thus, GAPSS4B based model shows improved response, having lesser amplitude of angular speed deviation under fault and damp out the system oscillation as seen in all the figures. Thus, GAPSS4B found to be much less affected to faults (robust) because PSS4B stabilize in all the state deviations to zero much faster than GAPSS as shown in Table 3.

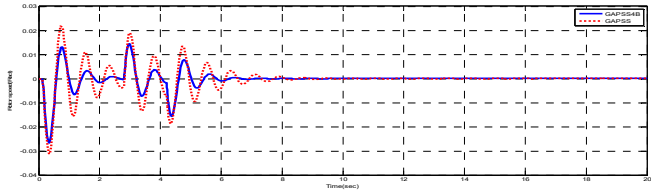


Figure 16. Comparison between transient responses of both the PSSs

X. CONCLUSION

A novel approach for designing Dual input Genetic Algorithm power system stabilizer for a Single Machine Infinite Bus system has been conferred. Genetic Algorithm has been explained as the optimization tool for tuning the parameter of the power system stabilizer, the GAPSS4B enhances both transient and dynamic stabilities. The dynamic performance of the propose technique GAPSS4B improved, both in terms of peak deviation and settling time, than the performance of the system obtained with GAPSS and CPSS. Also the performance of the system with GAPSS4B is quite robust over various operating point.

ACKNOWLEDGMENT

I wish to express my sincere gratitude to the Almighty Allah who is above everything (Alhamdulillah) for giving me the health, strength and ability to accomplish this research work. My thanks and sincere appreciation go to my supervisor and motivator Engr. Prof. Nuraddeen Magaji, Electrical Engineering Department, Bayero University, Kano

REFERENCES

[1] Magaji, N. and Bashir, U. (2015). "Application of Power System Stabilizer on Stability Studies." Applied Mechanics and Materials Trans Tech Publications, Switzerland Vol. 793, pp 149-153. doi:10.4028/www.scientific.net/AMM.793.149.

[2] S. Vijaykumar, B. Vijeta, P. Vishal, B. Jagadish, and J. Kittur, "Modeling of Synchronous generator connected to," pp. 56–63, 2016.

[3] M. G. Jolfaei, A. M. Sharaf, S. M. Shariatmadar, and M. B. Poudeh, "A hybrid PSS-SSSC GA-stabilization scheme for damping power system small signal oscillations," Int. J. Electr. Power Energy Syst., 2017.

[4] A. Andreoiu and K. Bhattachawa, "Robust tuning of power system stabilisers using a Lyapunov method based genetic algorithm." vol. 75, pp. 99–107, 2017"

[5] A. Kharrazi and A. Kharrazi, "Tuning of power system stabilizer in Single Machine Infinite Bus (SMIB) using genetic algorithm and Power Factory Modal Analysi," no. September, 2016.

[6] F. Rashidi and M. Rashidi, "*Tuning of Power System Stabilizers via Genetic Algorithm for Stabilization of Power Systems*," pp. 1210–1219, 2004.

[7] D. K. Sambariya, R. Gupta, and R. Prasad, "Design of optimal input–output scaling factors based fuzzy PSS using bat algorithm," Eng. Sci. Technol. an Int. J., 2016.

[8] S. Shakerinia, "The Power System Stabilizer (PSS) Types And Its Models," vol. 4, no. 3, pp. 154–159, 2016.

[9] G. Y. R. Vikhram and S. Latha, "Design of Power System Stabilizer for Power System Damping Improvement with Multiple Design Requirements," no. 5, pp. 43–48, 2012.

[10] R. Kamdar, M. Kumar, and G. Agnihotri, "Transient Stability Analysis And, Enhancement Of IEEE- 9 Bus System," vol. 3, no. 2, pp. 41–51, 2014.

L. H. Hassan, M. Moghavvemi, H. A. F. Almurib, and K. M. Muttaqi, "A coordinated design of PSSs and UPFC-based, stabilizer using genetic algorithm," IEEE Trans. Ind. Appl., 2014.

Proceedings of the 2nd International Conference of the IEEE Nigeria Computer Chapter:
IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

Machine Learning Classification Algorithms for Phishing Detection: A Comparative Appraisal and Analysis

Noah Ndakotsu Gana
Department of Cyber Security Science
Federal University of Technology
Minna, Nigeria
noah.gana@st.futminna.edu.ng

Shafi'i Muhammad Abdulhamid
Department of Cyber Security Science
Federal University of Technology
Minna, Nigeria
shafii.abdulhamid@futminna.edu.ng

***Abstract*—Exponential growth experienced in Internet usage has paved the way to exploit users of the Internet, a phishing attack is one of the means that can be used to obtained victim confidential details unwittingly across the Internet. A high false-positive rate and low accuracy have been a setback in phishing detection. In this research 17 different supervised learning techniques such as RandomForest, Systematically Developed Forest (SysFor), Spectral Areas and Ratios Classifier (SPAARC), Reduces Error Pruning Tree (RepTree), RandomTree, Logic Model Tree (LMT), Forest by Penalizing Attributes (ForestPA), JRip, PART, Nearest Neighbor with Generalization (NNge), One Rule (OneR), AdaBoostM1, RotationForest, LogitBoost, RseslibKnn, Library for Support Vector Machine (LibSVM), and BayesNet were employed to achieve the comparative analysis of machine classifier. The performance of the classifier algorithms was rated using Accuracy, Precision, Recall, F-Measure, Root Mean Squared Error, Receiver Operation Characteristics Area, Root Relative Squared Error False Positive Rate and True Positive Rate using WEKA data mining tool. The research revealed that quite several classifiers also exist which if properly explored will yield more accurate results for phishing detection. RandomForest was found to be an excellent classifier that gives the best accuracy of 0.9838 and a false positive rate of 0.017. The comparative analysis result indicates the achievement of low false-positive rate for phishing classification which suggests that anti-phishing application developer can implement the machine learning classification algorithm that was discovered to be the best in this study to enhance the feature of phishing attack detection and classification.**

***Index Terms*—Phishing Attack, Classification Algorithm, Accuracy, Performance metric, RandomForest.**

I. INTRODUCTION

This Phishing attack is propositional to the exponential growth experienced on the internet [1] and its usage, irrespective of the various detection approaches that have been modeled to checkmate phishing attack exploit. Phishing refers to the act of illegally exploiting the confidentiality of an entity, the breach of confidentiality includes but not limited to username, credit card details, password and personal profile of target victim unwittingly [2], [3]. Fig. 1 shows a typical phishing life circle. Phishing can also be likened to "fishing" in variation in respect to luring victim through the use of "bait" and "fishes" to extract victim privacy by an attacker [4].

Phishing attack has increased tremendously over the years across the platform of computing as about 1.3 million malicious phishing site was discovered in 2017 [5] ,these has led to a high false-positive rate in terms of phishing attack detection which is a major setback in phishing detection likewise the level of accuracy achieved from classification algorithms, while phishing was tagged as attack that has been occurring in a ridiculous pace with evolving techniques that aids in concealing its heinous act to be executed unknowingly on target victim [6], phishing attack on smart devices have recorded a hit of 36% in 2015 [3] and 91% of cyber exploit was said to have been initiated through phish attack [7]. Financial institutions, social media, webmail, healthcare, eCommerce, software as a service, cloud storage/file hosting, government services, retailers, insurance firm, social network, dating site, as well as technology users are the key target of phishing attack [6]. Billions of dollar are reported to have been lost by various agencies, firm, origination, and industries within a short period resulting in economic sabotage as well as the loss of integrity at varying levels [4] not spearing loss of intellectual property and compromise of national security [8]. Computer ethics come to play as it serves to analyze the nature and social impact of computer technology and the corresponding formulation and justification of policies for ethical used [9]. According to [10] phishing attack has led to the loss of $500 million annually in businesses in the US

alone, while phishing refers to as the most organized 21st-century crime.

Machine learning has been implemented in phishing website detection [11]–[13], supervised method which entails classification or prediction of problems in other to indicate the hidden association between the target class and independent variable are known to be employed in data mining [10]. For supervised learning, classifiers permit tagging of an attribute to observation, this is to ensure the classification of data not observed on the training data. A phishing detection system is developed with the implementation of classification algorithms to distinguish between legitimate and phishing attack [14], [15]. The adoption of a 10 folds cross-validation was as a result of its best gauge for error as obtained from a broad test on various machine learning algorithms [15]–[19]. A specific number of folds is selected, partitioning the data attribute into 10 part in which the class is represented in the approximately same proportion in regards to the complete dataset. Each partition is holdout in turn and learning scheme trained on the remaining nine-tenths, then the error rate is processed on the holdout set. Hence, the learning procedure is performed 10 times in various training set, at the end the average of the 10 error estimate are taken in other to obtain an overall error estimate.

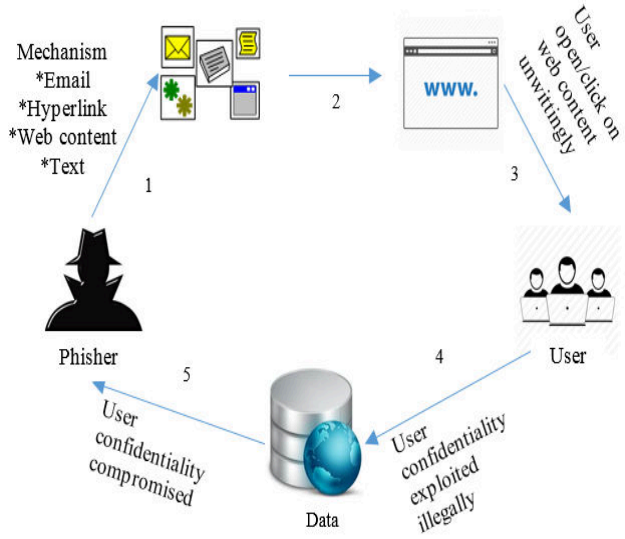


Figure 1. ***Phishing attack life circle***

Summary of the key contributions of this study are outlined below;

- We analyze the relevant classification algorithms used in related literature
- We identified and compared the standard parameters used for the evaluation of the machine learning algorithms
- We carried out a comparative analysis of machine learning classification algorithms

The research aims to comparatively analyze machine learning classification algorithms that can be employed for phishing attacks detection, employing the following machine learning classification algorithms RandomForest, Systematically Developed Forest (SysFor), Spectral Areas and Ratios Classifier (SPAARC), Reduces Error Pruning Tree (RepTree), RandomTree, Logical Model Tree (LMT), Forest by Penalizing Attributes (ForestPA), JRip, PART, Nearest Neighbor with Generation (NNge), One Rule (OneR), AdaBoostM1, RotationForest, LogitBoost, RseslibKnn, Library for Support Vector Machine (LibSVM) and BayesNet.

The remaining sections of the paper were arranged as follow Section II presents related literature in classification algorithms in the field of phishing detection. Section III reveals the methods employed in the research. Section IV presents the results obtained in the comparative analysis of machine learning classification algorithms and Section V shows the conclusion and recommendation.

II. RELATED LITERATURE

Due to the exponential rise of internet users, this has resulted in challenging issues as it relates to the mining of data and machine learning concerning a phishing attack. This has prompted for researches on comparative studies in terms of classification algorithms performance to accurately classify phishing using a combination of performance metrics [15]. It is, therefore, a matter of importance to determine algorithms that perform optimally for any chosen metrics to assist in proper classification of phishing and legitimate site. Support Vector Machine (SVM) [13], [20]–[23], K- Nearest Neighbors (KNN) [20], [21], Adaptive Neuro-Fuzzy Inference System (AFIS) [21], Decision Tree (DT) [20], [23]–[25], RandomForest (RF) [20], [24], [26], RotationForest (RoF) [13], [24], AdaBoost [20], Naïve Bayes (NB) [12], [13], [20], [23], [25], [27], Neural Network (NN) [13], [28], J48, Instance Based Learner K- Nearest Neighbors (IBK) and Reinforcement [13], Random Under-Sampling Boosting (RUSBoost), Gaussian Naïve Bayes and Perception [25], Sequential Minimum Optimization (SMO), Transductive Support Vector Machine (TSVM) and Phishing Hybrid Feature-Based Classifier (PHFBC) [23], Classifier and Regression Tree (CART), Credal Decision Tree (C-DT) and Gradient Boosted Machine (GBM) [24], Logistic Regression [20] the aforementioned classification algorithms in terms of their performance were employed in the reviewed related literatures. A summary of the classification algorithm used in the previous study is reflected in Table I

TABLE I. SUMMARY OF RELEVANT ALGORITHMS COMPARED IN RELATED RESEARCH WORKS

REFERENCE	ALGORITHMS																					
	SVM	K-NN	ANFIS	Decision Tree (DT)	Random forest (RF)	rotation forest (RoF)	AdaBoost	Naive Bayes	Neural Network	J48	IBK Lazy Classifier	Reinforcement	RUSBoost	Gaussian Naïve Bayes	Perception classifier	SMO	TSVM	PHFBC	Classification and Regression Tree(CART)	Credal Decision Tree (C-DT)	Gradient Boosted Machine (GBM)	Logistic Regression
[21]	√	√	√																			
[24]				√	√	√													√	√	√	
[20]	√	√		√	√		√	√														√
[12]								√														
[22]	√																					
[13]	√					√		√	√	√	√											
[28]									√			√										
[27]								√														
[23]	√			√				√								√	√	√				
[26]					√																	
[25]				√				√					√	√	√							

In a bid to protect online internet user from revealing sensitive information to malicious entity, [12] proffer a data mining method that will detect a phishing site and notify user about such malicious site, focus of the research was to be able to analyze on distinct email and offer a precautionary measure against phishing attacks, therefore a naïve based classifier algorithm was used to train classifier model, while Naïve Bayesian classification was used to determine between a real and fake emails. It was asserted that the proposed method accuracy outperformed previously known detection algorithms of phishing attacks.

Accuracy of 98.63%, True Negative Rate (TNR) of 98.19% and True Positive Rate (TPR) of 99.07% was achieved after an experiment was carried out on a well-known dataset with a False Positive Rate (FPR) of 1.81% and False Negative Rate (FNR) of 0.93% respectively in the research [28], a model that uses Neural Network(NN) algorithm enhanced with reinforcement learning to detect online email phishing was postulated, the reinforcement learning algorithm aids in self-remodeling of the phishing detection system in order to checkmate metamorphic nature of email phishing as well as enhance detection of unknown behavior of malicious traits which will lead to the fortification of self-dynamism of the system over time, it was stated the proposed model performed better compared to other existing methods as it possessed the capability of detecting zero-day phishing attack, however, the research fall short of attaining an accuracy of 100%. The research work by [27] used a Bayesian algorithm to enable the detection of email phishing as well as spam mail, anti-phishing simulator was developed with a major concentration on the evaluation of text content and the defining of messages that accommodates phishing attribute as well the application of cloud to offer a more accurate classification and results.

The quest for a productive detection system for fake website [22] proposed a statistical learning-based system for detection of fake websites, to this ends, AZprotect was developed which was incorporated alongside with SVM that serves as a classification algorithm, furthermore, an experiment was performed in four series in other to evaluate the efficacy of the following; the utilization of extended fraud cues, the comparison of the new system against the existing fake website detection systems, the evaluation of the new system algorithm as against other existing learning methods, and the performance rate as they relate to the new system. A total of 900 websites was used for the experiment on a testbed, out of which 200 authentic websites 350 fabricated websites and 350 spoofed websites. The following metrics were used in evaluating the performance; overall accuracy, classified precision, classified level recall, F-Measure, and ROC (Receiver Operating Characteristic. It was noted that the overall accuracy rate for concocted and spoofed site detection outperforms other existing fake website detection systems, same applies to F-Measure, recall, and precision as it was said to be close to 100% however not 100%, while the comparison against other learning classifiers of fake website detection as against the proposed SLT-based algorithm SVM was said to have outperformed Logistic Regression, Bayesian Network, Naïve Bayes, Neural Network and Winnow classifier in terms of overall accuracy, F-Measure, precision, recall and ROC all in respect to spoofing detection and concocted website detection, however, J48 was acclaim to be effective in

detection of concocted site as compared to SVM which have a 3% recall rate higher. While in the t-test comparison SVM result was stated to have significantly outperform other six learning algorithms in terms of overall accuracy and recall on detection of concocted and spoof websites, while it falls short in terms of recall when compared to Logistic Regression and Bayesian Network on concocted website detection, however, the use of multiple classifiers could have offer better result.

The research work by [26] proposed a combined method of Resource Description Framework (RDF) and Random Forest in other to identify phishing website, the posit method implores a training technique of supervised learning to train system after which a true positive rate of 98.8% was achieved, while further integration of RDF and ensemble learning method offer an accuracy rate of 98.68 and a false positive rate of 1.5% indicating that the proposed system generates a less false positive and a high accuracy rate. [13] suggest a Random Forest classification algorithm for the training of the proposed research model, as the research focus on URLs to detect phishing websites as well as its attributes. A cloud-based classification model was opined in other to enhance the detection of both old and newly generated phishing URLs, furthermore, publicly accessible dataset that comprises of 6157 legitimate websites and 4898 phishing websites culminating to 11055 records was used and then experiment on comparison of different classifier algorithms was carried out using WEKA, Random Forest recorded an accuracy of 97.259% make it the best compare to Naïve Bayes, J48, SVM, Neural Network and IBK Lazy classifiers, Random Forest was later preferred for the proposed model.

An integrated feature of images, frames, and text of phishing websites approach was used in presenting an intelligent phishing detection and protection, through the implementation of a develop AFIS algorithm. Further experiment was carried out on a phishing website dataset, while it was established that the presented approach renders efficient integrated solution to the detection of phishing websites and also protection against phishing websites through the achievement of accuracy rate of 98.3% based on the experiment performed, the research emphasis was cap on integration of hybrid feature extraction from text, frames, and images. The proposed research asserts that existing phishing detection methods do not have the algorithm to detect accurately new phishing attacks as a result of features such as text, images, and frames integration evaluation process. However, the developed algorithm has a challenge in term of identifying features using linguistic variable to establish intelligent model for the detection of phishing websites, the following classification algorithm was used to compare the accuracy rate; SVM, K-NN, and ANFIS, while precision, recall, accuracy, and F-Measure serve as the performance matrix with the following scores 98.31%, 98.26%, 98.3%, 98.28% respectively and an error rate of 1.7%, the developed model was said to be superb [21].

The research work by [25] Random Under-Sampling Boosting algorithms (RUSBoost) was used as a classifier in the proposed research that aims at addressing email phishing, SAFE-PC system was developed in other to aid detecting new phishing pranks, as it evolves from existing phishing techniques, furthermore, an evaluation was performed on real-world corpus, which aid in comparing SAFE-PC against Sophos an email protection software and Spam Assassin, the performance of SAFE-PC eluded and outperform both Sophos and Spam Assassin in term of email detection by 70%, Gaussian Naïve Bayesian, Decision Tree, and Perceptions classifiers were implored to complement RUSBoost, while RUSBoost alongside Gaussian Naïve Bayes learner resulted in a better performance achieving overall accuracy of 97%.

The performance as obtained after the analysis carried out on the proposed classifier Phishing Hybrid Feature-Based (PHFB) by [23] justify PHFB classifier as excellent in comparison to SMO, SVM, TSVM, NB and DT classification algorithms, 97%, 0.7% , 0% and 98.07% representing TPR, FPR, FNR and AUC respectively was achieved. The proposed classifier encompasses a hybrid feature of both Naïve Base and Decision Tree algorithm classifiers alongside a Recursive Feature Subset Selection Algorithm (RFSSA) that supports the distinction of phishing explicitly. Contracting misclassification of phishing algorithm in real-time as well as achieving a strong classification accuracy was the basis of this research and likewise detection of new phishes, however, the accuracy of 100% was not achieved.

The comparison of different ensemble approaches was performed based on the use of Random Forest, Rotation Forest, Gradient Boosted Machine (GBM) and Extreme Gradient Boosting which are known to be ensemble algorithms against single classifiers namely Decision Tree, Classification and Regression Tree and Credal Decision Tree as it relates to website phishing detection, the deploy performance matric in respect to different resampling approaches was Area Under ROC Cure (AUC), while to achieve baseline indication as it relates to evaluation of the importance of the classifier, a statistical test was used which further revealed that the research added benchmarking classifier of ensemble to the existing model with respect to website phishing detection, Random Forest was said to have outperformed other classifiers algorithms, however, in other to achieve a standard bench-marking other web phishing dataset are to be put into [20], [24] The representation of numeric phishing mails was achieved through the implementation of distributive representation of phishing mails known as Term Frequency Inverse Document Frequency (TF-IDF), furthermore, Random Forest, AdaBoost, Naïve Bayes, Decision Tree, SVM, KNN and Logistic Regression machine learning classification algorithm techniques was use in other to achieve a comparative study of the classifier algorithms, performance metrics such as accuracy, precision, recall and F1- Score was used in this research, while Non-Negative Matrix Factorization (NMF) and Simple Value Decomposition

(SVD) serves as a method for feature extraction and dimensional reduction respectively. Furthermore, the classification of email in terms of legitimate or phishing was achieved through the implementation of machine learning method on TFIDF + SVD as well as TFIDF + NMF representations, the highest accuracy was attained using Decision Tree and Random Forest classification algorithms, however, the imbalance structure of dataset affected the accuracy of the proposed methodology.

A combination of some performance metrics was employed in the reviewed of related literature such as AUC [23], [24], Accuracy [13], [20]–[22], [25], [26], [28], Precision [20]–[22], [28], F1-Score [24],Recall [20], [21], [28], F-Measure [21], [22], [28], FPR and FNR [12], [23], [25], [26], [28], ROC [22], TPR [23], [25], [26], [28], TNR [26], [28]. Table II shows performance metrics employed by previous research works.

TABLE II. SUMMARY OF RELEVANT PERFORMANCE METRICS USED FOR COMPARISON IN RELATED RESEARCH WORK

Reference	Performance Metric										
	AUC	Accuracy	Precision	F1-Score	Recall	F-Measure	FPR	FNR	ROC	TPR	TNR
[24]	√										
[20]		√	√	√	√						
[12]							√	√			
[22]		√	√			√			√		
[13]		√									
[28]		√	√		√	√	√	√		√	√
[21]		√	√		√	√					
[25]		√					√	√		√	
[23]	√						√	√		√	
[26]		√					√	√		√	√

III. METHODOLOGY

In this research, publicly available phishing attack dataset obtained from UCI machine learning repository [29] comprising of 11055 instances with 31 attributes of different variable type text, image and frames available in .arff format (a machine learning compactable format) used in WEKA to serve as input data for further analysis was employed which has been preprocessed and void of null value, machine learning classifier was employed, thereafter, comparative evaluation of machine learning classifier was carried out.

To remarkably classify phishing website dataset, RandomForest, SysFor, SPAARC, RepTree, RandomTree, LMT, ForestPA, JRip, PART, NNge, OneR, AdaBoostM1, RotationForest, LogitBoost, RseslibKnn, LibSVM, and BayesNet were employed, a 10folds cross-validation and a 70% split (70% for training and 30% for testing) were used as a test option.

IV. RESULT

A 10 fold cross-validation and 70% split test option was applied on the machine learning classification algorithms in the experiment using the complete dataset. Accuracy, Precision, Recall, F-Measure, Receiver Operator Characteristic (ROC) Area, Kappa Statistics, Root Mean Squared Error (RMSE), True Positive (TP), False Positive (FP) and Root Relative Squared Error (RRSE) were used in order to determine the effectiveness, validate and relay key information of the research [30], [31], the comparison of classification algorithms is summarized below.

A. Accuracy

Accuracy indicates the prediction level correctness, the value 1 indicates the highest accuracy rate, in this research, the highest accuracy of 0.9838 was obtained when 10 folds cross-validation was employed on RandomForest algorithm and the lowest accuracy of 0.892 was obtained from 10 folds cross-validation of OneR, little variation is observed compared to 70% split test option as 90% of classification algorithms scored well above 90% accuracy. Fig. 2, indicates the accuracy.

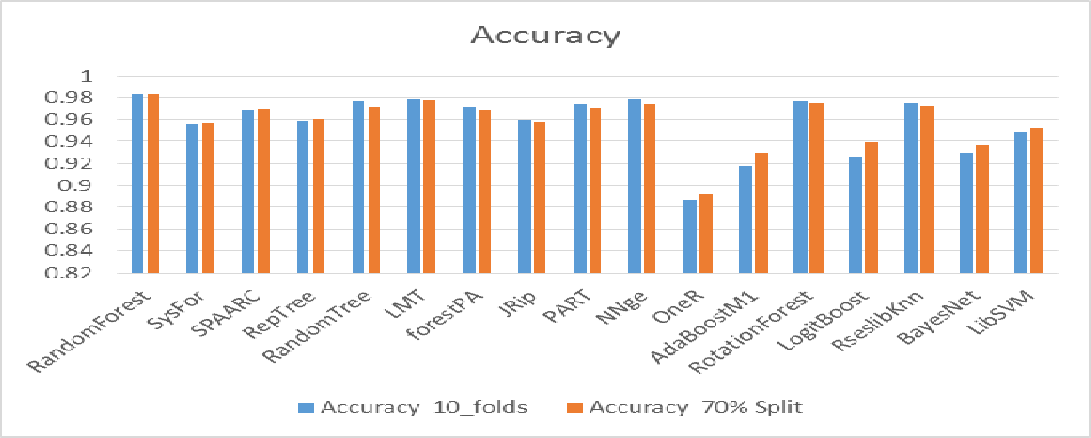


Figure 2. ***Comparison of accuracy***

B. Precision, Recall and F-Measure

Precision indicates the fraction of important recollected instances, RandomForest with a precision of 0.984 under 10 fold cross-validation outperformed other classification algorithms with little variation compared to the other test option, Fig. 3, Recall represents the relevant instances which are recalled RandomForest attained the best recall rate of 0.984 under 10 fold cross-validation test option outperforming other classification algorithms Fig. 4, while F-Measure is used in this research, a high F-Measure is needed since both precision and recall are needed to achieve a high score and RandomForest has the highest F-Measure of 0.984 as presented in Fig. 5, through the application of a 10 fold cross-validation test option.

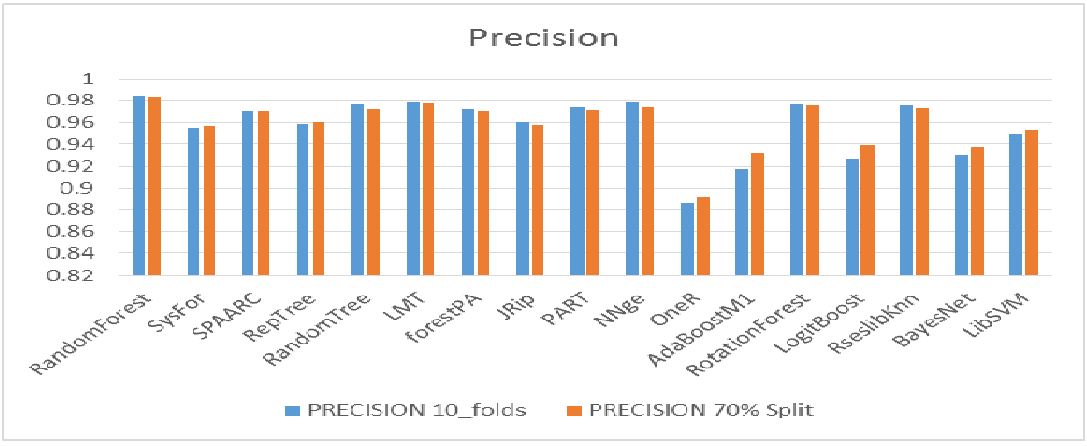


Figure 3. ***Comparison of precision***

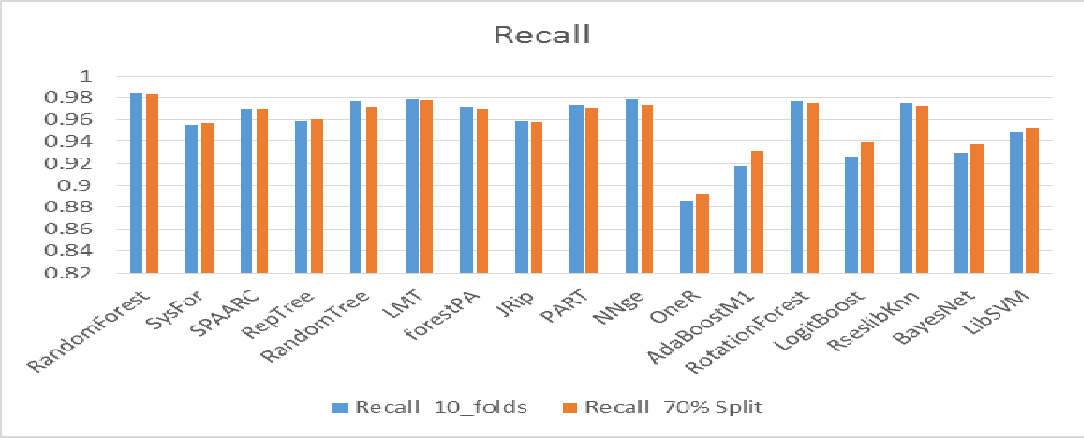


Figure 4. ***Comparison of recall***

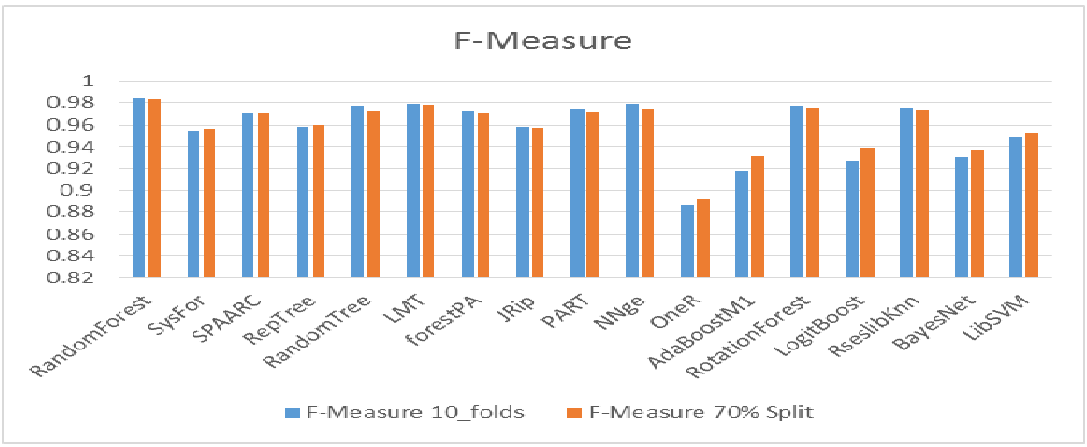


Figure 5. ***Comparison of the f-measure***

C. ROC Area

The ROC (AUC) Area of a classifier/algorithm is the probability of the classifier ranking a random selected positive instance higher than a randomly selected negative instance Fig. 6, represents the area under ROC curves of classifier used in this research with RandomForest having the highest of 0.997 same for test options 10 fold cross-validation and 70% split and OneR having the lowest with 0.884 with a 0.005 variation between the test options.

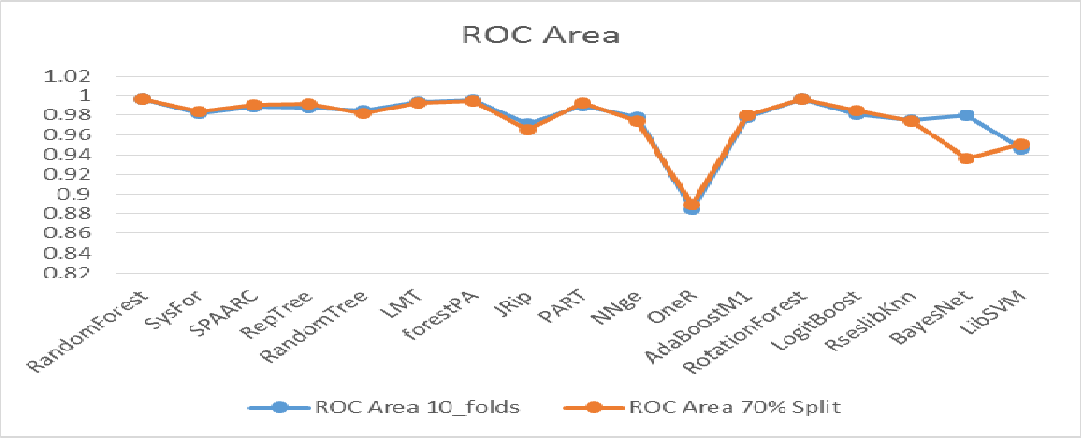


Figure 6. ***Comparison of roc area***

D. Kappa Statistics

The kappa characteristic represents the level of agreements between the true classes and the classifications. The highest value represented as 1 indicate total agreement, in this research, the highest kappa characteristic is 0.9672 obtained from the test performed on RandomForest with 10 folds cross-validation, Fig. 7, indicates the respective kappa characteristic.

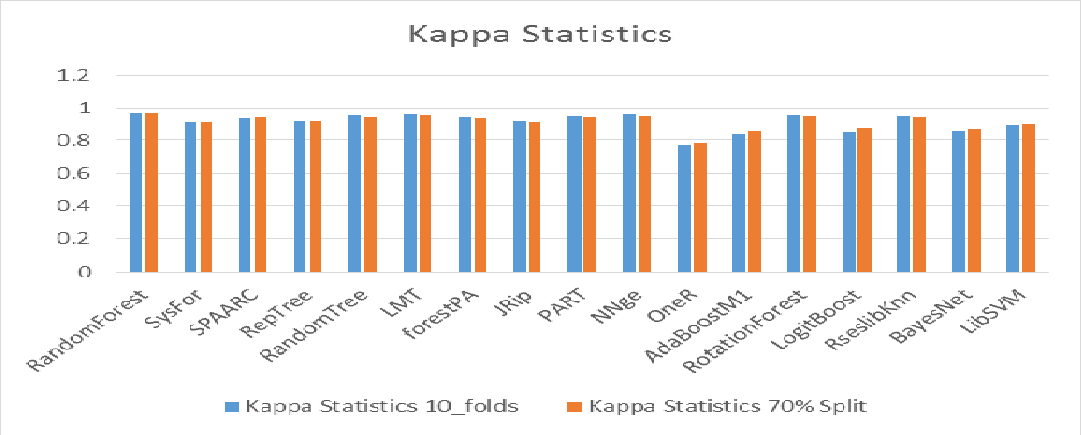


Figure 7. ***Kappa statistic***

E. Root Mean Squared Error

A low value is an indication of an excellent classifier in respect to the root mean squared error, a low value for the root mean square error was recorded for RandomForest using 10 folds cross-validation with 0.117, Fig. 8 indicates the RMSE.

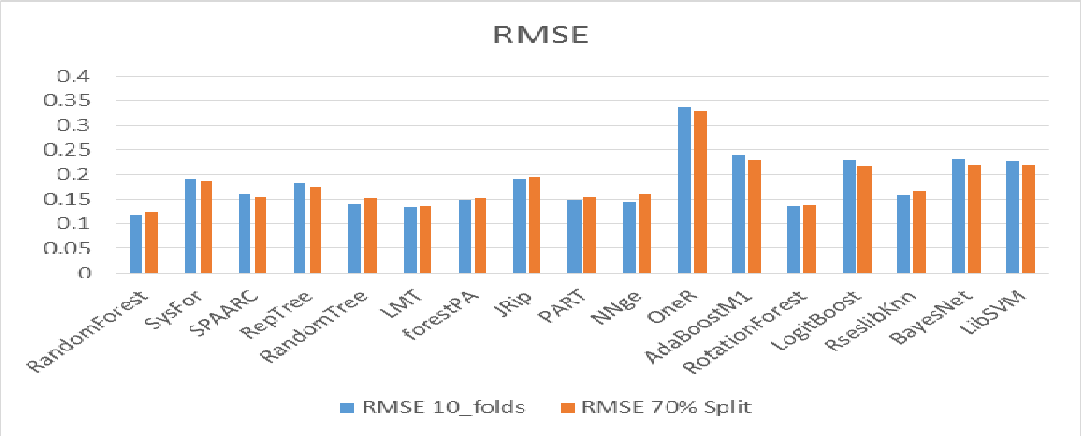


Figure 8. ***Comparison of Root Mean Squared Error***

F. Root Relative Squared Error, False Positive and True Positive

The relative squared error normalizes the total squared error by dividing it by the total squared error of the simple predictor. The error is reduced to the same dimension as the quality being predicted by taking the square root of the relative squared error, RMSE of 0.117 was achieved from the test option of 10fold cross-validation. Fig. 9, indicates the values of the RRSE of 0.2356 obtained from 10 fold cross-validation test option. The False Positive (FP) is the number of legitimate sites incorrectly classified as phishing while True Positive (TP) is the number of phishing emails correctly classified as phishing, both FP and TP achieved 0.017 and 0.984 respectively as indicated in Fig 10, Fig. 11, while a variation of 0.001 was discovered in true positive as regards to test options employed, there was no variation for false-positive rate.

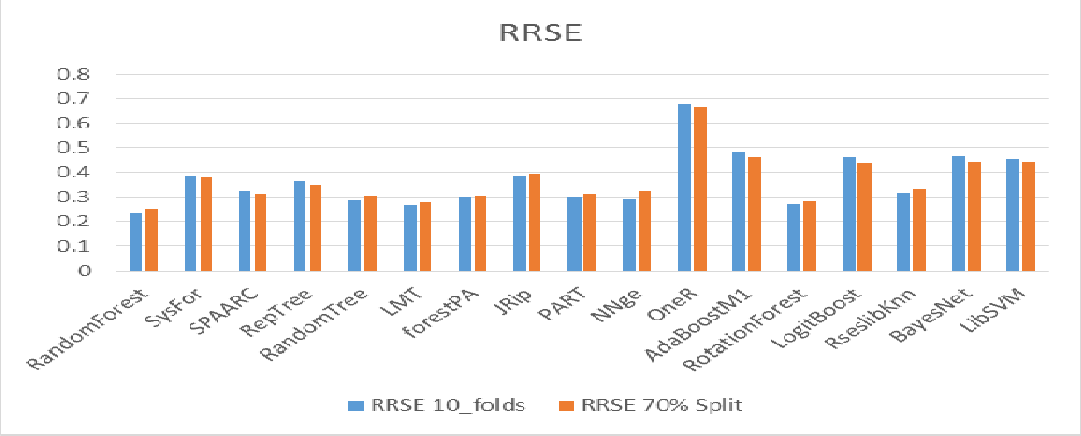


Figure 9. ***Root Relative Square Error***

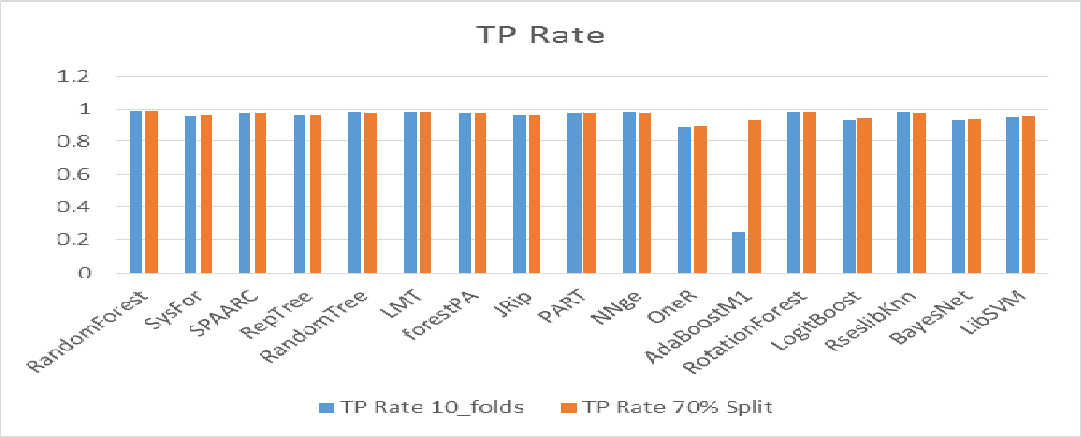


Figure 10. ***True Positive Rate***

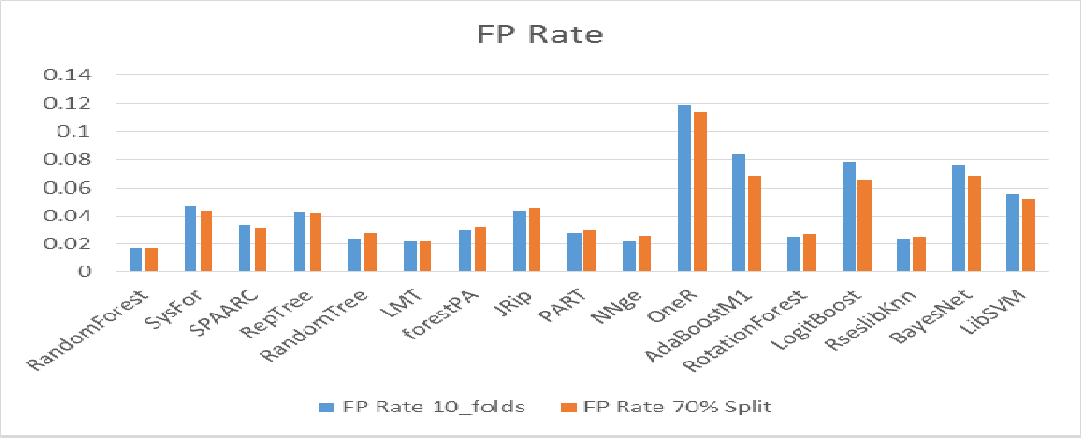


Figure 11. ***False Positive Rate***

V. CONCLUSION AND RECOMMENDATION

This research that was simulated as a result of the exponential rate of phishing attack across Internet platform and also as a result of knowledge derived from the reviews of literature available on classification algorithms that are yet to be compared in terms of their performance metric on phishing attack datasets.

Some classification algorithms performed optimally out of the 17 different machine learning classification algorithms used in performing experiment on phishing dataset involving two test options, RandomForest outperform other machine learning classification algorithms proving its efficiency in performance based on the fact that RandomForest is distinguished with low bias, reduced variance, and overfitting, meanwhile RandomForest attained an excellent classification through developing multiple decision tree at the training period and obtains a mode [32]. The indication from the result obtained from the experiment reveals that RandomForest with a false positive rate of 0.017 and 0.9838 accuracy was excellent for phishing detection classification, which is effective than other commonly recognized classification algorithms such as RandomTree, AdaBoostM1, and RotationForest with 0.9771, 0.9181, and 0.9769 accuracies respectively. An anti-phishing application developer can implement the machine learning classification algorithm that was discovered to be the best in this study to enhance the feature of phishing attach detection and classification, thereby reducing high false-positive rate associated with some methodology employed in phishing attack detection furthermore, helping in filtering phishing attack.

The author recommends further experiment be performed on more phishing attack dataset from different sources, while

future research should investigate deep learning techniques compared with the method described in this study.

REFERENCES

[1] N. Ashraf, W. Ahmad, and R. Ashraf, "A Comparative Study of Data Mining Algorithms for High Detection Rate in Intrusion Detection System," *Ann. Emerg. Technol. Comput.*, vol. 2, no. 1, pp. 49–57, 2018.

[2] P. Pella, T. S. Latinovi, and D. M. Preradovi, "A Review on Cybersecurity Threats and Statistical Models A Review on Cybersecurity Threats and Statistical Models," 2018.

[3] N. Zlatanov, "Computer Security and Mobile Security Challenges," 2015. [Online]. Available: https://www.researchgate.net/publication/298807979.

[4] K. L. Chiew, K. Sheng, C. Yong, and C. L. Tan, "PT," *Expert Syst. Appl.*, 2018.

[5] "Phishing Trends and Intelligence (PTI) Report," 2018.

[6] "Phishing Activity Trends Report 3 Quarter," 2018. [Online]. Available: http://www.apwg.org.

[7] B. J. Hope and D. Ph, "Report provides insights on cybersecurity," 2019.

[8] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: a survey," *Comput. Secur.*, 2017.

[9] S. O. Ogunlere and A. O. Adebayo, "Ethical Issues in Computing Sciences," *Int. Res. J. Eng. Technol.*, vol. 2, no. 7, pp. 10–16, 2015.

[10] R. Wang, W. Ji, M. Liu, X. Wang, J. Weng, and S. Deng, "Review on mining data from multiple data sources," *Pattern Recognit. Lett.*, pp. 1–9, 2018.

[11] Z. Hu, C. Raymond, P. Ilung, B. Yukun, and L. Yuqing, "Malicious web domain identification using online credibility and performance data by considering the class imbalance issue," *Ind. Manag. Data Syst. Artic. Inf.*, 2018.

[12] P. K. Sahoo, "Data mining a way to solve Phishing Attacks," *2018 Int. Conf. Curr. Trends Towar. Converging Technol.*, pp. 1–5, 2018.

[13] M. Thaker, M. Parikh, P. Shetty, V. Neogi, and S. Jaswal, "Detecting Phishing Websites using Data Mining," *2018 Second Int. Conf. Electron. Commun. Aerosp. Technol.*, no. Iceca, pp. 1876–1879, 2018.

[14] R. Islam and J. Abawajy, "A multi-tier phishing detection and filtering approach," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 324–335, 2013.

[15] S. Abu-nimeh, D. Nappa, X. Wang, and S. Nair, "A Comparison of Machine Learning Techniques for Phishing Detection," in *In Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit (pp. 60-69). ACM.*, 2007, pp. 60–69.

[16] I. H. Witten and E. Frank, *Data Mining Practical Machine Learning Tools and Techniques*, Second Edi. 2005.

[17] G. Varoquaux, "Cross-validation failure: small sample sizes lead to large error bars," *Neuroimage*, 2017.

[18] L. E. O. Breiman, "Random Forests," *Mach. Learn. 45(1), 5-32.*, pp. 5–32, 2001.

[19] S. Varma and R. Simon, "BMC Bioinformatics," *BMC Bioinformatics*, vol. 8, pp. 1–8, 2006.

[20] A. V. Harikrishnan NB, Vinayakumar R, Soman KPVidyapeetham, "A Machine Learning approach towards Phishing Email Detection," 2018.

[21] M. A. Adebowale, K. T. Lwin, E. Sánchez, and M. A. Hossain, "Intelligent web-phishing detection and protection scheme using integrated features of Images , frames and text," *Expert Syst. Appl.*, vol. 115, no. December 2017, pp. 300–313, 2018.

[22] A. Abbasi, Z. Zhang, D. Zimbra, H. Chen, and J. F. Nunamaker, "Detecting Fake Websites: The Contribution of Statistical Learning Theory," *Manag. Inf. Syst. Res. Center, Univ. Minnesota*, vol. 34, no. 3, pp. 435–461, 2010.

[23] H. Zuhair and A. Selamat, *Phishing Hybrid Feature-Based Classi fi er by Using Recursive Features Subset Selection and Machine Learning Algorithms*, vol. 1. Springer International Publishing, 2018.

[24] B. A. Tama and K. H. Rhee, "A Comparative Study of Phishing Websites Classification Based on Classifier Ensembles," no. May, 2018.

[25] C. N. Gutierrez *et al.*, "Learning from the Ones that Got Away : Detecting New Forms of Phishing Attacks," *IEEE Trans. Dependable Secur. Comput.*, vol. PP, no. c, p. 1, 2018.

[26] V. Muppavarapu, A. Rajendran, and S. Vasudevan, "Phishing Detection using RDF and Random Forests," *Int. Arab J. Inf. Technol.*, vol. 15, no. 5, pp. 817–824, 2018.

[27] M. Baykara, "Detection of phishing attacks," *IEEE*, 2018.

[28] S. Smadi, N. Aslam, and L. Zhang, "Detection of Online Phishing Email using Dynamic Evolving Neural Network Based on Reinforcement Learning," *Decis. Support Syst.*, 2018.

[29] "Phishing Website," *Phishing Website Dataset: UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/machine-learningdatabases/ 00327/Training%20Dataset.arff (Accessed: 20 February2019). .*

[30] T. M. Ibrahim *et al.*, "Recent Advances in Mobile Touch Screen Security Authentication Methods : A," *Comput. Secur.*, no. April, 2019.

[31] M. Shafie *et al.*, "A Review on Mobile SMS Spam Filtering Techniques," *IEEE Access*, vol. 5, pp. 2169–3536, 2017.

[32] U. R. Hodeghatta and U. Nayak, *Supervised Machine Learning — Classification*. Apress, Berkeley, CA, 2017.

Development of a True Random Number Generator based on Computer Mouse Wait time and Datetime

Idakwo M.A
Dept.of Computer Engineering, Ahmadu Bello University Zaria.
mondabutu@gmail.com

Mu'azu M.B
Dept. of Computer Engineering, Ahmadu Bello University Zaria.
muazumb1@yahoo.com

Adedokun E.A
Dept. of Computer Engineering, Ahmadu Bello University Zaria.
adewaleadedokun@yahoo.co.uk

Sha'aban Y.A
Dept. of Computer Engineering, Ahmadu Bello University Zaria.
shaabanyusuf@yahoo.com

Abstract— Random number generators are pivotal in information theory especially in cryptography (stochastic, mathematical and quantum), financial institution, lottery, numerical simulations, Monte Carlo calculations, statistical research, randomized algorithms, amongst others. For a random number to be unpredictable, the environment producing such sequences must not be easily subjected to manipulation or reproduction. Therefore, this paper developed a true random number generator that is low cost, universal and convenient using the non-deterministic computer mouse wait time and system datetime. The generated sequences passes all the 15 NIST SP 800-22 test suites with P-value > 0.0001 which is within the expected proportion rate of 0.98056 and also with an entropy of 0.999984. Furthermore, the proposed generator p-value were compared with other existing generators in order to evaluate the generator statistical tests. The result obtained clearly shows that the proposed generator outperform existing systems and as such can be adopted as a true random number sources.

***Keywords*—** entropy source, mouse wait time, seed, system datetime, true random number generator, pseudo random number generator

I. INTRODUCTION

Random numbers are basically generated using random number generator (RGN) and forms a pivotal components in various applications including scientific simulations, cryptography, recreational entertainment, industrial testing amongst others [1]. This generated numbers can be shared and used in securing communication. Hence two or more systems can generate the same sequences of non-repeating numbers known as key for the purpose of communicating securely. True random number generators (TRNGs) [2] and pseudo-random number generators (PRNGs) [3] are basically the two types of random number generators that exist. PRNGs are based on assumptions of some certain mathematical difficulties such as linear feedback shift register [4], discrete logarithm problem [5], non-linear congruencies [6], quadratic residuosity problem [7], cellular automata [8], amongst others. PSNGs are popularly used in application with lower security demands owing to high flexibility, lower cost, and faster number generating time with their uncertainty dependent on the secrecy of their seed [3]. A seed is a number (or vector) use for initializing PRNG. Once the seed is known every other keys becomes predictable [9]. Contrarily, a TRNGs mechanism relies on sampling entropy inherent in non-deterministic physical phenomena and guarantees unpredictable numbers even with full knowledge of the underlying mechanism [2]. TRNGs are sub divided into physical and nonphysical true random number generators. The physical true random number generators are implemented in hardware based on non-deterministic effects of electronic circuits such as thermal noise, shot noise, ring oscillators or physical experiments like quantum random processes [11]. Nevertheless, there are practically more expensive to implement due to additional hardware requirements, subjected to malfunction, slower due to harvesting of entropy sources and the solutions are hardware dependent [9]. Several researchers have utilized the nonphysical true random number generator by adopting the computer system [10]. Xingyuan *et al.* [24] proposed and improvement of Hu *et al*. [15] mouse movement generator by evaluating the x and y-coordinates of the computer mouse movement as iteration. Chaotic map was adopted to post process the iterated numbers in order to generate a true random sequences. Similarly, Zhu *et al.* [25] adopted the uniqueness of human iris as a source of entropy. The computer system was used to capture the iris image while a canny edge detection was adopted to extracts the edges of the captured image to form a binary image of the iris information which were post process using chaotic function. Also Teh *et al.* [26] utilized the graphics processing units simultaneous access of a memory location and chaotic maps in generating an unpredictable random numbers. Equally, Kiamehr *et al.* [28] implemented a random number generator using the static random access memory cells start-up value while Kim *et al.* [1] utilized the dynamic random access memory cells. Equally, the noise from flash memory has been exploited by Ray and Milenković [11], likewise the android mobile device camera which was equally implemented as entropy source by Yeoh *et al.* [27]. These systems created the possibilities of exploring the computer system as an entropy

sources in generating random numbers which are non-deterministic, more secured, cheaper and convenient than using hardware true random number generator. Nonetheless, the entropy sources which are solely from the computer random processes are extracted and converted into an image format before post processing in order to eliminate similar patterns. This increases the computational requirement as a result of feature extraction and post processing phases therefore making it slower for practical usage as evident by the researchers. While entropy sources from peripheral devices can be subjected to either software or hardware malfunctioning. This infer that the entropy sources are dependent on the user's environment, acquisition sources and methods. This paper developed a true random number generator based on computer mouse wait time and system clock datetime that is cheaper, convenient and universal for all personal computer. The generated random numbers obtained from the generator were subjected to the 15 national institute of standards and technology special publication (NIST SP) 800-22 test suite [11] and passed all the statistical tests. The contribution of this paper are highlighted as follows:

- The mouse wait time and system clock datetime generated a high entropy source for the TRNG
- The developed TRNG can be used in all personal PC as it is flexible, cheaper and required no external circuit, convenient and generate random numbers in shorter time.
- The TRNG passed all the NIST SP 800-22 test and thus can be adopted to the numbers can be used as secret keys for secure communication.

II. BACKGROUD OF THE STUDY

A random number generator is an algorithm or a device that produces an unbiased sequential and statistically independent number. The random number are not only limited to cryptography but likewise essential in lottery [20], numerical simulations [22], Monte Carlo calculations [21], statistical research [23], randomized algorithms [24], amongst others [12]. The determinant of a number's randomness is known as uncertainty or entropy and it is mathematically given by [13]

$$H(y) = -\sum_{j=1}^{n} p_{(\sigma j)} \log_2 p_{(\sigma j)} \quad (1)$$

Where; p is the probability of the value numbered j, and $1 \le j \le n, \sigma_j = \{a_0, ..., a_{j-1}\}, a \in \{0,1\}$

True random number generators and pseudo-random number generators are basically the two types of random number generators that exist.

A. True Random Number Generator

The true random number generators uses unpredictable physical sources or natural phenomenon like atmospheric noise, thermal noise, radioactive decay, amongst others for generating random numbers. There are hardware based and constructed using entropy source, harvesting techniques, and post-processing with architectural structure as shown in Fig. 1.

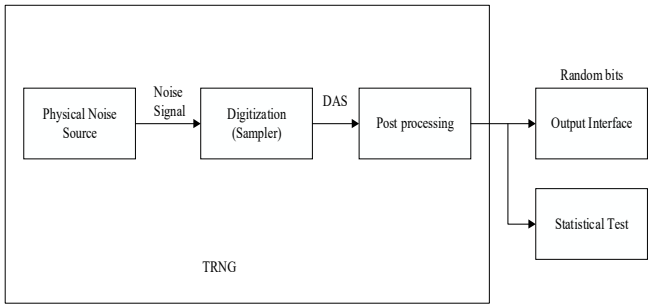


Fig. 1: True Random Number Generator Architecture

The determination of available entropy and its sources of randomness is the sole responsibility of the entropy source. The process of collecting nondeterministic data is refer to as entropy gathering. While the harvesting techniques helps in sampling the entropy source to gather as much entropy as possible without disrupting the physical process [13]. The post-processing is optional in designing TRNG but always adopted in practice for strengthening the robustness and elimination of bias in the generated random bits [12]. The TRNG are non-deterministic in nature and thus have the advantage of producing a higher entropy that guarantees higher security, but nonetheless they are practically more expensive to implement due to additional hardware requirements, subjected to malfunction, slower due to harvesting of entropy sources and the solutions are hardware dependent [9]. In a computer system, some of the standard entropy sources available to the operating system includes mouse motion, keyboard strokes, system clock, sound card data, disk access times, interrupt timing , buffers input/output contents, user input, system load and network statistics [16]. Nonetheless, developing a software program to exploit this randomness in order to generate a bit sequence that is free of biases and correlations is a difficult task. Any random number generators exploiting the computer entropy sources are referred to as non-physical non-deterministic [9].

B. Pseudo random number generator

Pseudorandom numbers generator are deterministic in nature and are solely dependent on mathematical difficulties assumptions such as linear feedback shift register, discrete logarithm problem, non-linear congruencies, quadratic residuality problem, cellular automata, amongst others. All PSNGs uses seed as initializing vector to produce a pseudo-

random bit sequence as output during computation. The seed is the only source of uncertainty and this makes periodicity a common property of pseudorandom number generators [9]. Hence, a pseudo random number generator is cryptographically secured if there is no known polynomial-time algorithm which can use the first bit of the sequence as an input to predict the next, previous state bit of the sequence with probability significantly greater than 1/2. [1]. The architecture of pseudo random number generator is as shown in Fig. 2.

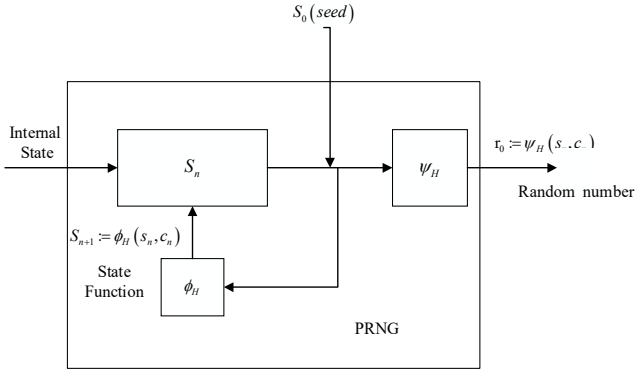


Fig. 2: Pseudo Random Number Generator Architecture

The major advantage of PRNGs are their rapidity in generating the random sequences using lesser memory for algorithm storage, cheaper as they require no hardware, and can easily be implemented [13]. For a number to be regarded as been true random, it must satisfy the following requirement [14].

i. The random numbers generated must not have any statistical weakness
ii. An attacker who knows the sub-generators of random numbers, must not be able calculate or predict predecessors and successor
iii. There must be no possibility of predicting or calculating previously generated random numbers with high accuracy using the known current internal state value of the random number generator.
iv. There must be no possibility of predicting or calculating subsequent random numbers with high accuracy without requiring its internal state information.

The differences between PRNG and TRNG is as shown in Table 1.

Table 1: Comparison of PRNG and TRNG

	PRNG	**TRNG**
Type of Realization	software or hardware	Hardware
Periodicity	Periodic	Aperiodic
Application design cycle	Easy because of standard structures	Complex
Efficiency	Perfect	Weak
Numbers nature	Deterministic	nondeterministic
Change of theoretical calculation limit	Time dependency	Constant (independent of time)

C. *Mouse and System Clock*

The computer mouse is an input device used in controlling mouse movement on the computer screen [15]. Its movement on the screen is a physical motion that can traced and measured and treated as analog signals while computer programs can only process the digital signals. Interestingly with a call of its application programming interfaces a serial of the movement coordinates are returned from the operation system [16]. System clock is an electronic device in a computer system that issues a steady high-frequency signal used in synchronizing all the internal components. The system clock measures system time. The system time is a simple count of the number of tick that have occurred since some arbitrary starting date, referred to as an epoch. System time is more suitable for human understanding once converted into a calendar time [17].

D. *Random Number Statistical Test*

The NIST SP 800-22 test suite which is a statistical package comprising of 15 tests is generally adopted to ascertain the randomness of (arbitrarily long) binary sequences produced by either software or hardware based random number generators. The 15 tests are [18]:

a) The Frequency (Monobit) Test: This test is used in determining whether the number of ones and zeroes in a sequence are almost the same as expected for a truly random sequence and given mathematically as:

$$P - value = erfc\left(\frac{|S_n|\sqrt{2}}{\sqrt{n}}\right) \quad (2)$$

Where n is the bit string length and $n \geq 100$, S_n is the sum of sequence bits value, *erfc* is the complimentary function. If the calculated *P-value* is < 0.01, then the sequence is concluded as non-random else it is random. All subsequent tests rely on passing this test.

b) Block Frequency Test: This test ascertain the occurrences of ones in an $M-bit$ block that is approximately $M/2$, and it is represented mathematically as

$$P - value = igamc\left(\frac{N}{2}, 4M\sum_{i=1}^{N}\left(\pi - \frac{1}{2}\right)^2\right) \quad (3)$$

Where M is the block length, π is the proportion of ones in every M-bit block and defined

as $\pi_i = \frac{\sum_{j=1}^{M} \varepsilon\,(i-1)M + j}{M}$, N is non over lapping bit and computed as $N = \left\lfloor \frac{n}{M} \right\rfloor$, ε is the generated sequence bits and $igamc$ is the incomplete gamma function

c) The Runs Test: This sum up the number of runs in the sequence and determines if the oscillation between zeroes and ones are too slow or fast. A run is an identical bits with an uninterrupted sequence. The run test is mathematically expressed as

$$P-value = erfc\left(\frac{\left|V_n(ob) - 2n\pi(1-\pi)\right|}{2\sqrt{2n\pi(1-\pi)}}\right) \quad (4)$$

Where $V_n(ob)$ is the number of observed runs in a sequence of length n and it is defined as $V_n(ob) = \sum_{k=1}^{n-1} r(k) + 1$ and $\pi = \frac{\sum j\varepsilon j}{n}$

d) Tests for a Block Longest-Run-of-Ones: The reason for this test is to ascertain the consistency of ones. Any irregularity in the longest run length of ones also indicates an irregularity in zeroes length longest run. So, only a test for ones is required and the test is mathematically given as follows

$$P-value = igamc\left(\frac{k}{2}, \frac{\chi^2(ob)}{2}\right) \quad (5)$$

Where $\chi^2(ob) = \sum_{i=0}^{k} (v_i - N\pi_i)^2 / \mathrm{N}\,\pi_i$ and measures how closely related the observed longest run length in an *M*-bit blocks and the expected longest length are, k is the degree of freedom, k and N are determine by *M* in accordance to Table 2.

Table 2: M, k, N values

M	k	N
8	3	16
128	5	49
10^2	6	75

e) The Binary Matrix Rank Test: This determines the original sequence linear dependences in a given fixed length substrings. The mathematically representation is as given

$$P-value = e^{-\frac{\chi^2(obs)}{2}} \quad (6)$$

Where e is exponential function and χ^2 is the same as in equation (5)

f) The Spectral Test: This test is for periodic features detection and it shows the deviation between the peak numbers above the 95% threshold and described using.

$$P-value = erfc\left(\left|\frac{(N_1 - N_0)}{\sqrt{n(.95)(.05)/4}}\right|\sqrt{2}\right) \quad (7)$$

Where N_1 is the actual number of peaks in *M* observed *that* are less than *T, T*is the 95% peak height threshold value and computed as $T = \sqrt{(\log 20)n}$, *M* is the modulus functions consisting of the first n/2 elements in S, $M = |S'|$, N_0 is computed as $N_0 = 0.95n/2$ and it is the theoretical (95%) expected number of peaks (under the assumption of randomness) that are lesser than *T*

g) The Non-overlapping Template Matching Test: This is for detecting generators that produces higher occurrences of a given aperiodic (non-periodic) pattern. The *m*-bit pattern is searched using m-bit window and shifts a bit position once the pattern is not found but reset the window and resumes searches once the pattern is found. The test is mathematically given as follows

$$P-value = igame\left(\frac{N}{2}, \frac{\chi^2(ob)}{2}\right) \quad (8)$$

Where $\chi^2 = \sum_{j=1}^{N} (W_j - \mu)^2 / \sigma^2$, W_j is the template number of occurrences within the block j, μ is the mean and computed as $\mu = (M - m + 1)/2^m$, m is each templates length in bits and it is the target string, σ is the variance and computed as $\sigma^2 = M\left(\frac{1}{2^m} - \frac{2m-1}{2^{2m}}\right)$

h) The Overlapping Template Matching Test: This is identical to non-overlapping template matching test and focuses on pre-specified target strings number of occurrences. The difference between this tests is that once the pattern is found, the window shift one bit only before continuing the search. It is mathematical given as

$$P-value = igamc\left(\frac{k}{2}, \frac{\chi^2(ob)}{2}\right) \quad (9)$$

Where χ^2 is the same as equation (5), but $n = 2(M - m + 1)/2^m$

i) Maurer's Universal Statistical Test: This shows the extent by which a sequence can be compressed without losing any information and it is represented mathematically as

$$P-value = \left(\frac{f_n - L}{\sqrt{2\sigma}}\right) \quad (10)$$

Where L is the expected value, σ is computed as $\sigma = c\sqrt{L}/k$, c is a constant defined as $c=0.7-\frac{0.8}{L}+\left(4+\frac{32}{L}\right)\frac{K^{-3/L}}{15}$ while f_n is the statistics test computed $f_n = \frac{1}{k}\sum_{i=Q+1}^{Q+K}\log_2\left(i-T_j\right)$ and Q is the numbers of blocks in the initialization sequence.

j) The Linear Complexity Test: This is used in determining how complex enough a sequence is to be considered random.

$$P-value = igamc\left(\frac{k}{2},\frac{\chi^2(ob)}{2}\right) \quad (11)$$

Take χ^2 as g and $\mu = M/2$

k) The Serial Test: This test is similar to frequency test when m=1 and it concentrates on the occurrence rate of all possible overlapping *m*-bit patterns across the entire sequence. All *m*-bit pattern has the same probability of appearing as every other *m*-bit pattern due to the uniformity of random sequence. It is mathematically expressed as

$$P-value1 = igamc\left(2^{m-2},\nabla\psi^2{}_m\right) \quad (12)$$

$$P-value2 = igamc\left(2^{m-3},\nabla^2\psi^2{}_m\right) \quad (13)$$

Where $\nabla\psi^2{}_m(obs)$ is the observed frequencies, $\nabla^2\psi^2{}_m(obs)$ is the expected frequencies of the *m*-bit patterns.

l) Approximate Entropy Test: **This** test is just like the serial test but compares the frequency of overlapping blocks of two neighboring lengths (*m* and *m+1*) against the expected result for a random sequence.

$$P-value = igamc\left(2^{m-1},\frac{\chi^2}{2}\right) \quad (14)$$

Where; $\chi^2 = 2n\left[\log 2-(\varphi^{(m)}-\varphi^{(m+1)})\right]$

, $\varphi^{(m)} = \sum_{i=0}^{2^m-1}\pi_i \log \pi_i$, $\pi_i = C_j$ and $j = \log_2 i$

m) The Cumulative Sums Test: This test is use in determining how large or small the tested sequence cumulative sum is in relative to the expected behavior of a cumulative sum for random sequences. The cumulative sum is referred to as a random walk. A random walk is a sequence of random steps of unit length taken at the beginning before returning to origin and mathematically expressed as

$$P-value=1-\sum_{k=\left(\frac{-n}{z}+1\right)/4}^{\left(\frac{n}{z}-1\right)/4}\left[\phi\left(\frac{(4k+1)z}{\sqrt{n}}\right)-\phi\left(\frac{(4k-1)z}{\sqrt{n}}\right)\right]+\sum_{k=\left(\frac{-n}{z}+1\right)/4}^{\left(\frac{n}{z}-1\right)/4}\left[\phi\left(\frac{(4k+3)z}{\sqrt{n}}\right)-\phi\left(\frac{(4k+1)z}{\sqrt{n}}\right)\right] \quad (15)$$

Where ϕ is the cumulative probability distribution function

n) The Random Excursions Test**:** This test determines the number of cycles with approximately *K* visits in a cumulative random walk from a partial sum (0, 1) sequences to the corresponding (-1, +1) sequences and mathematically expressed as:

$$P-value = igamc\left(\frac{k}{2},\frac{\chi^2(obs)}{2}\right) \quad (16)$$

Where $\chi^2(obs) = \sum_{k=0}^{k}\frac{\left(v_k(x)-J\pi_k(x)\right)^2}{J\pi_k(x)}$, *J* is the sum of zero crossings in *S′,* the zero crossing is a value of zero in *S* ' that occurs after the starting zero

o) The Random Excursions Variant Test: This test is for detecting the deviations from the expected total number of visits to various states in the random walk and the mathematically expression as

$$P-value = erfc\left(\frac{|\xi(x)-J|}{\sqrt{2J\left(4|x|-2\right)}}\right) \quad (17)$$

III. DEVELOPED SYSTEM

The true random number generator was implemented using python programming language [19]. The mouse wait time was evaluated by measuring the serial movement between the coordinates returned from the mouse operation using equation (18)

$$M_t = t_{1x}^{y} - t_{0x}^{y} \quad (18)$$

Where; x refers to the mouse position in the x-coordinate, y is the y-coordinate of the mouse, t_1 is the final time (milliseconds) spent, t_0 is the initial start time (ms) at the coordinate position
and M_t is the Mouse wait time

To enhance the entropy of the random number produced by the mouse wait time, the mouse wait time were exclusive-OR (XOR) with the system time and calendar using equation 19.

$$s_{dt}XORm_t = \sum_{k=0}^{[\log_2(s_{dt})]}2^k\left[\left(\frac{s_{dt}}{2^k}+\frac{m_t}{2^k}\right)\bmod 2\right] \quad (19)$$

Where; S_{dt} is the system current datetime (ms)

The TRNG generates 256-bits random number which can be increase to 1 Mb bit stream by increasing the mouse wait time A total of ten different random sequences were generated and

subjected to the 15 NIST SP Test suite in order to effectively and efficiently evaluate the bits randomness. Furthermore, the acceptable range of proportion was evaluated using equation (20)

$$R = \rho \pm 3\sqrt{\frac{\rho(1-\rho)}{m}} \qquad (20)$$

Where R is the acceptable range, ρ is the proportion and $\rho = 1 - \alpha$ while m is the sample size.

A. *Experimental Result*

The tests were performed on 1000 samples of 1-Mbit at a significance level of $0.01(\alpha = 0.01)$ with minimum passing ratio evaluated using equation (21) and the result is as shown in Table 3

Table 3: NIST SP 800-22 Test Result for the Generated bits

N=10^6	**P-value**	**Proportion**	**Result**
Frequency (Monobit)	0.98228	0.998	Pass
Block Frequency Test	0.98796	0.978	Pass
Runs Test	0.39984	0.911	Pass
Block Longest-Run-of-Ones	0.84193	0.985	Pass
Binary Matrix Rank	0.97681	0.987	Pass
Spectral Discrete Fourier Transform Test	0.84297	0.992	Pass
Non-overlapping Template Matching (m = 9, B =000000001)	0.75872	0.991	Pass
Overlapping Template Matching Test (m = 9)	0.99086	0.988	Pass
Maurer's Universal Statistical Test (L = 7, Q = 1280)	0.86540	0.982	Pass
Linear Complexity	0.65245	0.990	Pass
Serial Test	0.71296	0.985	Pass
Approximate Entropy	0.82365	0.991	Pass
Cumulative Sums Test	0.67248	0.989	Pass
Random Excursions (x = +1)	0.69768	0.993	Pass
Random Excursions Variant Test	0.76398	0.984	Pass

A perfect randomness is obtained when *P*-value is equal to 1 while a null randomness has a 0 as *P*-value. To pass all the required tests, the value of P must be greater than the threshold predefined δ value. Once all the 15 tests are passed, then the sequence is considered to be random with a confidence of $1-\delta$ else the tested sequence is considered not random. A α of 0.01 shows that 1 sequence in 100 sequences is expected not to be accepted. A *P*-value ≥ 0.01 indicates a 99% confidence on the generated random sequence while *P*-value < 0.01 meant that the sequence is nonrandom with a confidence level of 99%. From the evaluated range of acceptable proportion using a significant level of 0.01 is between 0.98056 and 0.99943. From Table 1 it can be clearly seen that the generated bits passed all the 15 NIST statistical test with P-value > 0.0001 and are within the expected proportion rate of 0.98056.This clearly indicates that the developed system has a good statistical randomness.

Furthermore equation (1) was used to establish the uncertainty of the generated bits randomness. Each number of bits is represented as a byte (i.e. j = 8).Where the maximum Shannon entropy is equal to 1. Instead of producing a continuous 1 Mb bits stream to obtain the desired length, the generator was made to produce a 64-bit sequences per time before been reset. In producing the 1-Megabytes sequences, the 64-bit sequences was repeated for 131072 times. The developed generator produced an array elements of 1024 of 64 bits in each run. The first element was chosen before resetting the generator. This was to avoid producing values that have relationship with the previous or next values $\left(\frac{64*131072}{1024*1024*8}\right)$.

In order to evaluate the proposed random number 15 statistical test suite level, the p-value for (Ray & Milenković, 2018), Yeoh *et al.* (2019) and Kim *et al.* (2019) were compare and the chart is as shown in Fig. 3

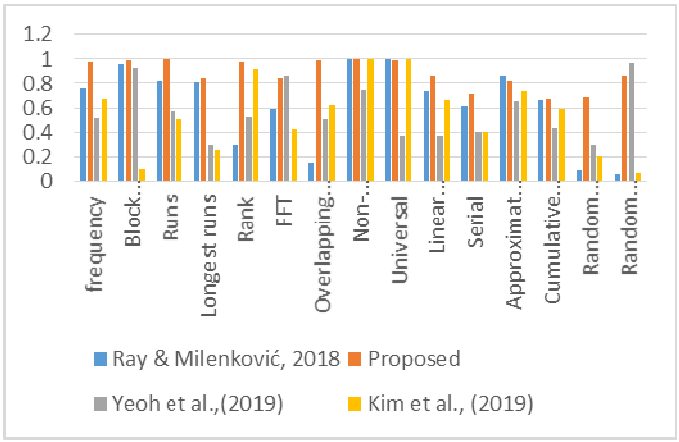


Fig.3: Comparison of the Developed True Random Numer Generator P-value with current generators.

From Fig.3 it can be clearly seen that the developed TRNG passes all the 15 NIST 800-22 statistical test suites with higher p-value when compared with others. Therefore the generator

is highly unpredictale and can be adopted in generating random numbers for secure communication.

IV. CONCLUSION

This paper presented a new true random number generator based on mouse wait time and system clock datetime. The developed algorithm takes advantage of uncertainties in determining mouse movement wait time and the dynamic system clock datetime. The generated bits were evaluated using the NIST SP 800-22, and the generator passed all tests indicating its capability to generate high-quality true random numbers. Furthermore, an entropy analysis was performed in order to prove that the generator bits are non-deterministic by satisfying the backward and forward unpredictability requirements. The developed generator is cheaper, convenient and universal for all personal computer as entropy source is from the personal computer and thus can be adopted for generating bits that can be used as secret keys in secure applications.

V. REFERENCES

[1] Kim, J. S., Patel, M., Hassan, H., Orosa, L., & Mutlu, O. (2019). D-range:Using commodity dram devices to generate true random numbers with low latency and high throughput. *ser. HPCA, 14*.

[2] Kiamehr, S., Golanbari, M. S., & Tahoori, M. B. (2017). *Leveraging aging effect to improve SRAM-based true random number generators.* Paper presented at the Proceedings of the Conference on Design, Automation & Test in Europe.

[3] Tang, H., Qin, T., Hui, Z., Cheng, P., & Bai, W. (2018). *Design and Implementation of a Configurable and Aperiodic Pseudo Random Number Generator in FPGA.* Paper presented at the 2018 IEEE 2nd International Conference on Circuits, System and Simulation (ICCSS).

[4] Morrison, D., Delic, D., Yuce, M. R., & Redouté, J.-M. (2019). Multistage Linear Feedback Shift Register Counters With Reduced Decoding Logic in 130-nm CMOS for Large-Scale Array Applications. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 27*(1), 103-115.

[5] Granger, R., Kleinjung, T., & Zumbrägel, J. (2018). On the discrete logarithm problem in finite fields of fixed characteristic. *Transactions of the American Mathematical Society, 370*(5), 3129-3145.

[6] Qureshi, C., Campello, A., & Costa, S. I. (2018). Non-existence of linear perfect Lee codes with radius 2 for infinitely many dimensions. *IEEE Transactions on Information Theory, 64*(4), 3042-3047.

[7] Ateniese, G., Fech, K., & Magri, B. (2018). *A Family of FDH Signature Schemes Based on the Quadratic Residuosity Assumption.* Paper presented at the International Conference on Cryptology in India.

[8] Wolfram, S. (2018). Cellular automata and complexity: collected papers: CRC Press.

[9] Herrero-Collantes, M., & Garcia-Escartin, J. C. (2017). Quantum random number generators. *Reviews of Modern Physics, 89*(1), 015004.

[10] Song, P., Zeng, Y., Liu, Z., Ma, J., & Liu, H. (2018). *True Random Number Generation Using Process Scheduling of Android Systems.* Paper presented at the 2018 International Conference on Networking and Network Applications (NaNA).

[11] Ray, B., & Milenković, A. (2018). True random number generation using read noise of flash memory cells. *IEEE Transactions on Electron Devices, 65*(3), 963-969.

[12] Buchovecká, S., Lórencz, R., Kodýtek, F., & Buček, J. (2017). True random number generator based on ring oscillator PUF circuit. *Microprocessors and Microsystems, 53*, 33-41.

[13] Avaroğlu, E., Koyuncu, İ., Özer, A. B., & Türk, M. (2015). Hybrid pseudo- random number generator for cryptographic systems. *Nonlinear Dynamics, 82*(1-2), 239-248.

[14] Arslan Tuncer, S., & Kaya, T. (2018). True Random Number Generation from Bioelectrical and Physical Signals. *Computational andmathematical methods in medicine, 2018*.

[15] Street, R. L., Liu, L., Farber, N. J., Chen, Y., Calvitti, A., Weibel, N., Rick,S. (2018). Keystrokes, mouse clicks, and gazing at the computer: how physician interaction with the EHR affects patient participation. *Journal of general internal medicine, 33*(4), 423-428.

[16] Hu, Y., Liao, X., Wong, K.-w., & Zhou, Q. (2009). A true random number generator based on mouse movement and chaotic cryptography. *Chaos, Solitons & Fractals, 40*(5), 2286-2293.

[17] Comer, D. (2017). Essentials of computer architecture: Chapman and Hall/CRC.

[18] Bassham, L. E., Rukhin, A. L., Soto, J., Nechvatal, J. R., Smid, M. E., Leigh, S. D., Banks, D. L. (2010). A statistical test suite for random and pseudorandom number generators for cryptographic applications.

[19] Sun, Q., Berkelbach, T. C., Blunt, N. S., Booth, G. H., Guo, S., Li, Z., Sharma, S. (2018). PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 8*(1), e1340

[20] Lonardoni, D., Lovato, A., Gandolfi, S., & Pederiva, F. (2015). Hyperon puzzle: hints from quantum Monte Carlo calculations. *Physical review letters*, *114*(9), 092301.

[21] Sillitoe, N., & Hilico, L. (2016). Numerical Simulations of Ion Cloud Dynamics. In *Trapped Charged Particles: A Graduate Textbook with Problems and Solutions* (pp. 161-177).

[22] Krishnamoorthy, K. (2016). *Handbook of statistical distributions with applications*. Chapman and Hall/CRC.

[23] Zhang, L., & Suganthan, P. N. (2016). A survey of randomized algorithms for training neural networks. *Information Sciences*, *364*, 146-155.

[24] Xingyuan, W., Xue, Q., & Lin, T. (2012). A novel true random number generator based on mouse movement and a one-dimensional chaotic map. *Mathematical Problems in Engineering, 2012*.

[25] Zhu, H., Zhao, C., Zhang, X., & Yang, L. (2013). A novel iris and chaos-based random number generator. *Computers & Security, 36*, 40-48.

[26] Teh, J. S., Samsudin, A., Al-Mazrooie, M., & Akhavan, A. (2015). GPUs and chaos: a new true random number generator. *Nonlinear Dynamics, 82*(4), 1913-1922.

[27] Yeoh, W.-Z., Teh, J. S., & Chern, H. R. (2019). A parallelizable chaos-based true random number generator based on mobile device cameras for the Android platform. *Multimedia Tools and Applications, 78*(12), 15929-15949.

[28] Kiamehr, S., Golanbari, M. S., & Tahoori, M. B. (2017). *Leveraging aging effect to improve SRAM-based true random number generators.* Paper presented at the Proceedings of the Conference on Design, Automation & Test in Europe.

Proceedings of the 2nd International Conference of the IEEE Nigeria Computer Chapter:
IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

Implementation of Timing and Synchronization in Digital Clocks: A Wireless Communication Design

S. Garba, B. Yahaya, B.O Sadiq, D. A Udekwe, Z. M. Abubakar
Department of Computer Engineering
Ahmadu Bello University, Zaria
sgarba@abu.edu.ng, byahaya@abu.edu.ng, bosadiq@abu.edu.ng

***Abstract*— This paper presents an improved method in solving the problem of time lag experienced by digital clocks, which are often due to power loss. The design concept in this paper was validated by designing a prototype. A master clock connected to a real time clock transmits time through a transceiver to slave clocks within a defined region to ensure that time displayed on the clocks is synchronized. This results in the master clock updating the slave clocks every second through the transceivers to maintain non-varying time. The setup in this paper verified the possibility of having synchronized time which is always accurate with the real time clock.**

Keywords— Wireless Communication, Clock Synchronization and Timing

I. INTRODUCTION

Clocks does not only serve as a means of indicating time but it also plays a very important role in the operation of all services that run on a digital network. Devices often communicate with each other and time is exchanged [1][2]. Times on hardware clocks are initially set correct. But as each clock counts time differently; this results in a clock drift [3]. Clock drift causes an awkward moment whereby times do not agree on devices that uses timestamps when files are transferred between a client and a server. This makes it inconvenient in comparing logs on two systems. As such, there comes the need for clock synchronization. Clock synchronization is the coordination of multiple independent clocks that differ in time due to clock drift [4-6]. When there is a clock drift, independent clocks will not run at precisely the same rate as a reference clock. Power loss amongst other factors are the causes of clock drift in hardware clocks [7].

In order to solve the problem of clock drift, this paper proposed the use of a wireless communication technology whereby time coordination and synchronization is achieved. Validation of the proposed solution is achieved by developing a prototype for wireless time synchronization between hardware digital clocks. Notwithstanding, a number of works in literature have developed techniques of timing and synchronizing hardware digital clocks. However, limitations exist with their design approach and are reviewed as follows:

The authors in [8] designed a digital clock using a microcontroller. The clock was able to retain time as long as it was powered. However, no provision was made to set the clock after a power failure. [9] designed a digital clock with multiple faces using a microcontroller. The clock was able to display time in multiple directions. However, the work done was insufficient as no provision was made to take care of power outages which causes clock drift. The work of [10] designed a digital clock using a microcontroller. A separate power supply source was provided along with user buttons to set the display of the clock thereby reducing the effect of clock drift. Nonetheless, the design was prone to communication failures. Authors in [11] designed a microcontroller based digital clock with calendar and message display. However, the operation of the clock seemed too complex as the clock was difficult to set. [12] designed and constructed a digital clock. The designed approach of the clock was reliable, portable and fully functional with an alarm system.

However, from the reviewed sister papers, there is a need to design a hardware digital clock that will incorporate Real Time Clock (RTC) module to ensure that the clocks never lose time, a wireless transceiver module to ensure that time is synchronized between multiple clocks located in various places and a communication module to aid easy setting of time via a remote location. As such, the contributions of this paper are as follows:

i. We designed an improved digital synchronizing clock that can transmit and receive information updates wirelessly.

ii. We implemented a digital synchronizing clock that can be managed remotely via the use of a mobile application.

The remaining part of this paper as arranged as follows: section two presents the proposed design in details, section three presents the discussion of the result obtained as well as the prototype demonstration. Section iv presents the conclusion.

II. PROPOSED DESIGN

Modern embedded systems that uses microcontrollers are mostly used to perform a wide range of functions. Since the embedded system is dedicated to specific tasks, design engineers can optimize it to reduce the size and cost of the product and increase the reliability and performance. Due to this fact, this work employs the use of the ATmega325P microcontroller in implementing timing and synchronization

of digital clocks. The overall systems design is separated into three constituent parts which are: the hardware design methodology, electrical connections and software design methodology. The hardware utilized in this work consists of LED strips mounted on a Vero board which serves as the clock display, a power supply unit which consists of transformer and a voltage regulator and supporting hardware which consists of an Arduino Nano microcontroller, a Real Time Clock, a Bluetooth module and a wireless transceiver. The flow chart of the design methodology is presented in Figure 1.

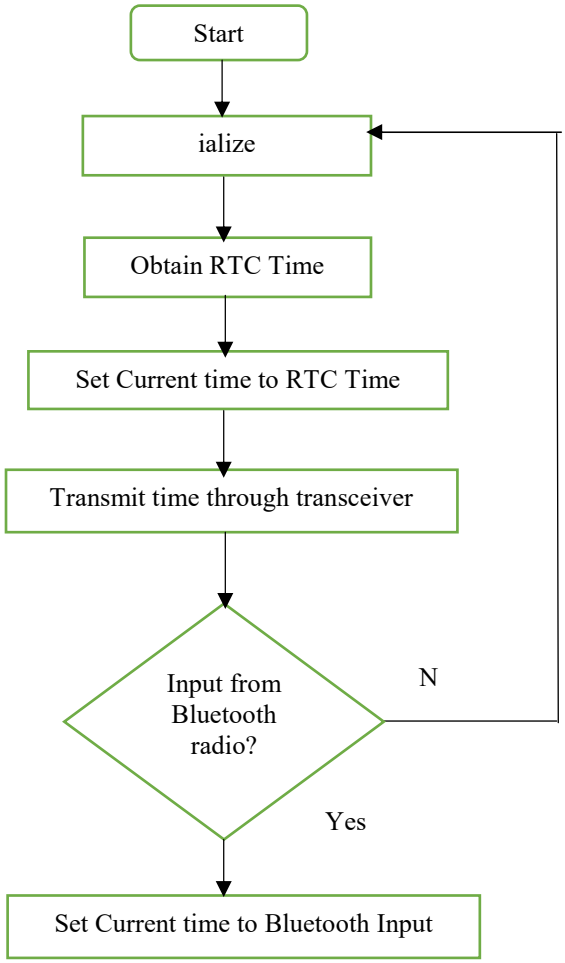


Figure 1: Flow Chart of the Design Methodology

A. Power Supply

To provide power to the entire system, specific considerations had to made. The Arduino Nano requires a 5v supply for reliable operations. Besides this, the LED strips require 12v, the Real Time Clock requires 5v, the nRF24L01 transceiver requires 3.3v while the HC- 05 Bluetooth module also requires 5v. Therefore, the power supply to be used needed to be able to power all the components listed earlier. Most importantly the power supply also needed to be able to supply power for long periods of time without failure as the clocks need to be always on. As a result, mains AC was determined to be the best source for this work. After due considerations a power supply consisting of AC supply of 220v transformed down to 12v, then rectified to DC using a full wave bridge rectifier circuit and smoothing capacitors and then further regulated to 5v using an LM7805 voltage regulator to provide a steady regulated voltage of 5v for the microcontroller and also 12v for the LED strips. The average voltage or the DC voltage available after rectification is given by [9]:

$$vdc = \frac{1}{\pi}\int_0^{\pi} Vm * sinwtdwt = 2\frac{Vm}{\pi}$$

(1)

Also, the maximum ripple voltage present for a full wave rectifier circuit is not only determined by the value of the smoothing capacitor but by the frequency and load current as shown in equation (2)

$$Vripple = \frac{I}{F*C} volts$$

(2)

Where, I is the DC load current in amps, F is the frequency of the ripple or twice the input frequency in Hertz and C is the capacitance in Farads. The circuit diagram of the developed power supply is presented in figure 2.

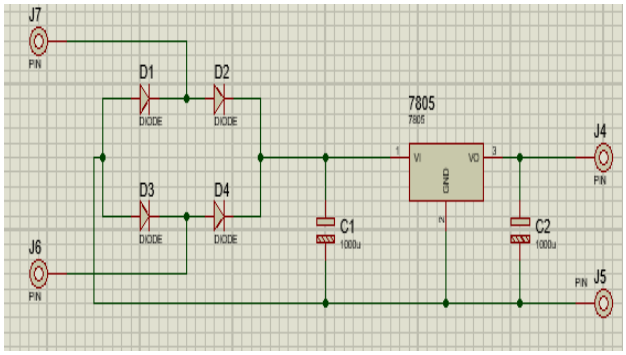


Figure 2: Power Supply Circuit Design

B. Base Supporting Hardware

The base supporting hardware consists of the following components: a DS3231 Real Time Clock, an Arduino Nano microcontroller, an nRF24L01 wireless transceiver and a HC-05 Bluetooth module. In order to ensure the clocks don't lose time in the event of a power outage, a separate source of time had to be sought for. This necessitated the use of the DS3231 Real Time Clock, which would be sufficiently small and low powered but would provide a backup. The Arduino Nano was used, as it is readily available, low cost and provides sufficient computing power to handle all operations at a relatively small size as compared to other microcontrollers of such capacity. In situations where multiple clocks exist at different points within a defined region, it is important for all clocks to display accurate time which is also synchronized across all point in that region. Therefore, an nRF24L01 wireless transceiver was used to send time from a master to a slave clock. This particular transceiver was chosen because it low powered and can be easily interfaced with the Arduino microcontroller. Also, a HC-05 Bluetooth module was installed on the master clock

to aid easy setting of time through a built Android application. The HC-05 module was chosen because it is easy to use and is readily available. After due considerations, the above-mentioned components were selected and mounted on a printed circuit board connecting the various devices to the microcontroller as depicted in Figure 3.

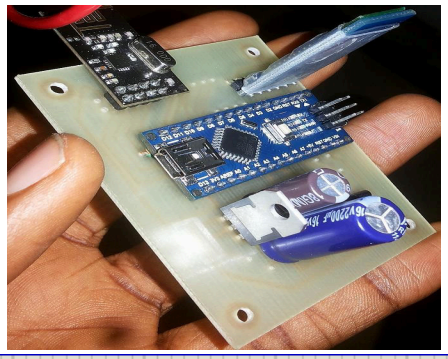


Figure 3: Base Supporting Hardware Component

C. *Electrical Connections*

Once the hardware had been selected and setup, connections were then made between various components as depicted in Figure 4.

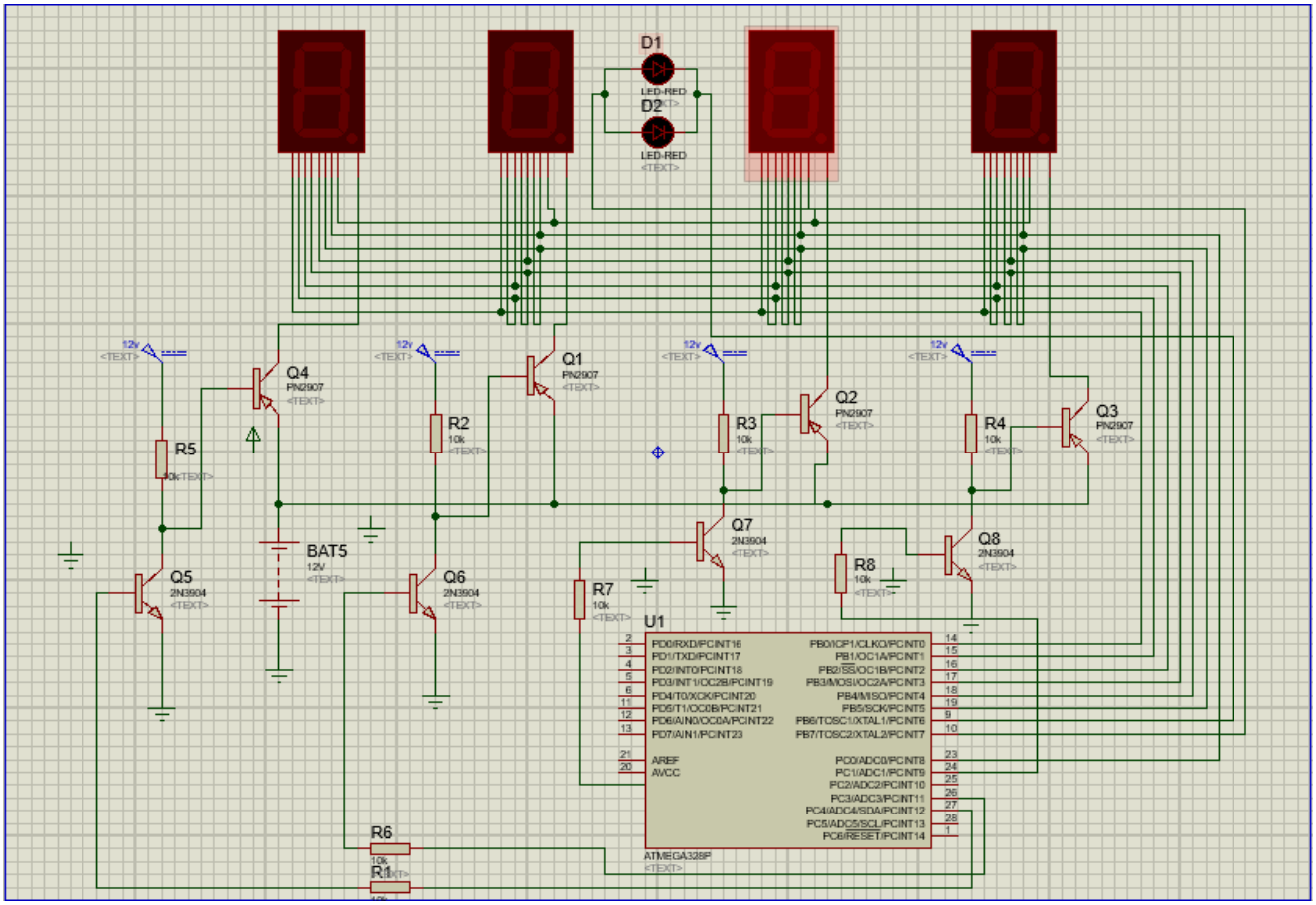


Figure 4: Electrical Connections Between Components

When a signal is sent from the microcontroller to the clock through any of its digital pins, such signal travels to the base of an NPN transistor through a current limiting resistor.
Since the emitter of the NPN transistor is grounded (connected to 0v), the signal flows to the connector of the NPN transistor, which then switches off a PNP transistor causing no signal to flow from the collector of the PNP transistor to the clock. The flow or non-flow of signals from the microcontroller determines which digits of the clock turn on or off.

D. *Software Design*

The software used in this paper consists of two major parts. The first is the base control Arduino program which comprises of all code interfacing the LED displays with the microcontroller, Real Time Clock and Bluetooth module.

The second is the android software which runs on an android device that is used to set the clocks. The android software designed to set the clocks was composed and written using the MIT App Inventor software and it consists of just three buttons as shown in figure 5.

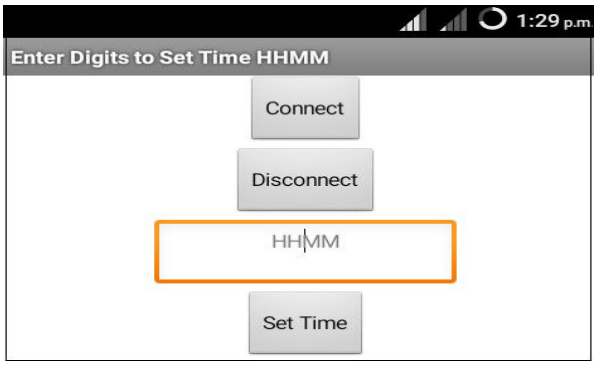


Figure 5: Android Mobile Interface

The first button of the software called the connect button scans through all neighboring devices and establishes a connection between the mobile device and the microcontroller. The second button of the clock is used to terminate the connection between the mobile device and the microcontroller. The last button of the clock sends the value of time imputed in the textbox to the microcontroller.

Table 1 shows the component used in the implementation of timing and synchronization in digital clocks.

Table1: Design Component

a. S/N	b. Components	c. uantity
d.	e. ED strips	f. 6 pieces
g.	h. rduino Nano	i.
j.	k. RF24L01	l.
m.	n. ransistors	o. 0
p.	q. iodes	r.
s.	t. apacitors	u.
v.	w. oltage Regulators	x.
y.	z. ero Boards	aa.
bb.	cc. eal Time Clock	dd.
ee. 10	ff. luetooth radio	gg.

III. RESULT AND DISCUSSION

After assembling the system, the objective of the paper which was to design an improved digital clock to synchronize with one another thereby solving the problem of clock drift was largely met. The completed prototype of the digital clock is presented in figure 6. The digits constructed on the clock display the time received from the Real Time Clock via the microcontroller every second. This time is then transmitted from the microcontroller on the sending node to the receiving node via the wireless transceiver every second. Whenever there is an input from the Bluetooth radio attached to the sending node, the time in the Real Time Clock's memory is updated and then entire process of updating the clock's display and sending this time through the microcontroller continues.

Figure 6: Clock Synchronization Result

IV. CONCLUSION

This work proposed and implemented an improved timing and synchronization of digital clock through the use of wireless communication. The design eliminated the effect of clock drift cause by power failures. Though, inexpensive hardware component was used in the design but the accuracy is high and the objective of the design was largely met.

REFERENCES

[1] Shulong Wang, Yibin Hou, Fang Gao, Songsong Ma, "A novel clock synchronization architecture for IoT access system", *Computer and Communications (ICCC) 2016 2nd IEEE International Conference on*, pp. 1456-1459, 2016.

[2] R. Cochran, C. Marinescu, C. Riesch. Synchronizing the Linux system time to a PTP hardware clock. *IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*, 2011, doi:10.1109/ispcs.2011.6070158.

[3] Jingmeng Liu, Xuerong Li, Min Liu, Xikui Cui, and Dong Xu. A New Design of Clock Synchronization Algorithm. *Hindawi Publishing Corporation Advances in Mechanical Engineering*, 2014, 1-9.

[4] S. K. Mani, R. Durairajan, P. Barford, J. Sommers. A System for Clock Synchronization in an Internet of Things, Retrieved from: *arXiv:1806.02474v1 [cs.NI],* 2018, 1-18.

[5] N. Khandare, M. Bhavsar, P. Kumare, S. Raksha. Clock Synchronization in Distributed System. *International Journal of Computer Science and Information Technologies*, 2013, 4(3), 457-460.

[6] M. Buevich, N. Rajagopal, A. Rowe. Hardware Assisted Clock Synchronization for Real-Time Sensor Networks. *IEEE 34th Real-Time Systems Symposium.* 2013, doi:10.1109/rtss.2013.34.

[7] A. Ciuffoletti. Power-Aware Synchronization of a Software Defined Clock. *Journal of sensors and actuators networks*, 2019, 8(11), 1-19.

[8] N. Muslim, M. T. Adnan, M. Z. Kabir, M. H. Kabir, S. M. Islam. Design and implementation of a digital clock showing digits in Bangla font using microcontroller AT89C4051, 2012. Retrieved December, 2018 from *arXiv:1208.0995v1 [cs.AR].*

[9] L. A. Ajao, M. A. Adegboye, E. M. Dogo, S. O. Aliyu, D. Maliki. Development and Implementation of Microcontroller-based Improved Digital Timer and Alarm System. *International Conference on Information and Communication Technology and Its Applications (ICTA)*, 2016, 28-30.

[10] I. S Khalid, N. Afrin, R. Mikail. Design and Implementation of a Microcontroller Based LCD Screen Digital Stop Watch. *International Journal of Electrical and Information Engineering,* 2011, 5(2), 183-187.

[11] M. M. Rahman, M. A. Islam, R. S. Rajan, D. M. Parvez, M. R. Islam, M. N. Islam. Microcontroller and LCD Based Digital Bangla Clock cum Calendar. *International Journal of Engineering Research and Development, 4(1),* 2012, 99-103.

[12] Victor Kathan Sarker, M. Ataur Rahman, and M. A. Matin. Design and Development of Microcontroller Base Digital Bangla Clock. *International Journal of Computer Theory and Engineering*, 2012, 4(6), 1-3

Proceedings of the 2nd International Conference of the IEEE Nigeria Computer Chapter:
IEEEnigComputConf'19: Ahmadu Bello University, Zaria, Nigeria, October 14-17, 2019

Development and Testing of Q-Basic Computer Software for Newton-Raphson Power Flow Studies

***Abstract*—There had been obvious system insecurity in Nigerian power systems' grid which often lead to frequent power system collapse. Studies have shown that the Privatized power companies that emanated as a result of power sector reform carried out recently in Nigeria are not fully on ground in network argumentation, planning and analysis of the network. For example, it is required that for any proposed capital project execution which would add to the existing Networks; certain analysis via system simulation need to be carried out. Such analysis would display the likely situation after the project is completed and commissioned. These analyses include investigating the bus voltage profile, power flow, losses, overvoltage condition etc. Consequently, this paper presents a development of Q-basic computer software package on Newton- Raphson power flow algorithm and subsequent testing on a 3 - Bus power system network with known solutions. Steady state operating conditions of busbars, generation, branch power flows and circuit system losses were determined. The obtained results are presented and discussed. These results are accurate and reliable, because they agree with the known solutions of the network. Conclusions are drawn, and necessary recommendations are presented.**

***Index Terms*—System Collapse, Simulation, Grid, Argumentation, Newton-Raphson, Power Flow, Q-Basic Software Package, Busbars.**

I. Introduction

Electric power has become virtually indispensable in modern societies in the sense that most industrial, commercial, domestic and social activities are heavily dependent on availability of electric energy. Electric power generating stations are often located far from the load (or power consumption) centres. Conventionally, several generating stations are connected and to the load centres via transmission lines to form an electric power system also known as the grid [1], [2]. A power system is required to maintain a continuous balance between electrical generation and a varying load demand so that system frequency and voltage are maintained at their statutory levels, while guaranteeing system security.

The performance of the electrical power systems is continually and extensively analyzed to assess its quality. Such analysis is used in planning expansion schemes on the system to cater for the ever-increasing power demand [3]–[7]. Some of the analysis problems in system planning, requiring attention include power (or load) flow, fault level, stability and reliability.

However, power flow studies frequently precede and, in most cases, provide the starting conditions for complicated system analysis such as optimization. Fault level, stability and outage security assessment. Generally, the problem encountered in transmission network planning has always been the multitude of situations to be examined in order to guarantee that a projected network is necessary and enough for a given year. Such situations arise from load variations, hydraulic variations, generation unit's unavailability leading to a reorganization of the generation program with every failure, and finally unavailability's due to transmission equipment (i.e lines and transformers) [8], [9].

In Nigeria, Electricity Companies comprising Genco, TCN and Disco are responsible for planning, installation, generation, transmission and distribution of electrical energy in the country and some neighbouring countries. They have several generating stations including a mix of hydro and thermal power plants mostly located in the southern and middle parts of the country. Several papers had been written and research had been carried out on power flow analysis for power system grid using the Gauss, Gauss-Seidel, Fast decoupling as well as other several iterative approaches which are all slow in convergence [10]–[18]. Newton-Raphson Iterative method is generally believed to be faster in convergence [1], [4], [5], [7]. Thus, a suitable software package for power flow method is developed.

II. Newton-Raphson Method

The method involves the idea of finding a vector $x \in IR^N$ such that,

$$F(x) = 0 \quad (1)$$

Where F is a vector valued function of dimension N.
Expanding equation (1) in a Taylor Series and due to the error in neglecting the higher order terms in the Taylor series, Value x can be iteratively determined as,

$$x^{(k)} = x^{(k-1)} - J^{-1}F\left(x^{(k-1)}\right) \quad (2)$$

Where, k = iteration, J = A square matrix of same dimension as x and F, and its entries are partial derivatives defined as [4], [8], [18], [19],

$$J_{pq} = \frac{\partial F_p}{\partial x_q}\Big|_{x^{(k-1)}} \quad (3)$$

Where, p and q denote buses. This matrix is the Jacobian matrix of the system of equation (1).

Applying this method to power flow problem, the derivatives of the mismatch form of the power flow equation are used to force the bus power mismatch to zero. An equation is written for each bus p

$$P_p + jQ_p = E_p I_p^* \quad (4)$$

Where,

$$I_p = \sum_{K=1}^{NB} Y_{pk}\, E_k \quad (5)$$

Then,

$$P_p + jQ_p = E_p \sum_{K=1}^{NB} Y_{pk} E_k^* = |E_p|^2 Y_{pp}^* + \sum_{\substack{K=1 \\ K \neq p}}^{NB} Y_{pk}^* E_p E_k^* \quad (6)$$

For p= 1, 2, 3 …… NB, excluding the slack bus.

The parameters contained in the above equations and the subsequent ones are clearly defined in table 1.

A set of starting voltages is often used to commence the iteration process. The power P $+jQ$ calculated using equation (6) is subtracted from the scheduled power $P_s + jQ_s$ at the bus and the resulting errors stored in a vector. Polar co-ordinates are generally employed for the voltages. The voltages' magnitudes and phase angles are adjusted as separate independent variables. Each bus injection equation is differentiated with respect to all independent variables. Thus, for each bus the power mismatch is calculated as [17], [18], [19],

$$\Delta P_p = \sum_{K=1}^{NB} \frac{\partial P_p}{\partial \theta_k} \Delta\theta_k + \sum_{K=1}^{NB} \frac{\partial P_p}{\partial |E_k|} \Delta |E_k| \quad (7)$$

$$\Delta Q_p = \sum_{K=1}^{NB} \frac{\partial Q_p}{\partial \theta_k} \Delta\theta_k + \sum_{K=1}^{NB} \frac{\partial Q_p}{\partial |E_k|} \Delta |E_k| \ldots \quad (8)$$

The partial differentials are arranged in a Jacobian matrix so that the equations (7) and (8) can be written in matrix vector form as,

$$\begin{bmatrix} \Delta P_1 \\ \Delta P_{NB-1} \\ \hline \Delta Q_1 \\ \Delta Q_{NB-1} \end{bmatrix} = \left[\begin{array}{ccc|ccc} \frac{\partial P_1}{\partial \delta_1} & \cdots & \frac{\partial P_1}{\partial \delta_{NB-1}} & \frac{\partial P_1}{\partial |E_1|} & \cdots & \frac{\partial P_1}{\partial |E_{NB-1}|} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial P_{NB-1}}{\partial \delta_1} & \cdots & \frac{\partial P_{NB-1}}{\partial \delta_{NB-1}} & \frac{\partial P_{NB-1}}{\partial |E_1|} & \cdots & \frac{\partial P_{NB-1}}{\partial |E_{NB-1}|} \\ \hline \frac{\partial Q_1}{\partial \delta_1} & \cdots & \frac{\partial Q_1}{\partial \delta_{NB-1}} & \frac{\partial Q_1}{\partial |E_1|} & \cdots & \frac{\partial Q_1}{\partial |E_{NB-1}|} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial Q_{NB-1}}{\partial \delta_1} & \cdots & \frac{\partial Q_{NB-1}}{\partial \delta_{NB-1}} & \frac{\partial Q_{NB-1}}{\partial |E_1|} & \cdots & \frac{\partial Q_{NB-1}}{\partial |E_{NB-1}|} \end{array}\right] \begin{bmatrix} \Delta\theta_1 \\ \Delta\theta_{NB-1} \\ \hline \Delta |E_1| \\ \Delta |E_{NB-1}| \end{bmatrix} \quad (9)$$

The solution to Newton-Raphson problem runs according to flowchart in Figure 1.

TABLE I
DEFINITION OF PARAMETERS

S/N	Symbol	Description
1	P	Active (or real power) in megawatts
2	Q	Reactive (or imaginary power) in megavolt ampere
3	J	Jacobian matrix
4	K	Iteration count
5	I	Current
6	E	Bus voltage in complex form
7	| |	Modulus
8	Y	Admittance
9	NB	Number of buses
10	*	Conjugate
11	∑	Summation
12	Δ	Incremental value
13	∂	Partial differential
14	Є	Mathematical representation of member
15	kV	Kilovolt
16	Genco	Generation Company
17	TCN	Transmission company of Nigeria
18	DISCO	Distribution Company of Nigeria
19	AF	Acceleration Factor
20	NR	Newton Raphson

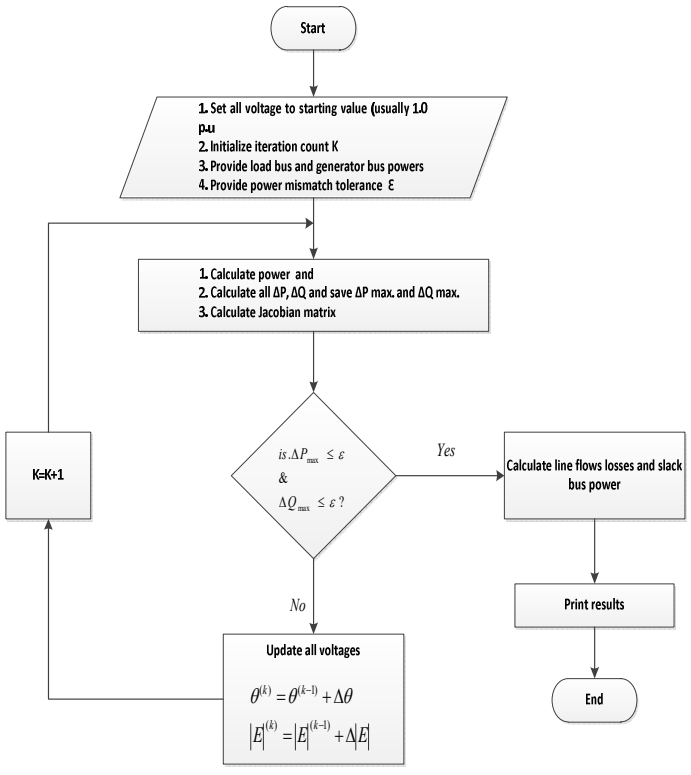


Fig. 1. Flowchart of the Newton – Raphson power flow solution

For a fast convergence [7], [12], common method used is to multiply the voltage corrections by a constant at the end of each iteration. This constant is called acceleration factor [2], [3], [9], [12], i.e

$$Q_p^{(k)}(\text{Accelerated}) = Q_p^{(k-1)} + \propto \Delta Q_p^{(k-1)} \quad (10)$$

$$|E_p|^{(k)}(\text{Accelerated}) = |E_p|^{(k-1)} + \propto \Delta |E_p|^{(k-1)} \quad (11)$$

Where, $\propto$ = acceleration Factor 0.7 <$\propto$> 1.4

III. DEVELOPMENT OF A NEWTON-RAPHSON POWER FLOW (NRPF) SOFTWARE PACKAGE

To enable the solution of power flow problem using the N- R method, a suite of computer programs has been written to create a power flow software package. This NRPF package was developed and implemented using a personal computer (PC). Iterative conversational mode was employed to interface the package with its users. The package was structured in such a way as to fit into the limited core memory of PC. This is achieved by use of the overlay technique described in section A.

A. Structure of the power flow package

Overlaying together with interactive conversational mode are used. Overlaying permits reduction of the working memory of the computer and enables modularization of the package. For the purpose of overlaying, the package is divided into three stands – alone modules such that only one module is resident in core memory at any given time. The CHAIN command of Quick Basic is used to interface the supervisor module NRPF to each of the modules LAMBE 1 and LAMBE 2 as shown in Figure 2. In the next section the functions of the modules are described.

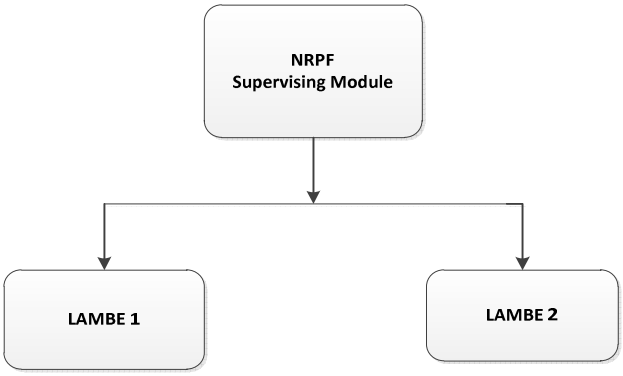


Fig. 2. Structure of NRPF Package

B. Package Description and Functions of the Modules

All programs are coded in Microsoft's Quick Basic (QB). Each module together with its subroutine procedures (if any) is complete. The modules are:

1. NRPF Supervising module
2. LAMBE 1 Power system data entry and Management
3. LAMBE 2 Newton – Raphson power flow solution

1) Supervising Module (NRPF)

NRPF is the package supervisor. It controls the solution procedure by enabling the user to choose which path of the overlaying structure is to be followed. At system level the command

C: \>Quick Basic NRPF

Loads the module into memory and executes it. When the module is active the display screen is as shown in Figure 3.

NEWTON- RAPSON POWER FLOW PROGRAM
1. System data entry
2. Power Flow solution
3. Exit form package

Enter the number (1- 3) of your choice.

Fig. 3. NRPF menu display screen

2) System Data Entry Module (LAMBE 1)

The main function of this module is to permit the power system data to be entered into the computer by the user. The module is loaded and executed by selecting option 1 available on NRPF's menu. The module interactively solicits for the following data from user.

1. Number of buses (NB) in the network
2. Number of transmission lines (NL)
3. Transmission line data
   - Start bus, (SB)
   - End Bus, (EB)
   - Series resistance, R
   - Series reactance, X\
   - Line charging admittance, Y_{sh}
4. Bus scheduled active and reactive powers (P_s and Q_s). The module then formulates bus admittance matrix, Y_{bus}. Due to the non – availability of in – built complex arithmetic functions in Quick Basic, the real and imaginary parts of the elements of Y_{bus} are formed separately. The Y_{bus} building algorithm is as follows;

   Y_{pp} The diagonal entries of Y_{bus} are formed by summing the primitive admittance of lines and line charging admittance.

   Y_{pq} The off-diagonal entries are the negatives of the admittance buses p and q. if there is no line between p and q this term is zero.

All data entered or generated in this module are stored into appropriate data files for future use, so that this module need not be called if the data files are already available. On exit from the module, the NRPF module is reloaded into memory and executed.

3) Power Flow Solution Module (LAMBE 2)

This module implements the Newton- Raphson power flow solution method. It is loaded into memory and executed when option 2 of module NRPF's menu is

selected. It makes use of data files created by LAMBE 1. If such data files are not available, the program returns an error message to the effect that the file is not available. Thus, before this module is run, the user must be sure that LAMBE 1 has already been executed.

Again, since Quick Basic does not have in – built complex arithmetic library functions, the Newton – Raphson power flow solution has to be reformulated in rectangular co-ordinates (rather than the polar co-ordinates) so that real and imaginary parts of variables and parameters are handled separately.

The module consists of a main programme and the following five subroutine procedures.

1. Sub Powers
2. Sub Currents
3. Sub Jacobian
4. Sub Voltage
5. Sub Power flows

The flowchart for the program's logic is shown in Figure 4.

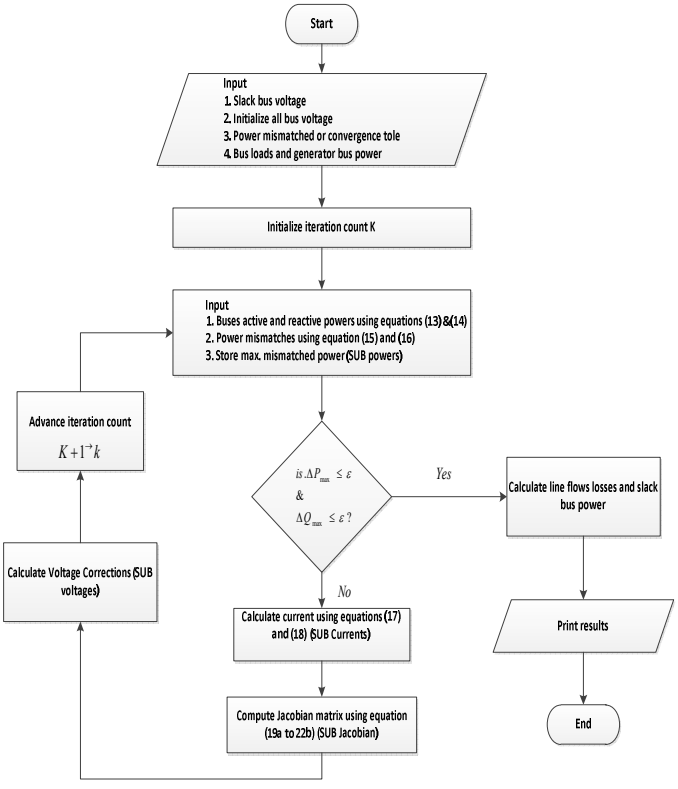


Fig. 4. Flowchart for module LAMBE 2

Consider now the application to an n- bus power system. The solution starts with the iteration counter K set to zero and all the buses except the slack bus being assigned voltages,

$$\mathbf{V}p = 1 + j\,0 \; per unit \qquad (12)$$

For a line connecting buses p and q of admittance,

$$P_p = \sum_{q=1}^{NB} [\, e_p(e_q G_{pq} + f_q B_{pq}) + f_p(f_q G_{pq} - e_q B_{pq})\,]$$

(13)

$$Q_p = \sum_{q=1}^{NB} [\, f_p(e_q G_{pq} + f_q B_{pq}) - e_q(f_q G_{pq} - e_q B_{pq})]$$

(14)

It has been assumed that the slack bus is the bus number 1 of the network. The main program then calls SUB Powers which computes P and Q using equations (13) and (14). It also calculates the power mismatches using equation (15) and (16);

$$\Delta P_p^{(k)} = P_{sp} - P_p^{(k)} \qquad (15)$$

$$\Delta Q_p^{(k)} = Q_{sp} - Q_p^{(k)} \qquad (16)$$

For p = 2, 3, - - - - - - - NB, p ≠ Generator buses.

Furthermore, the maximum mismatches ΔPmax and ΔQmax are stored in memory. If ΔPmax ≤ ϵ per unit and ΔQmax ≤ ϵ per unit, then the iterative process has converged and control transfers to SUB Power flows otherwise control transferred to SUB currents to compute the bus currents using equation (17) and (18)

$$C_p^{(k)} = \frac{P_p^{(k)} e_p^{(k)} + Q_p^{(k)} f_p^{(k)}}{(e_p^{(k)})^2 + (f_p^{(k)})^2} \qquad (17)$$

$$d_p^{(k)} = \frac{Q_p^{(k)} e_p^{(k)} - P_p^{(k)} f_p^{(k)}}{(e_p^{(k)})^2 + (f_p^{(k)})^2} \qquad (18)$$

For p = 2, 3, ---------- NB, p ≠ Generator buses.

After the currents have been calculated, control transfers to SUB Jacobian which computes the elements of the Jacobian matrix using equations (19a) to (22b).

The entries of the Jacobian matrix are computed using equations (19a) to (22b) [19], [20]

Submatrix J^1

$$J_{pq}^1 = \frac{\partial P_p}{\partial e_q} = e_p G_{pq} - f_p B_{pq}, \; p \neq q \qquad (19a)$$

$$J_{pq}^1 = \frac{\partial P_p}{\partial e_q} = e_p G_{pp} - f_p B_{pp} + C_p \qquad (19b)$$

Submatrix J^2

$$J_{pq}^2 = \frac{\partial P_p}{\partial e_p} = e_p B_{pq} + f_p G_{pq}, \; p \neq q \qquad (20a)$$

$$J_{pq}^2 = \frac{\partial P_p}{\partial f_p} = e_p B_{pp} + f_p G_{pp} + d_p \qquad (20b)$$

Submatrix J^3

$$J_{pq}^3 = \frac{\partial Q_p}{\partial e_q} = e_p B_{pq} + f_p G_{pq}, \quad p \neq q \qquad (21a)$$

$$J_{pp}^3 = \frac{\partial Q_p}{\partial e_p} = e_p B_{pq} + f_p G_{pp} - d_p \qquad (21b)$$

Submatrix J^4

$$J_{pq}^{4} = \frac{\partial Q_p}{\partial f_q} = -e_p G_{pq} + f_p B_{pq}, \quad p \neq q$$

(22a)

$$J_{pp}^{4} = \frac{\partial Q_p}{\partial f_p} = -e_p G_{pp} + f_p B_{pp} + c_p$$

(22b)

SUB Voltages is then called to solve the linear system of equations (23) for the voltages Δe and Δf. This subroutine makes use of triangular factorization and forward and back substitution to solve for the voltages.

$$\begin{bmatrix}\Delta P \\ \Delta Q\end{bmatrix} = \begin{bmatrix}\frac{\partial P}{\partial e} & \cdots & \frac{\partial P}{\partial f} \\ \vdots & \vdots & \vdots \\ \frac{\partial Q}{\partial e} & \cdots & \frac{\partial Q}{\partial f}\end{bmatrix}\begin{bmatrix}\Delta e \\ \Delta f\end{bmatrix} = [J]\begin{bmatrix}\frac{\Delta e}{\Delta f}\end{bmatrix}$$

(23)

Where J is the Jacobian matrix which for convenience may be partitioned as:

$$J = \begin{bmatrix}J^1 & \vdots & J^2 \\ J^3 & \vdots & J^4\end{bmatrix}$$

The entries of the Jacobian matrix are computed using equations (19a) to (22b) [19], [20]. This completes the solution for the voltage corrections, $\Delta e^{(p)}$ and $\Delta f^{(p)}$. The subroutine then updates the voltages for the next iteration as,

$$e_p^{(k+1)} = e_p^{(k)} + \Delta e_p^{(k)}$$

(24)

$$f_p^{(k+1)} = f_p^{(k)} + \Delta f_p^{(k)}$$

(25)

The iteration count is increased by 1, i.e, k = k+1, and control is transferred to SUB Powers. When the power mismatch becomes negligible, then control is transferred to SUB Power flows which evaluate the power flows as well as the power losses in the lines. In SUB Power flows the power flowing from Bus p to Bus q and measured at Bus p is given by equation (26),

$$P_{pq} = G_{pq}\left(e_p^2 + f_p^2\right) + \left(e_p f_q - f_p e_q\right) - G_{pq}\left(e_p e_q + f_p f_q\right)$$

(26a)

$$Q_{pq} = B_{pq}\left(e_p^2 + f_p^2\right) + G_{pq}\left(f_p e_q - e_p f_q\right) - B_{pq}\left(e_p e_q + f_p f_q\right) + \frac{Y_{sht}\left(e_p^2 + f_p^2\right)}{2}$$

(26b)

Similarly, the power from q to p and metered at q is given by:

$$P_{qp} = G_{pq}\left(e_q^2 + f_q^2\right) + B_{pq}\left(e_q f_p - f_q e_q\right) - G_{pq}\left(e_q e_p + f_q f_p\right)$$

(27a)

$$Q_{pq} = B_{pq}\left(e_q^2 + f_q^2\right) + G_{pq}\left(f_q e_p - e_q f_p\right) - B_{pq}\left(e_q e_p + f_q f_p\right) + \frac{Y_{sht}\left(e_q^2 + f_q^2\right)}{2}$$

) (27b)

Power losses in each branch of the network are obtained as the algebraic sum of the power calculated using equations (26a) and (27a) and that calculated using equations (26b) and (27b) for active power P and reactive power Q respectively. While the total power losses in the system is the sum of the absolute values of the calculated power losses in each branch of the network. The slack bus power is determined by summing the power flows on the lines terminating at the slack bus.

IV. Units Application of the Developed Software Package to a Known Solution of Power System Network

In this section, usage of the NRPF package is illustrated by testing it for a simple three bus power system whose exact solution is known [19]. The network is shown in Figure 5. The example was used to test each step of the program. Per unit impedances of the lines are clearly written in the Network diagram of figure 5. Other given data are,

v_1 = 1.05 + j 0 (slack bus)

Number of buses = 3

Number of lines = 3

The scheduled power for buses 2 and 3 are

$P_{s2} = 0.096$, $Q_{s2} = -2.07pu$

$P_{s3} = 3.15$, $Q_{s3} = 2.85pu$

Power mismatch tolerance ϵ = 0.001 per unit

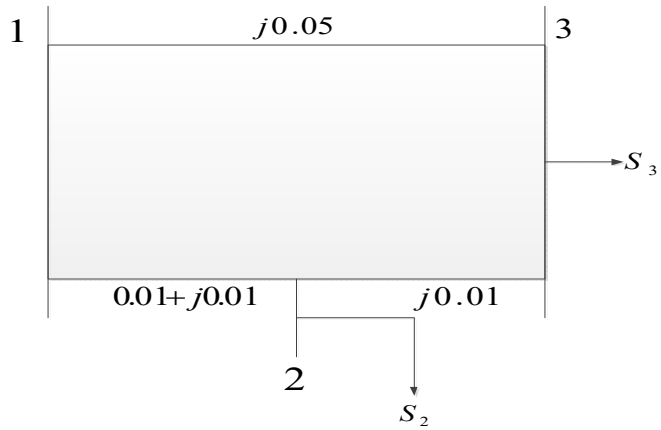


Fig. 5. Simple power system to illustrate usage of the power flow software package [19].

LAMBE 1 is used to input and store the system data:

Number of buses, NB; Number of lines, NL; Start bus, SB, and End bus, EB for each line; Line resistance, R; Line reactance, X; Line susceptance, B; Busbar scheduled power P_s and Q_s.

LAMBE 1 then assembles the bus admittance matrix, Y bus. A11 data are stored in data files (e.g Y2BUS. DAT and PQ2BUS.DAT).

LAMBE 2 then executes with bus voltages initialized to the slack bus voltage value V = 1.05 + j 0 per unit. The Newton- Raphson iteration process converged with power mismatch tolerance ϵ = 0.001 per unit in 5 iterations. The results are as follows:

Number of Buses = 3; Number of Lines = 3; Slack Bus Number = 1, Slack Bus Voltage = 1.05 + j 0; Program Converged in Iteration IT = 5, Power Mismatch Tolerance = .001.

TABLE 4
POWER FLOWS BETWEEN BUSES OF THE TEST NETWORK

Bus		Power flow per unit	
	To	Active	Reactive
ɔm			
	2	1.65	-5.25
	1	-1.953	5.553
	3	-1.08	-8.4
	1	1.08	0.93

TABLE 2
DEFINITION OF PARAMETERS BUSBAR VOLTAGES OF THE TEST NETWORK IN PER UNIT

Bus	Voltage
1	1.05 + j 0
2	1.069 + j - .036
3	1.042 + j - .05

TABLE 3
DEFINITION BUSBAR VOLTAGE MAGNITUDES, PHASE ANGLES, ACTIVE AND REACTIVE POWER OF THE TEST NETWORK

	Voltage		Power	
Bus No	Magnitude	Phase Angle (Degree)	Active	Reactive
1	1	0	0	0
2	1.07	-1.929	9.61	-2.07
3	1.043	-2.967	3.15	2.849

V. DISCUSSION OF RESULTS

In power flow analysis, a generator bus is often used as the slack bus. Hence, bus 1 was chosen as the slack bus and its voltage is 1.0 < 0 per unit for the analysis. The voltage of the generator buses would have been constrained to be 1.0 < 0 per unit but unfortunately none of the buses are with generator.

From table 3, there is a wide variation of voltage profile, thus, for any meaningful analysis, it becomes necessary to identify buses with minimum voltage profile and buses with maximum voltage profile. The minimum and maximum voltage values acceptable for a normal power system network are 0.95 and 1.05 per unit respectively [17]–[20]. On this basis, buses whose voltage magnitude falls above maximum acceptable value are can also be seen from the same overloaded [5]. It would have also been seen on the same table that generator buses in the network and whose voltages values are fixed at 1.0 < 0 per unit falls within the accepted range and therefore not overloaded. Similarly, based on these results, loaded buses such as bus 1 and 2 in the network have their calculated voltages within the permissible range and hence, they are not overloaded. Therefore, there exist no overloaded bus in the network perhaps the distance of the buses from the source of supply are not far.

From the flow pattern shown in Table 4 of the results, it is observed that for a steady state situation, there are no overloaded lines. For example, on Table 4, the active and reactive power flow from bus 2 to bus 3 measured at bus 3 are - 2.021 and 2.821 per unit respectively. While the returned power flow measured at bus 2 are 2.021 and - 2.716 per unit respectively. This implies that there is a negligible active power loss on this line. Furthermore, there is a reduction in reactive power flow when measured at bus 2, which accounts for the losses. However, it is 'unhealthy' of any power system to have higher reactive power flow than active power flow as is the case for bus 2 – bus 3 and bus 1 – bus 2 transmission lines in the network.

Also in Table 4, power flow analysis shows that the slack bus active power is 1.65 per unit. This suggests that for the Practical feasibility of the network as a power system, the rated capacity of the Generator to be installed at bus 1 should be far greater than 1.65 per unit (165MVA).

VI. CONCLUSION

The paper has presented a review of Newton Raphson power flow analysis. A computer software package is developed for Newton- Raphson power flow studies. The package was tested for a three Bus-bar network as case study. The results of this analysis converged at the fifth iteration. The quick convergence of the solutions of the power flow analysis for the network considered in this paper satisfied the theoretical reports of Newton Raphson power flow iterative method. Newton – Raphson approach to power flow solutions was observed to be the most sophisticated and very difficult to program. However, it converges faster than the other methods and the results are often more reliable. Thus, it is worth mentioning that the obtained power flow analysis results are in conformity with the known solution of the Network.

REFERENCES

[1] L. M. Adesina and O. A. Fakolujo,"Power Flow Analysis of Island Business District 33kV Distribution Grid System with Real Network Simulations", International Journal of Engineering Research and Applications (IJERA), vol. 5, issue 7, pp. 46-52, July 2015.

[2] C. M. Wankhade and A. P. Vaidya "Optimal Power Flow Using Genetic Algorithm: Parametric Studies for Selection of Control and State Variables", British Journal of Applied Science & Technology, vol. 4, no. 2, pp. 279 – 301, 2014.

[3] A. Abdulkareem, C. O. A. Awosope, H. E. Orovwode and A. A. Adelakun, "Power flow Analysis of Abule – Egba 33kV Distribution Grid System with real network Simulations", IOSR Journal of Electrical and Electronics

Engineering, ISSN: 2320 – 3331, vol. 9, issue 2, pp. 67-80, Mar – April. 2014.

[4] M. A. Kusekwa, "Load flow Solution of the Tanzanian power network using Newton – Raphson method and MATLAB Software", International Journal of Energy and Power Engineering, ISSN: 2326-957X, vol. 3, no. 6, pp277-286, 2014.

[5] S. S. Meel, S. B. Sspm and M. S. Mohiuddin, "Load Flow Analysis on Statcom Incorporated Interconnected Power System Networks Using Newton - Raphson Method", IOSR Journal of Electrical and Electronics Engineering, e-ISSN: 2278-1676, p-ISSN: 2320-3331, vol. 9, issue 4, pp. 61-68, Jul. – Aug. 2014.

[6] I. A. Adejumobi, G. A. Adepoju and K. A. Hamzat, "Iterative Techniques for Load Flow Study: A Comparative Study for Nigerian 330kVGrid System as a Case Study", International Journal of Engineering and Advance Technology (IJEAT), ISSN: 2249 – 8958, vol. 3, issue 1, pp. 153-158, October 2013.

[7] N. Nayak and A. K. Wadhwani, Performance of Newton – Raphson Techniques in Load Flow Analysis using MATLAB", National Conference on Synergetic Trends in Engineering and Technology, International Journal of Engineering and Technical Research, ISSN: 2321 – 0869, special issue, pp.177 – 180. 2014.

[8] T. Tingting, L. Yang, L. Ying and J. Tong, "The Analysis of the Convergence of Newton – Raphson Method Based on the Current Injection in Distribution Network Case", International journal of Computer and Electrical Engineering, vol. 5., no. 3, pp. 288-290. June 2013.

[9] V. K. Shukla and A. Bhadoria, "Understanding Load Flow Studies by Using PSAT", International Journal of Enhanced Research in Science Technology and Engineering, ISSN: 2319 – 7463, vol. 2,iIssue 6, pp.50 – 57, June 2013.

[10] J. Chakavorty and M. Gupta, "A New Method of Load – Flow Solution of Radial Distribution Networks", International Journal of Electronics and Communication Engineering, ISSN: 0974 – 2166, vol. 5, no. 1, pp.9 – 22, 2012.

[11] C. M. Wankhade and A. P. Vaidya "Optimal Power Flow Using Genetic Algorithm, International Journal of Power System Operation and Energy Management, ISSN: 2231 – 4407, vol. 1, issue4, pp.67 – 72, 2012.

[12] M. Syai'in and A.Soeprijanto, "Improved Algorithm of Newton Raphson Power Flow Using GCC limit based on Neural Network", International Journal of Electrical & Computer Science (IJECS), vol. 12, no. 01, pp.7 – 12, 2012.

[13] P. Kumkratug, "Fast Decoupled Power Flow for Power System with High Voltage Direct Current Transmission Line System", American Journal of Applied Science, ISSN: 1546 - 9239, vol. 7, no, 8, pp.1115 – 1117, 2010.

[14] H. Kim, N. Samann, D. Shin, B. Ko, G. Jang and J. Cha, "A new Concept of Power Flow Analysis", Journal of Electrical Engineering & Technology, vol. 2, no. 3, pp. 312 – 319, 2007.

[15] R. Alqadi and M. Khammash, "An Efficient Parallel Gauss – Seidel Algorithm for the Solution of Load Flow Problems", The International Arab Journal of Information Technology, vol. 4, no. 2, pp.148 – 152, April 2007.

[16] S. B. Warkad, M. K. Khedkar and G. M. Dhole, "A Genetic Algorithm Approach for Solving AC – DC Optimal Power Flow Problem", Journal of Theoretical and Applied Information Technology, pp.27 – 39, 2005 – 2009.

[17] C. O. A Awosope and T. O. Akinbulire, "The Power Systems Analysis of the medium planning of the NEPA 33okV Network: part II – Transmission System Reliability Studies", The Nigerian Engineers, vol. 25, no 3, pp. 55-73, July – Sept. 1990.

[18] C. O. Awosope and A. A. Esan, "The Power Systems Analysis of the medium planning of the NEPA 330kV Network: Part I – Load flow studies", The Nigerian Engineer, vol. 18, no 1, pp. 9-23, January – March 1983.

[19] C. A. Gross, "Power system analysis", John Wiley and Sons, Inc., pp. 61- 25, 1990.

[20] G. T. Heydt, "Computer Analysis Methods for Power Systems", Macmillan Publishing Company, New York and Collier Macmillan Publishers, London, July 2005.

Smart Energy System: The Panacea to Nigeria Ecliptic Power Supply

Sylvester Ajah
Computer Engineering Technology,
Akanu Ibiam Federal Polytechnic Unwana,
Ebonyi State, Nigeria.
ajah.sylvester@gmail.com

Ogbonnia Ibe – Enwo
The Rector,
Akanu Ibiam Federal Polytechnic Unwana,
Ebonyi State, Nigeria
ibeenwo@gmail.com

Abstract – The ecliptic power supply in Nigeria is one of the major banes of the Nigeria industrialization, which consequently lead to very high unemployment rate, low tax revenue, the continual stagnation of Nigeria nation as a third world country, etc. The smart energy system is an efficient and cost effective energy generation, distribution and consumption that are intelligently and flexibly integrated to optimized energy generation, distribution and consumption. To generate sufficient electricity for Nigerians, the governments have to prioritize centralized and decentralized renewable and non – renewable energy sources simultaneously. This work evaluates different ways in which smart energy system paradigm can be the fast and sustainable solution to Nigeria ecliptic power supply. By looking at different ubiquitous cost effective sources through which power can be generated, distributed and used efficiently.

Keywords: *Smart Energy Systems, Smart Grid, Power to Gas, Vehicle to Grid, Nigeria Ecliptic Power Supply, Renewable and Non – Renewable Energy, Sustainable Energy*

I Introduction

Nigeria is a beautiful country, endowed with numerous resources by nature: human resources, mineral resources, weather, arable land, water resources, and any good thing of nature you can imagine. With these resources, the imperative question for any right thinking person is why Nigeria nation is still a third world country with tendency to a failed State since Nigeria got her independence from their Colonial Master(s)? Many accomplished Nigerians, Nigeria organizations, international citizens and international organizations have tried to answer this question in an attempt to aid Nigerians (Nigeria Government and citizens) in solving numerous challenges (insecurity, hunger, inadequate social amenities, etc) confronting them. Many of them said that the major challenge is corruption resulting from ineffective leadership. While we will partly agree with that view, we will rather say as a fundamentalists that the number one enemy of Nigeria / Africa development is spiritualism (Which we call 'Africa Magic / black & white'). Reason being that corruption resulting from ineffective leadership is a by – product of spiritualism as practice in Nigeria / Africa because it enslaves the mind and compromise to a greater extent the logical reasoning of Nigerians. While the spiritualism concept as practiced in Nigeria / Africa will be discussed in a different avenue, smart energy system is core in addressing many challenges in Nigeria.

The engine of life world over is uninterrupted electrical power supply, which is at the core in catalyzing the development of any nation. Stable electricity supply is at the core in ensuring that more industries are working; which aid in mitigating numerous challenges facing Nigeria, thereby leading to a stronger nation. The author in [1] enumerated the history of the electricity industries in Nigeria, the challenges and proposed some solutions which are based on the hierarchical electrical grid that currently exist in Nigeria. Agreed that adopting the proposed solutions will definitely enhance the quality of electricity in Nigeria, but the electricity generating capacity in Nigeria is far low to meet the demand of Nigerians, hence the need for a shift in paradigm to smart energy system in order to meet the energy demands in Nigeria. In figure 1, the current hierarchical structure of the electricity grid in Nigeria is shown.

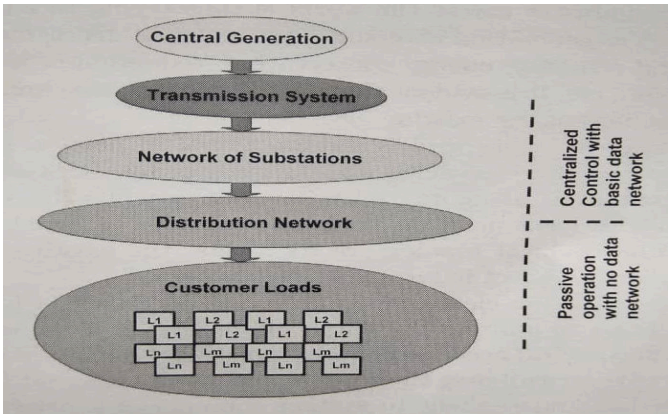


Fig 1: The existing structure of the hierarchical electricity grid in Nigeria [2]

To illustrate the impact of the ecliptic power supply on Nigeria economy, the Nigeria Director General of the National Association of Commerce, Industry, Mines and Agriculture (NACIMA), Ambassador Ayoola Olukanni stated that Nigeria private businesses spend over $14 billion annually on acquisition of power generators, their maintenances and fuelling [3]. If the aforementioned amount is spent by the private businesses annually, imagine the amount that is being spent by private sector, public sector and individuals combined annually on power, also imagine the impact of the accumulative sum will have on the Nigeria economy assuming the Nigeria state have stable electricity.

In addition, the Manufacturers Association of Nigeria (MAN) has stated that the investments in Nigeria are being lost as a result of the continued poor power supply across Nigeria, resulting to companies relocating their manufacturing plants to Ghana with relatively stable electricity in West Africa Region [4]. Also, MAN in [5] stated that poor power supply had remained the bane of the manufacturing sector in Nigeria, they maintained that it was one of the core reasons why the cost of the locally produced goods were beyond the reach of the average Nigerian. No wonder the Nigeria economy is import dependent and there is over 23% unemployment rate, if combined with underemployment, could reach 50% [6]. Based on the reports above, it is now obvious why there is high level of insecurity in Nigeria.

Nigeria currently have over 15000 megawatts installed capacity of electrical power generation plants across Nigeria, consisting of thermal plants with install capacity of over 10,000 megawatts, hydroelectric electric power plant with installed capacity of over 5,000 megawatts and wind farm with installed capacity of 10 megawatts [7][8]. Despite these available power generation plants capacities, only around 7,000 megawatts of power is currently being generated and the power that actually reaches the industries and other final users is around 4,000 megawatts of energy. No wonder Nigeria is now the poverty capital of the world, and the insecurity in Nigeria is worsening daily. Furthermore, they end users of the power distributed hardly pay their electricity bills, especially the public sectors which result to the loss of revenue for the distribution companies due to the absence of prepaid meters in most locations. And the distribution companies allotting exorbitant bills to the few users that pay their bills, while the workers from these distribution companies resorts to all sorts of illegal activities to extort the final users at the expense of the distribution companies. Making power sector in Nigeria unattractive to the private investors, which could have been the driver of the power sector in Nigeria as a result of the big market.

Smart energy system is defined as new technologies and infrastructures that create new forms of flexibility, primarily in the conversion stage of the energy system. This definition entails merging electricity, thermal and transport sectors so that the flexibility across these different areas can compensate for the absence of flexibility renewable energy resources like wind and solar [9]. As opposed to smart grid concept that takes sole focus on the electricity sector, the smart energy system include the entire energy system in its approach to identifying suitable energy infrastructure designs and operation strategies. The smart grid only focus on the electricity sector usually lead to the conclusion that transmission lines, electricity storage and resilience electricity demands are the main technique to deal with the integration of the unstable renewable power sources. But the nature of wind, solar and other similar sources have the consequence that these measures are not effective and cost – efficient. The most effective and cost – efficient solution is combining the electricity sector with the transportation, cooling and heating sectors simultaneously [10].

Smart grid on the other hand is the integral part of smart energy system. The smart grid involves the application of advanced electrical engineering and services technologies, facilitated by the application of ICT (information and communications technology) and concomitant solutions to more effectively and efficiently managed complex infrastructure investments [11]. While [12] stated that the

smart grid are the technologies that observes the state of power system and intelligently take decisions to quickly clear faults, restores power and monitor demand to preserve stability and performance of the electric power network formerly done by engineers. Smart grid is explained by the author as comprise a broad and evolving range of advance technologies that can be applied along the full electricity supply chain – from generation, via transmission, distribution and metering to end users. The smart grid is also explained as the future for the development of future generation, transmission and distribution networks that is highly flexible, reliable and sustainable. The smart grid concept encompasses the integration of renewable energy sources; application of innovative technologies; use of communication technologies to improve observability and controllability of networks; development of intelligent applications, protection, and automation concepts; high security of supply and overall suitable network performances; and design of new network structure like micro-grids, DC networks, and overlay transmission grids [13].

This work is organized as follows; section II discusses smart grid on sustainable energy generations, transmissions and distributions as practiced in the developed world in terms of centralized and distributed energy systems, section III discussed how to transit from the existing silo hierarchical grid structure to smart grid, section IV look at Smart Energy System for Nigeria Stable Power, while section V concludes the work and makes suggestions on how Nigeria can transit to a nation with stable power .

II Smart Grid

Smart grid paradigm had briefly been discussed earlier in the introduction part of this paper, but for emphasis sake; smart grid concept is a consolidation of all technologies, concepts solutions and methodologies that allow the isolated hierarchies of generation, transmission and distribution to be replaced with an end – to – end organically intelligent and fully integrated environment where business processes, objectives and needs of all stake holders through efficient exchange of data, services and transactions. This is achieved through the use of smart sensors for faults detection, reporting and preventions via two way communications. It empowers the consumer to interact with the energy management systems to manage their energy use and reduce their energy costs [2]. Figure 2 shows the smart grid, while table 1 shows the difference between the existing silo hierarchical grid system and smart grid.

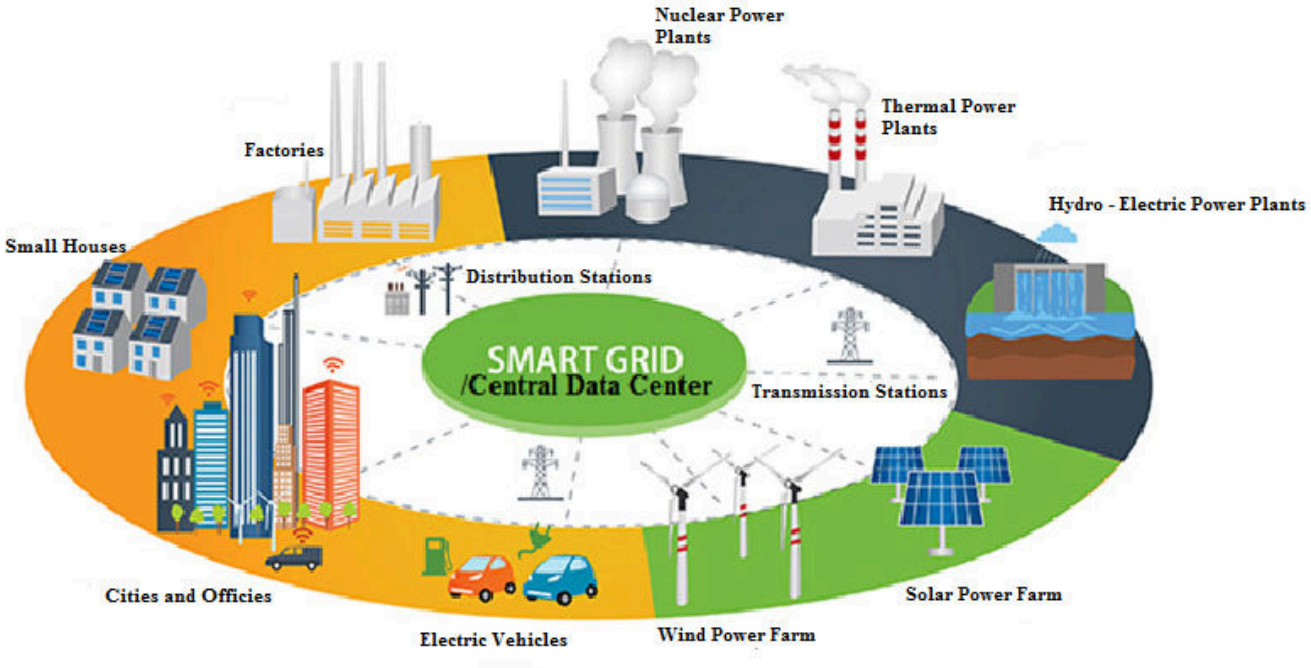


Fig 2: Smart grid Paradigm

Table 1: Comparison of Existing Grid and Smart Grid [2]

Existing Grid	Smart Grid
One – way Communication	Two – way communications
Electromechanical	Digital
Centralized Generation	Distributed Generation
Few Sensors	Sensors throughout
Hierarchical	Network
Blind	Self – monitoring
Manual Restoration	Self – healing
Failure – blackouts	Adaptive and Islanding
Manual check / test	Remote check / test
Limited control	Ubiquitous control
Few customer choices	Many customer choices

Many incidents in smart grid are local in nature; hence have to be solved locally via in-built localized intelligence. Though, higher systems and applications may need to be informed about the occurrence of such incidents and the local action taken to attend to those incidences, but the higher systems and applications do not have to assume the urgent and critical control in implementing remedies to local incidences.

III Transition from Existing Grid to Smart Grid

Now that we have gotten grip of the features of smart grid paradigm, and the differences between the existing power grid structure in Nigeria and smart grid, the question is, how can the existing grid structure be transformed to a smart grid? The answer to this question is that smart grid should not be a disruptive power grid technology (that is, replacing the existing silo hierarchical grid structure) but complementary to the existing grid structure, adding to its capabilities, functionalities and abilities via an evolutionary path. Hence, this necessitate a smart grid topology that allows for organic growth, inclusive forward looking technologies and full compatible with the existing grid structure. At the core of the transition from the existing grid structure to smart grid is through an ad-hoc integration of complementary components, subsystems and functions under ubiquitous control of highly intelligent, distributed commands and control systems. In addition, smart grid evolution is anticipated to necessitate plug and play smart Micro-grids. Micro-grids are interconnected network of distributed energy systems that can operate within or outside the existing grid [2]. The smartness of the Micro – grids is a function of the level of close – loop intelligence that allows the needed level of automation, energy management and protection to be built into the system.

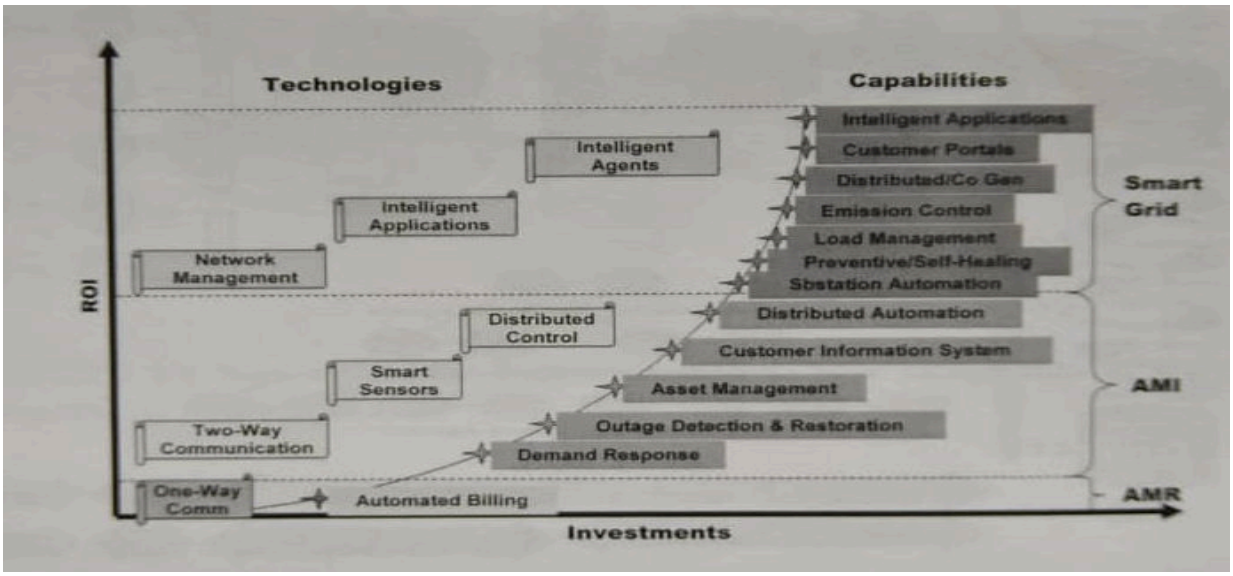


Fig 3: Transition for existing grid to smart grid [2]

Figure 3 shows the steps in transiting from the existing grid structure to smart grid. The horizontal axis is the required investment to be made, while the vertical axis is the expected return on the investment (ROI). Transiting to smart grid paradigm in Nigeria is very good in ensuring uninterrupted power supply which is the engine of development for any nation. But the power generation capacity in Nigeria is very low; hence canvassing for this transition currently seems unattainable, even though there is nothing impossible when there is the will. The focus now should be removing all the bureaucratic bottlenecks to ensure foreign direct investment into the power generation being that the Government of Nigeria claim not to have sufficient fund to invest in the power generation that will be sufficient for her citizens. Though, smart grid being a more sustainable solution should be the mission of the new investments in the grid.

Many authors have canvassed for the government investments into new generation plants to be able to meet the power need of her citizens, we will partly agree that it is the solution, based on the current economic situation of Nigeria government, it can be deduced from the reviewed literatures that the most effective solution to enhance power generation in Nigeria is that the Nigeria government should try to revive, repair and optimize all of the already existing installed power plants in Nigeria to ensure that they generate to their full installed capacities adding ad-hoc smart sensors network that will enhance the intelligence in ensuring their optimal performances. Followed by reviewing and optimizing the power lines with ICT based devices for monitoring and maintenance of the power equipments to ensure that the generated power reaches the end users. Coupled with the end users using energy efficient devices, to conserve the distributed-energy.

IV Smart Energy System for Nigeria Stable Power

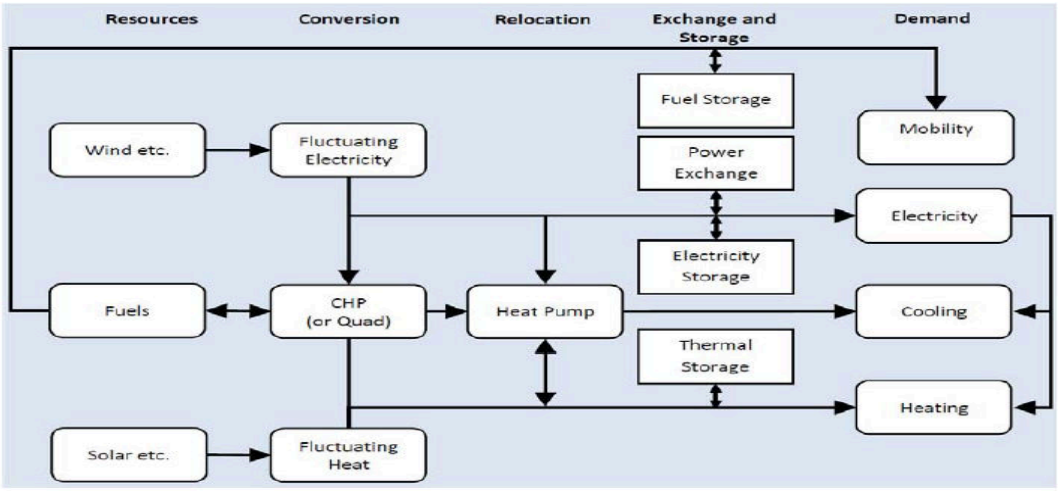


Fig 4: Smart Energy System [9]

The smart energy system as explained in the introduction part of this paper is a global paradigm shift into more sustainable and greener energy as shown in figure 4. From the global perspective, the smart energy system encompasses the following energy sub – systems: Vehicle – to – Grid (V2G) subsystem; The process in which the car batteries especially electric vehicles (EVs) are used to store the renewable energy sources during the peak period and returned back to the grid during the off – peak periods [14. While this concept is ideal in the developed nations with required infrastructures, it may not immediately be used in Nigeria because low / lack of renewable energy infrastructure, sources and EVs with required battery capacities to store sufficient energy that can make impact on the grid. Figure 5 shows the concept of V2G subsystem.

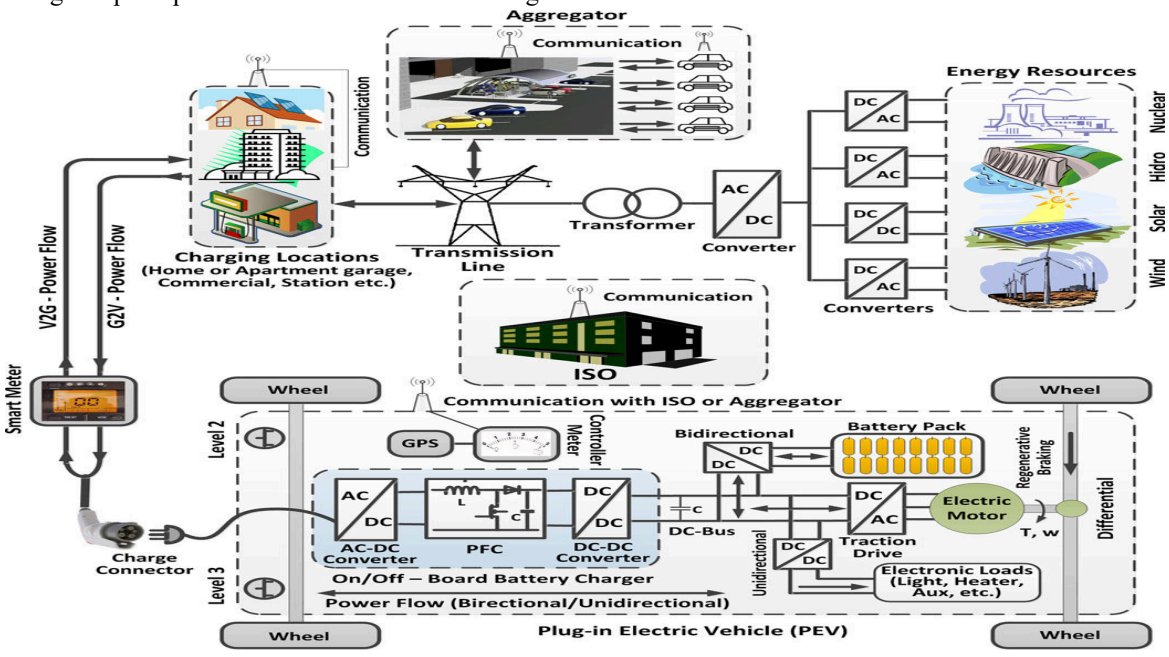


Fig 5: Vehicle – to – Grid Concept [14]

From the global perspective, another energy sub – system that made up of smart energy system is 4th generation district heating (4GDH). This is a comprehensive technological and institutional concept, through which smart thermal grids aids

the development of sustainable energy systems. It provides the heat supply of low energy buildings with minimal grid losses in a way that the use of low – temperature heat sources are blended with the operation of smart energy systems [15]. While this sub – energy system may sound unnecessary for those in the southern part of Nigeria with moderate climate / weather conditions, it is very useful for those in the Northern part of Nigeria with extreme weather conditions at different period of the year. Figure 6 shows the schematic diagram of 4GDH energy sub – system.

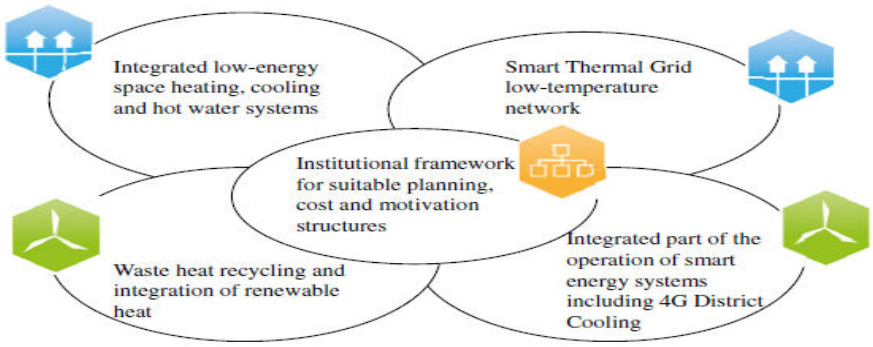


Fig 6: 4GDH Concept including smart thermal Grid [16]

Another energy sub – system that made up smart energy system paradigm is power to gas.

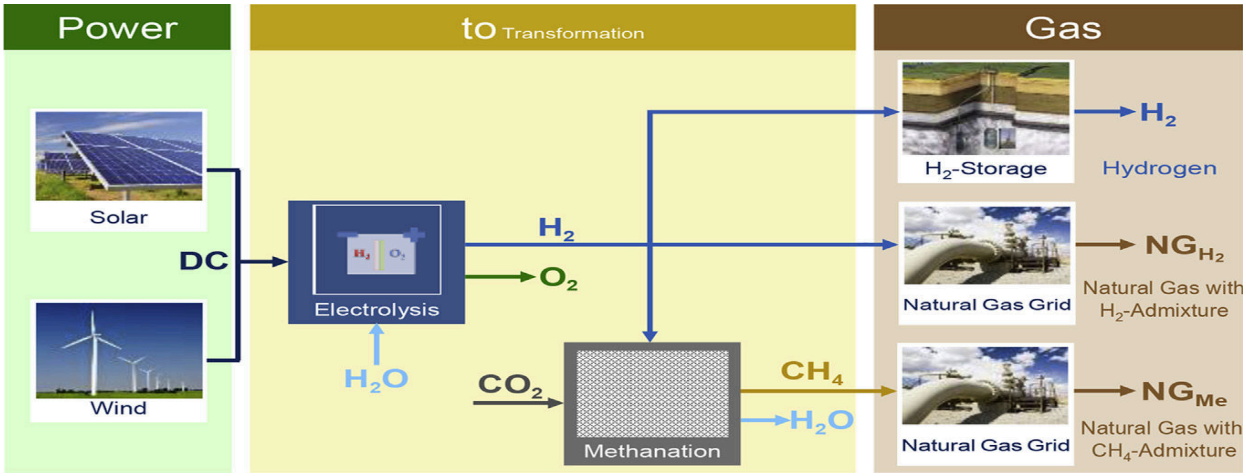


Fig 7: Power to Gas process [17]

Figure 7 shows the mechanism of power to gas process. It is applied to enable large scale storage of renewable energy sources during the peak period to save the excess power, which will compensate the short time or seasonal imbalance of renewable energy sources. It is achieved by converting electricity into hydrogen through a process called electrolysis. The hydrogen that is generated can serve as a link combining the electricity, traffic and heating sector in one market [17]. This process may seem farfetched by the level of energy sector development in Nigeria, but being futuristic, it will be good to draw a blue print of the energy sector in Nigeria toward smart energy system, because that is the global sustainable energy model.

V Conclusion

If Nigerians are serious about changing the sorry state of Nigeria nation, a concerted effort is required from all stakeholders in ensuring that there is a stable uninterrupted power supply in Nigeria. Take to piece numerous challenges confronting Nigeria state, stable electric power is at the core in mitigating those challenges. The smart energy paradigm can be started in Nigeria by various states and federal

governments of Nigeria drafting and enforcing polices that will enhance smart energy system. The federal government of Nigeria should decentralize the power sector, by granting permits to various states that can generate and distribute their own power, thereby removing excess loads on the limited power generated and distributed across Nigeria. The government of Nigeria should enforce distribution of prepaid meters to Nigerians, in that way, the distribution companies will know that the revenue they will generate is function of power they distribute, making them responsible stakeholder. Not the current exploitation of Nigerians by the distribution companies irrespective of the power they distribute, making them irresponsible stakeholders.

V. References

[1]. **A., Claudius A.** "*Nigeria Electricity Industry: Issues, Challenges and Solutions",* Ota, Ogun State, Nigeria, Convenant University, 2014.

[2]. **Jinsong, W., Sundeep, R. and Honggang, Z.** "*Green Communications: Theoretical Fundamentals, Algorithms and Applications",* Florida : CRC Press, 2013. 978-1-4665-0107-2.

[3]. **Okonji, Chares.,** "Businesses Spend N5trn Annually on Power Generation - NACCIMA Boss. *The Nation Newspaper"* [Online] The Nation Newspaper, July 21, 2019. [Cited: July 21, 2019.] https://thenationonlineng.net/businesses-spend-n5trn-annually-on-power-generation-naccima-boss/.

[4]. **Manufacturers Association of Nigeria (MAN),** "MAN Says Poor Power Supply Affacting Investments in Nigeria's Manufacturing Sector", *Energy Mix Report.* [Online] Energy Mix Report, June 7, 2019. [Cited: August 21, 2019.] https://www.energymixreport.com/man-says-poor-power-supply-affecting-investments-in-nigerias-manufacturing-sector/.

[5]. **MAN,** "Poor Energy Bane of Manufacturing in Nigeria - MAN", *Nigeria Tribune Newspaper.* [Online] Nigeria Tribune Newspaper, September 24, 2017. [Cited: August 21, 2019.] https://tribuneonlineng.com/poor-energy-bane-manufacturing-nigeria-%e2%80%95man/.

[6]. **National Bureau of Statistics (NBS),** "*Labour Force Statistics - Volume I: Unemployment and Underemployment Report",* National Breau of Statistics (NBS), 2018.

[7]. **Daminabo, I., Aloni, C. and Alexander, B. C., "***State of Power Supply in Nigeria, The Way Out"*, Port Harcourt , International Journal of Development and Sustainability, 2018 , Vol. 7.

[8]. **Habibu, U. A., "**Katsina Wind Farm: N4.4bn Down the Drain, Yet Zero Electricity", *Daily Trust Newspaper.* [Online] Daily Trust Newspaper, April 28, 2019. [Cited: August 21, 2019.] https://www.dailytrust.com.ng/katsina-wind-farm-n4-4bn-down-the-drain-yet-zero-electricity.html.

[9]. **Herik, L.; Poul, A. O.; David, C.; Brian, V. M.,** "*Smart Energy and Smart Energy Systems",* Elsevier, 2017.

[10]. **Herik, L., Brian, V. M., David, C. and Poul, A. O. , "***Renewable Energy Systems - A Smart Energy Systems Approach to the Choice and Modelling of 100% Renewable Solutions"* ISBN 978-88-95608-30-3, Chemical Engineering Transactions, 2014, Vol. 39. ISSN 2283-9216.

[11]. **Fortfas,** "*Priority Area K Smart Grids and Smart City Development Action Plan",* Forfas, 2013.

[12]. **Tsado, J., Imoru, O. and Segun, O. D., "***Power System Stability Enhancement through Smart Grid Technolopgies with DRS"*, 2012, International Journal of Engineering and Technology, Vol. 2, pp. 621 - 629.

[13]. **Siemens, "***Smart Grid Network Concepts",* Erlangen, Germany, Siemens AG and Siemens Industry Inc, 2014.

[14]. **Murat, Y. and Philip, T. K., "***Review of the Impact of Vehicle - to - Grid Technologies on Distribution Systems and Utility Interfaces"*, Illinios : IEEE Transactions on Power Electronics, 2013, Vol. 28.

[15]. **Peter, J.,** "4GDH Definition", *4dh.eu.* [Online] 4dh.eu, 2012 - 2018. [Cited: July 15, 2019.] https://www.4dh.eu/about-4dh/4gdh-definition.

[16]. **Henrik, L.; Sven, W.; Robin, W.; Svend, S.; Jan, E. T.; Frede, H.; Brain, V. M.**, "*4th Generation District Heating (4GDH) Integrating Smart Thermal Grids into Sustainable Energy Systems"*, Energy 68, Elsevier Limited, 2014, Vol. 1.

[17]. **Sebastian, S.; Thomas, G.; Martin, R.; Vanessa, T.; Bhunesh, K.; Detlef, S.**, "*Power to gas: Tecnological Overview, systems analysis and economic assessment for a case study in Germany",* Elsevier, 2015, Vol. 1.

Polynomial Reduction of TSP to Freely Open-loop TSP

[1]Muhammad Bashir Abdulrazaq, [2]Yusuf Suleiman Tahir, [1]Suleiman Sha'aban, [3]Muhammed Sani Jibia
[1]Department of Computer Engineering, Ahmadu Bello University Zaria, Nigeria
[2]Department of Electrical Engineering, Eastern MediteraneanUniversity, North Cyprus.
[3]Department of Communications Engineering, Ahmadu Bello University Zaria, Nigeria
Emails: mbashiray@yahoo.com, yutaone1@gmail.com, shaabansuleiman@gmail.com, muhammedengrjibia@gmail.com.

***Abstract* – Travelling Salesman Problem (TSP) is one of the earliest combinatorial problem that is identified to be NP-hard problem. It is a problem that seeks to find the shortest possible route in a graph problem which passes through all nodes only once and return to the starting point. A variant of TSP is the Freely Open-loop TSP (FOTSP) which seeks to find the shortest route in the graph without having to return to starting point and with no specific starting or end node. In this paper, a reduction of polynomial complexity for TSP problem into FOTSP is presented and vice versa. This reduction proves that FOTSP is also NP-complete just as TSP.**

***Keywords* – FOTSP, NP-complete, NP-hard, Polynomial Reduction, Travelling Salesman Problem.**

I. INTRODUCTION

The Travelling Salesman Problem (TSP) is a combinatorial problem which has application in a lot of areas of life. It is the problem of a salesman who wish to deliver goods to his costumers located at different locations scattered around. The aim of the solution is to find a route that has the shortest length (weight) while visiting each of the costumers (nodes) only once and returning to the starting point [1].

The surest way to find the route with the shortest distance in TSP is to try all possible routes, find their distances and determine the one that has the shortest route. This approach is the brute force approach. In a fully connected TSP (when all nodes are connected to every other node by a direct path/edge), where the path from one node to another is not necessarily the same as the return path, the number of possible routes is N!. This scenario is called the asymmetric TSP (aTSP). If the distance (weight) of the path from one node to another is the same as the return path, the number of possible routes is (N-1)!. This is the symmetric TSP (sTSP or simply TSP) scenario [2].

Brute force becomes unbearable, even with high computing capacity, when the number of nodes reach a few tens. For just 20 nodes, there are 121,645,100,408,832,000 possible routes in a fully connected symmetric case. A computing device capable of testing one route per clock cycle at 3GHz rate will take more than 469 days to complete the tests. This necessitates the search for other smarter means of finding a solution or an approximate solution.

Another version of TSP asks the question of, "Given a route weight, say L, and a connected graph, whether there is a TSP route in the graph which has a weight less than L?" This is the decision version of TSP and is a subject of a lot of research because of its relationship to other combinatorial problems [2]. It proves to be difficult to deterministically solve, but when given the answer, it is very easy to verify that the route weight is actually less than L.

This kind of problems such as TSP are termed NP-hard problems. They are problems that cannot be solved by any known algorithm using a deterministic Turing Machine (TM) but once a solution is provided it is easy to verify [3]. This means TSP is not easier than any other NP-hard problem, and therefore, a solution to it implies a solution to other NP-hard problems exist. This is proved through demonstrating that a problem, such as TSP, can be reduced using an algorithm of polynomial complexity to another known NP-complete problem[4].

The various real life applications of TSP also make it popular. Logistic supply, plane route planning, microchip manufacturing, circuit board drilling, DNA sequencing and many others are areas in which solution to TSP is used to minimise cost of providing the services or determine solutions [2] [5] [6].

Regular TSP where return to origin point is desired is termed closed loop TSP. Some other cases may not need the return to the origin. Example is when a delivery firm hire a van that is going to do deliveries, and the cost of travelling from the depot up to the last costumer is borne by the delivery firm. In this case there is a fixed starting point (the depot), but it does not matter where it ends. All that matter is to minimise the journey during delivery. This case of TSP is therefore an Open-loop TSP (OTSP). Sometimes the final destination is also fixed even though it does not have to return to base [5] [7].

Another scenario is when the beginning and the end does not matter. This is Freely Open-loop TSP (FOTSP). An example is when drilling circuit board. It does not matter where the machine starts or ends. All that matter is minimisation of distance which translates to energy savings. Other variants of TSP exist which include n-OTSP where the traveller is constrained to visit only n nodes which is less than the total nodes [7].

FOTSP received little attention from researchers as such there is no proof known to us that it is NP-hard or there is some algorithm for solving it. In this paper a reduction of FOTSP into TSP will be presented which will serve as a proof that FOTSP is NP-complete.

The rest of the paper is arranged as follows. Section II gives the theoretical background of TSP, its complexity and some available solutions to TSP and FOTSP. Section III describes the method of reducing TSP to FOTSP. Section IV gives an example of the procedure. Section V discusses the complexity and implication of the procedure. Lastly, Section VI concludes the paper.

II. THEORETICAL BACKGROUND

A. *Complexity and TSP*

In theory of complexity, problems that are easy to solve and easy to verify a given solution are categorised as P class of complexity. Easy here means the complexity of the solution scales polynomially (P) as the size of the problem increases [3].

There is another class of problems which are difficult to solve. But when a solution is provided, it is easy to verify that it is a valid solution. These problems are mostly combinatorial search problems. The size of the search space normally increases exponentially with respect to size of the problem. When there is no known polynomial algorithm to solve the problem using deterministic TM, but a solution is easily verifiable in polynomial time, the problem is said to be NP-hard (N for Non-deterministic) [3] [8].

There are dozens of problems that are proved to be NP-hard, among them are: Boolean Satisfiability, Hamiltonian Circuit problem, TSP and many others. An NP-hard problem that is confirmed to be at least as hard as any other NP-hard problem is termed NP-complete. This confirmation is done by reducing the NP problem to another NP-complete problem using an algorithm of polynomial complexity. The decision version of TSP is proved to be NP-complete through similar technique[4].

It is still the subject of research whether NP problems are really never solvable in polynomial time using determinist TM or there are some unknown techniques for doing such. This problem is coined the P vs NP problem. Many attempts were made to find efficient solution to NP problems or solution to P vs NP problem [9]. Some solutions available for TSP are discussed in the subsection ahead.

B. *Solving TSP*

The attempt to have efficient solution to TSP can be categorised into three basic categories. One is to have an exact solution to the problem more efficiently, the second is to have approximates, while the third category is the use of heuristic algorithms solutions for a more efficient algorithm [9].

Example of exact solution is the branch and bound algorithm. It is a dynamic programming technique that reduces the complexity of TSP solution to $N^2 2^N$ [10]. Though still exponential, this is a huge improvement over brute force. It uses bounds to discard any branch of the solution space that has a weight larger than the global or branch bound or when it is not feasible. Different methods exist for choosing the first global bound.

Among the early solutions to TSP is the Held-Karp algorithm. It was presented in 1962 as an exact algorithm that solves TSP with computational complexity of $O(N^2 2^N)$ and memory requirement of the order $O(N2^N)$. This was achieved by a dynamic programming method which splits the problem into smaller sub-problems. It advances the solution to bigger sub-groups recursively and whenever a smaller sub-problem is required, it is recalled from memory [11].

Kruskal algorithm is used in an approximation algorithm to create upper bound through the use of minimum spanning tree (MST). The MST is created by arranging the weights of all the edges in ascending order, then an edge is picked from the smallest upward provided it does not create a cycle. Twice the total weight of the MST is the upper bound used with a dynamic programming technique to replace some of the edges until a valid circuit is formed. The complexity of such algorithm is of the order $O(E\log(N))$, where E is the number of edges and N is the number of nodes. In fully connected situation, $E \alpha N^2$ [12].

Among the approximate algorithms is the nearest neighbour algorithm. Nearest Neighbour is a very quick greedy algorithm in which the salesman starts from any city and visits the city with smallest distance among the cities not visited yet. He continues that until all the cities are visited, after which it returns to the starting point [13]. It does not normally produce optimal result, but it is a very quick way of getting a short route.

Metaheuristic algorithms use techniques that mimic nature to find solution to search problems. They provide good solutions with low complexities. Among the popular ones is the Ant Colony Optimisation (ACO). ACO was used to solve TSP in [14]. The algorithm mimics the behaviour of ants when seeking the best path. Those who traversed the path before drop some pheromone for others that will follow to know the path already traversed by others. Experimental tests show it can achieve 99% accuracy with a complexity of the order $O(N^4)$.

Genetic algorithm is another metaheuristic algorithm used in solving TSP. It uses the way genes mutate and go through selection of the fittest [15].

In general, as shown above, the exact algorithms are exponential, while the approximates do not guarantee optimal result.

C. *FOTSP*

2-Opt, 3-Opt etc. provide approximate solutions to TSP. In 2-Opt, two edges are removed and alternative edges, which can replace them with lower weight, are used to replace them. This continues until there is no more replaceable two edges. Some algorithms for implementing 2-Opt can have complexity of about $O(N^3)$. This makes it suitable as a good heuristic algorithm and gives very good results especially in euclidean cases, where the distances obey triangular inequality.

In [16] it was shown that 1-Opt can be used to optimise FOTSP with a complexity of the order $O(N^2)$. This is because in 1-Opt, only one edge is replaced by another. Therefore, a method is presented in this paper with which TSP can be converted to FOTSP. The conversion has complexity $O(N^2)$ and hence TSP can be optimised using 1-Opt. The resultant complexity is therefore, $O(N^2)$ which is better than 2-Opt. The conversion also serves as the proof that FOTSP is NP-complete.

III. Reduction of TSP to FOTSP

A. *Eliminating 1 Node*

TSP is generally represented in the form of adjacency matrix. This is a matrix that displays the weight of path between any two nodes in a row and column corresponding to the nodes. Fig. 1 Shows example of an adjacency matrix.

	A	B	C	D
A	∞	$w_{1,2}$	$w_{1,3}$	$w_{1,4}$
B	$w_{2,1}$	∞	$w_{2,3}$	$w_{2,4}$
C	$w_{3,1}$	$w_{3,2}$	∞	$w_{3,4}$
D	$w_{4,1}$	$w_{4,2}$	$w_{4,3}$	$\infty_{4,4}$

Fig. 1: An Example of Adjacency Matrix

In Fig. 1, A, B, C and D are the nodes. $w_{1,2}$, for example, is the weight of the edge between A and B. In TSP, because it is a condition that the traveller can not move from a node to the same node, the distance of, say node A to A is written as infinity instead of zero.

If a constant is added to all elements of a row or to all elements of a column in an adjacency matrix of TSP, the optimal route of the original matrix remains the same as the optimal route of the new resultant matrix.

In the light of the above, we can equalise the weight of all edges from a particular node to all others by adding a given number to each column. Also we can equalise the weight of edges from other nodes to a particular node by similar additions. After these operations, the path with the minimum weight will still remain the same as before the operation.

If these equalisations are done, then it does not matter from which node you approach the particular node or to which node you move from that particular node, the weight effect of that node will be the same. This means the node, if removed, will leave an open edge in the routes which can start or end at any node. Therefore, the optimal FOTSP route of the graph will form the optimal TSP route of the graph if the equalised node is put back.

As a result, equalisation of an N node TSP problem can be reduced to an N-1 FOTSP problem through elimination of the equalised node. The steps needed in this reduction can be summarised and generalised as follows:

a) Let the minimum weight allowed on any edge be w_{min}.
b) Subtract the weights in the row of node to be removed, from the columns in which they appear except the column with ∞.
c) Subtract the weights in the column of node to be removed, from the rows in which they appear except the row with ∞.
d) Find the minimum weight of the new matrix as w_{MIN}. Then add w_{min}-w_{MIN} to all the elements of the matrix.
e) Remove the equalised row and column.

Then the resultant matrix will have an N-1 optimal FOTSP route which if the removed node is placed back in its open edge, an N optimal TSP route will be formed. In other words, solving the FOTSP gives solution to the TSP.

The example given in the next section gives more clarity to the reduction of TSP to FOTSP.

IV. Example of TSP to FOTSP Reduction

Fig. 2 shows a symmetric TSP adjacency matrix showing the weight of going from all nodes to other nodes.

$$\begin{pmatrix} \infty & 12 & 3 & 7 \\ 12 & \infty & 13 & 2 \\ 3 & 13 & \infty & 9 \\ 7 & 2 & 9 & \infty \end{pmatrix}$$

Fig. 2: Example Adjacency Matrix

$$\begin{pmatrix} \infty & 0 & 0 & 0 \\ 12 & \infty & 10 & -5 \\ 3 & 1 & \infty & 2 \\ 7 & -10 & 6 & \infty \end{pmatrix}$$ (a)

$$\begin{pmatrix} \infty & 0 & 0 & 0 \\ 0 & \infty & -2 & -17 \\ 0 & -2 & \infty & -1 \\ 0 & -17 & -1 & \infty \end{pmatrix}$$ (b)

$$\begin{pmatrix} \infty & 18 & 18 & 18 \\ 18 & \infty & 16 & 1 \\ 18 & 16 & \infty & 17 \\ 18 & 1 & 17 & \infty \end{pmatrix}$$ (c)

$$\begin{pmatrix} \infty & 16 & 1 \\ 16 & \infty & 17 \\ 1 & 17 & \infty \end{pmatrix}$$ (d)

Fig. 3: Example Reduction of TSP to FOTSP
(a) Row equalisation (b) Column equalisation
(c) Normalisation (d) Reduction

a) Let us assume the minimum weight allowed is $w_{min} = 1$.
b) To equalise first node, subtract 12 from 2nd column, 3 from 3rd column and 7 from 4th column (Fig. 3(a)).

c) Then subtract 12 from 2nd row, 3 from 3rd row and 7 from 4th row as shown in Fig. 3(b).
d) It can be seen that the minimum value from elements not on the equalised row and column is $w_{MIN} = -17$. Therefore, we will add $w_{min} - w_{MIN} = 1 - (-17) = 18$ to all elements to have Fig. 3(c).
e) The equalised row and column are removed, to get size 3 FOTSP (Fig. 3(d)). Its FOTSP solution gives solution to the size 4 TSP through insertion of the equalised node in the open edge of the FOTSP optimal route.

V. Discussion

Whichever position the removed node happens to be placed, it will add up the same value to the route. This is because all the weights to and from the node are the same. This means it can as well be placed in between the nodes at the open end in order to close the route. Any open-loop TSP route in the reduced problem can be closed with the removed node without affecting its rank. This means if the FOTSP route with the smallest weight is closed, it will remain the TSP route with the smallest weight. Therefore, optimal FOTSP route of the reduced case is automatically the optimal TSP route of the non-reduced TSP case.

The computational complexity of the reduction is at most quadratic. This is because it involves additions and subtractions to the elements whose number is N(N-1). It also involves finding the minimum in the reduced area whose number is (N-1)(N-2). The space complexity is also at most quadratic. That is the memory required to save the elements of the matrix.

This concept that a node with all equal weights to and from any node can convert a FOTSP problem to a TSP problem can be used as a procedure for the reverse conversion also. In that case, we can say the reduction procedure to convert FOTSP to TSP has linear complexity. This is because it involves just adding 2N equal weights.

VI. Conclusion

In this paper we have discussed a procedure of polynomial complexity (quadratic) for reduction of TSP to FOTSP. We have also seen that the concept can be used to convert FOTSP to TSP with linear complexity. This is, therefore, a proof that FOTSP is an NP-complete combinatorial problem.

References

[1] Marek Karpinski, Michael Lampis, Richard Schmied, "New Inapproximability bounds for TSP," Journal of Computer and System Sciences, vol. 81, no. 8, pp. 1665-1677, 2015.
[2] Ilavarasi K., Suresh Joseph K., "Variants of Travelling Salesman Problem: A Survey," 2014 International Conference on Information Communication and Embedded Systems (ICICES-2014), February 2014.
[3] Stephen A. Cook, "The Complexity of Theorem Proving," Proceedings of the Third Annual ACM on Theory of Computing, pp. 151-158, May 1971.
[4] Alan Cobham, "The Intrinsic Computational Difficulty of Functions," The Journal of Symbolic Logic, vol. 34, no. 4, pp. 657, February 1970.
[5] Akshatha P. S., Vasudha Vashisht, Tanupriya Choudhury, "Open Loop Travelling Salesman Problem using Genetic Algorithm," International Journal of Innovative Research in Computer and Communication Engineering, vol. 1, no. 1, pp. 112-116, March 2013.
[6] Saloni Gupta, Poonam Panwar, "Solving Travelling Salesman Problem Using Genetic Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no 6, pp 376-380, June 2013.
[7] Hock Hung Chieng, Noorhaniza Wahid, "Performance Comparison of Genetic Algorithm's Mutation Operators in n-Cities Open Loop Travelling Salesman Problem," First Ineternational Conference on Soft Computing and Data Mining (SCDM-2014), pp 89-97, June 2014.
[8] Eric Allender, Peter Burgisser, Johan Kjeldgaard-Pedersen, Peter Bro Miltersen, "On Complexity of Numerical Analysis" SIAM J, Comput, vol. 38, no. 5, pp. 1987-2006, 2007.
[9] Lance Fortnow, "The Golden Ticket: P, NP and the Search for the Impossible," Princeton; New Jersey: Princeton University Press, ISBN 9780691156491, 2013.
[10] Mirta Mataija, Mirjana Rakamaric Segic, Franciska Jozic, "Solving Travelling Salesman Problem Using Branch and Bound Method," Zbomik Veleucilista u Rijeci, vol. 4, no 1, pp. 259-270, 2016.
[11] Michael Held, Richard M. Karp, "A Dynamic Programming Treatment of the Travelling Salesman Problem," Journal of Association of Computing Machines, vol. 9, 1962.
[12] Joseph B. Kruskal Jr., "On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem," Proceedings of the American Mathematical Society, vol. 7, pp. 48-50, February 1956.
[13] Gregory Gutin, Anders Yeo, Alexey Zverovich, "Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP," Discrete Applied Mathematics ,vol. 117, pp. 81–86, 2002.
[14] Marco Dorigo, Luca Maria Gambardella, "Ant colonies for the travelling salesman problem," Journal of BioSystems, vol. 43, pp. 73–81, 1997.
[15] Shamima Akter, Nazmun Nahar, Mohammad Shahadat Hossain, Karl Anderson, "A New Crossover Technique to Improve Genetic Algorithm and its Application to TSP," 2019 International Conference on Electrical Computer and Communication Engineering (ECCE-2019), February 2019.
[16] Muhammad B. Abdulrazaq, Aminu Jafar A., Yusuf Abdul-Razaq, "1-Opt Optimisation for Freely Open-loop Travelling Salesman Problem," Bayero Journal of Engineering and Technology, vol. 14, no. 2, 2019 pp 123-127.

Design, Construction and Testing of an Electronic Hospital Queuing Solution System

U.I. Bature[1*], Abbas Musa Hassan[1], N. M. Tahir[2], Ajekiigbe Enoch Seun[1], A. Y. Nasir[1], U. S. Toro[1]
[1]*Deptment of Computer and Communications Engineering, Abubakar Tafawa Balewa UniversityBauchi (ATBU)*
P.M.B 0248 Bauchi, Nigeria
[2]*Department of Mechatronics and System Engineering,Abubakar Tafawa Balewa University Bauchi (ATBU)*
P.M.B 0248 Bauchi, Nigeria
Corresponding Author: email address: *biusman@atbu.edu.ng

***Abstract*— One of the major challenges for healthcare services all over the world is the queuing process, especially in developing countries. Application of queuing theory enhance decision making and, queuing theory is not commonly used by managers in developing countries in contrast to their counterparts in the developed world. In many hospitals, patients have to wait for a long period in the healthcare facility before they will be attended to by the health personnel. Therefore, there is a need for a portable and smart queue resolution system. This project presents the Design, Construction and Testing of an Electronic Hospital Queuing Solution System, the proposed system operates on an input/output basis; centered on a programmed ATMEGA328p microcontroller. The system receives the input from the client on the device through the use of a Radio Frequency Identification (RFID) card reader which reads the card of the patient and therefore a number will be assigned to the patient, whereby he/she will proceed to the appropriate assigned seat. When a doctor is available, a receptionist will then press the button of the next available doctor and this action then triggers a record of the next patient on the queue by calling it through a speaker and displaying it on the Liquid Crystal Display (LCD). This system was aimed at maximizing the Doctors competence and enhancing quality of their service and, the use of RFID card reader will minimize the waiting time and improves the queuing process.**

[1] ***Keywords—electronic system, hospital queue, solution system, RFID Card***

I. INTRODUCTION

Acute care is the treatment of severe medical conditions for only a short period and at a crisis level. Many hospitals are acute care facilities which mean there are no adequate doctors to attend to patients who require proper medical attention. Therefore, hospitality was not only given to the traveler but, it can also be given to the sick, result shows that hospital patients sometimes left feeling unsatisfied despite receiving the necessary medical treatment [1].

Queuing has been the major challenge in almost all healthcare systems. In the developed world, a significant amount of research has been carried out on how to improve queuing systems in various hospital settings. This, unfortunately, has never been the case in developing countries such as Nigeria. This paper tries to contribute to this topic by analyzing the queuing situation in government hospitals in Nigeria and also to bring its practical value as to how the process of decision making can be enhanced in hospitals around the world. Queueing theory is defined as the mathematical study of waiting for lines or queues. When the demands for services increases and exceeds the limit of the service, therefore, the queue is said to be developed [2].

Its increasingly becoming a common management tool in decision making in the developed world. This important tool, unfortunately, has been minimally used in most healthcare systems in Nigeria and other parts of African. Queuing theory has now been utilized in industries to analyze how resource sharing constrained systems respond to various demand levels [3].

Very lengthy queues are symptomatic of inefficiency in the hospital services. Unfortunately, this has been the case in many public hospitals in Nigeria and other developing countries. Capacity management decisions in Nigerian hospitals are usually based on skill rather than the use of strategic research model-based analysis. Like any hospitals in Nigeria, ST Gerald hospital Kaduna, ST Gerald receives a large number of patients daily has results to a long patient waiting times.

There have been a lot of publications on the Application of queuing theory used to model hospital settings [2]-[6]. In most healthcare settings, unless an appointment system is in place, the queuing discipline is either first-in-first-out or a set of patient classes having different priorities (as in an emergency department, which treats patients with life-threatening injuries before others).

Waiting has become part of human daily activities. When there are too many people waiting to be served and fewer service providers (i.e. when demand is higher than supply), a queue has to be formed to identify the client that comes before another. Hospitals are one of the major organizations that have a greater number of populations of client that are waiting to be served. For many, waiting on a queue is an infuriating experience that should not happen. Those negative experiences can cause a hospital a lot of negative image minds of their patients which may lead to those patients changing one hospital for another, sometimes even a good hospital for a bad one.

For several reasons, satisfying customers within the shortest time possible by providing a faster service has recently received managerial attention. These reasons maybe because in developed countries the standard of living is high, time is valued more than commodity hence, customers have less desire to wait for service. Also, it was unanimously agreed that organizations now get to know the effect of speed of service to the future of their organization.

The solution is to design a microcontroller-based hospital queue management system presented in this paper. It enables the coordination of patients to avoid end to end sitting in waiting without any assurance of who is

next in the queue or who the doctor will attend to next. This will allow patients to flow in an orderly manner seeing the doctor who is to attend to them.

II. Related Work

In a research conducted by [7], a Smart Queue Management System was proposed using Global System for Mobile (GSM) modem Technology; the system uses GSM technology, microcontroller (Atmel 89C51) and personal computer as the main server. Visual Basic (VB6) program serves as the master controller; the system accesses the PC via the RS232 protocol. Their design was for clinical application, which solely relies on the perfect synchronization between GSM, microcontroller and Personal Computer.

Outpatient Queuing Model Development for Hospital Appointment System [8] is a web-based queuing and appointment system that is proposing a lasting solution to the long waiting arrival rate, they stated that factors like medical services, employees experience, and availability of clinical technology affect the performance of an appointment. The system was designed using Adobe Dreamweaver, The system has the following pages: Reserves, Calendar, Login, Patients, and Staff. Traffic route selection for non-signaled junctions was another area of Portable Queuing theory application in real-life analyzed by [9]. Research in [10], designed a queue management system that can help customers to analyze queue status and take action, the research focuses more on the banks queuing system. The Mobile ticket dispenser system (MTDS) is a system designed by [11] it permits the users to place a request remotely and updates their expected waiting time. But it was challenged by; users unable to cancel the order after their request, users can also place multiple ticket requests within no non-mobile systems, this action will make the system to be unavailable. Queue control system with the notification is another method of managing long queues [12], this system was designed with an App interface accessible via smartphones that provide online tokens to the users. Queueing models are also applied in emergency departments (EDs) [3], they reviewed the perspective of demand and supply-side problems as well as other methodologies that are created to addressed complexities in ED operations. The queue management system that uses audio and real-time updates to service requests with smartphone is another avenue of managing a queue; this system has units for ticket registration and verification [13].

There are number of research on traffic route optimization in the signalized junction, and better describes how someone can efficiently manage queue [14], [15], Their proposed method revealed that the maximum length of a queue in each movement and for the entire intersection was greatly reduced both in the peak hours and in the off-peak hour. Queue can also be managed from the outpatient counter [16], using the waiting arrival time and service time of patients.

In conclusion, most of the systems reviewed are either completely software-based, or not completely standalone, some number of limitations, lacks voice call capability, does not call numbers simultaneously i.e. numbers are allocated to a specific unit, which makes the waiting time longer for some patients and does not address emergency case with higher priority which is a different case in this project. This project is designed to call numbers serially and displays them on a Nokia 5110 LCD; it also tells the status of the queue by displaying the numbers that are currently inside the consultation cabin and the cabin letters the consultation is taking place. Four push-buttons were constructed to serve as a trigger for calling the numbers. Each cabin has a push button for making a call.

III. Methodology

Due to the design approach, which is capable of enhancing the hospital queue management system in other for doctors to attend to patients in an orderly manner. It is necessary to determine the scenario for patients to be able to follow a queue in an orderly manner without them standing and waiting for long before been attended to, in some cases some patients might become angry and the said manual queue can be in a disorganized form. Based on discussion among colleagues, it was revealed that some patients might be in the hospital from morning (especially governmental owned) till the closing hours of the doctors and still won't be attended to. In some cases, where the queue in just arranged manually, some high profiled people can just come into the hospital and they will be attended to quickly because they know the doctor directly or the know someone who knows the doctor. Furthermore, since the design is based on a hospital queue management system, a research was carried out in General Hospital Kawo, Kaduna state, Nigeria, to study the queue and the hospital environment for several days, this help to come up with an approach in which the design was made. The system receives the input from the client on the device by the push of the entrance button. This action then triggers a record of the next patient by calling it through a speaker and displaying it on the LCD and therefore led will blink by the entrance of any available doctor.

A. Research Work

Having studied the hospital environment, it was realized three things namely; need to know the numbers of doctors available, is the environment convenient for the patients who are to be attended to, and the way in which the queue can flow in an orderly manner so as to eradicate inconvenience of flow in a manual form. And to achieve this, an evaluation was carried out to ascertain the time spent by each patient on the without been attended to, and the time it takes for each doctor to attend to one patient at a time. It was realized there was a convenience in the hospital and therefore that was neglected.

B. Survey

To determine the time spent by each patient on the queue without been attended to, as part of this research, questions were asked to the main people affected by this situation, and some part-time workers in the ATBU Clinic situated in Yelwa Campus, Abubakar Tafawa Balewa University Bauchi Nigeria. These targeted audiences can give a perfect response to the situation on ground.

C. Summary of Data Analysis

In this section, a summary of major indices in all data analyzed is presented in table I.

Results from table I were generated based on the people consulted who usually patronize the clinic. This result is

designed to minimize the time spent by patients in the queue before seeing the doctor.

The average time to scan the RFID card on the RFID card reader was realized by taking an average time for normal RFID card reader to read a card, process and produce the output data, 10 samples were used and produced outputs approximately between 3.5 to 6 seconds.

TABLE I. SUMMARY OF CONSULTATION

S/N	Summary of Consultation Time	
	Activity	*Time (s)*
1	Average time to scan the RFID card on the RFID card reader	5
2	Average time it takes for a patient to takes his or her seat	50
3	Average time for the receptionist to press the button of the next available doctor	2
4	Average time for the patient to walk to the available doctor's office	55
5	Average consultation time	600

IV. SYSTEM DESIGN

The design of this system was divided into Conceptual part and Algorithmic part, details of this division were narrated as:

A. Conceptual Design

The Hospital can be regarded as an organization based on high technology and information-intensive processes. According Such organizations are based on democratic control mechanisms with institutionalized stakeholder influence in decision processes, but not a hierarchically structured bureaucracies. It is also expected that health care budgets and funding will depend significantly on sophisticated patient and diagnosis classifications.

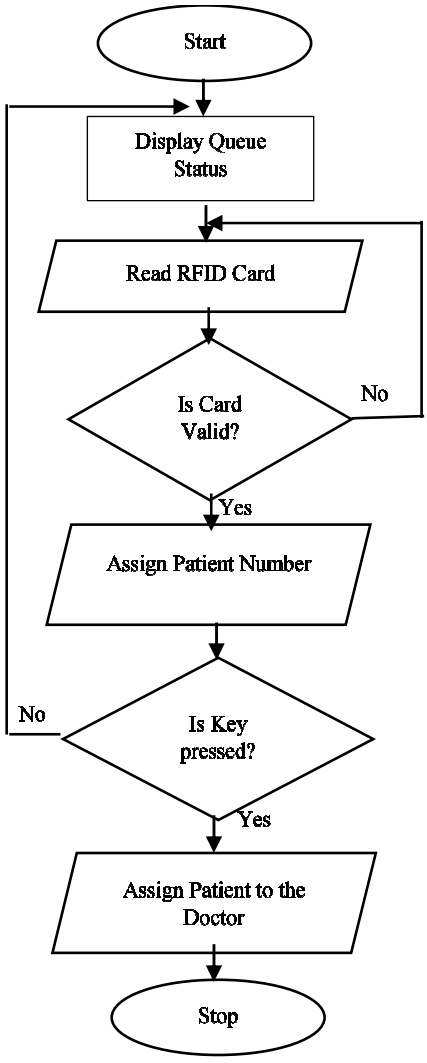
B. Algorithm Design

The algorithm of the system was stated as follows:

1. Start
2. Microcontroller Initializing
3. Tally number is generated for each patient at the entrance of the hospital by placing the RF card on the RFID card reader for it to read and assign a tally number.
4. If RFID card is valid a welcome greeting will be said to the patient and a patient will be told to check his/her number and therefore sit in the assigned place.
5. A patient will proceed to the seat assigned to him/her which are already labelled.
6. If RFID card is not valid no welcome message nor number assigned. Non dually registered RFID card will be an invalid, damaged or spoiled.
7. Doctors will be notified by the receptionist of the waiting patients.
8. Receptionist presses the button of the next available doctor.
9. A this will be voiced out from an audio output
10. In the case of emergency, the next patient will have to wait for the next doctor available to attend to the emergency patient
11. Else if there is no emergency the next patient on the queue will be attended to
12. If next patient which already have a number assigned is not on sit to see the doctor available, the next to him/her will proceed to see the doctor.
13. Stop

C. Flow Chart

A simple flowchart for the design algorithm is presented in fig.1, it is a diagrammatic depiction of the sequence of operation of the whole system.



Flowchart diagram of the design

The section read RFID card in the Flow chart is designed to read the card of the patient and assign number to the patient, so he/she could see the available doctor. The codes are shown as in fig. 2. Doc 1, Doc2, Doc3 and Doc4 means Doctor 1, Doctor2, Doctor3 and Doctor4 respectively. If the system read the patient RFID card and find out that Doc1 is assigned 0 means no patient was assign, therefore, the system will automatically assign this patient to him/her and, this process will continue for every

RFID card reading. The delay(6000) means the voice playout should be read between this allocated delay.

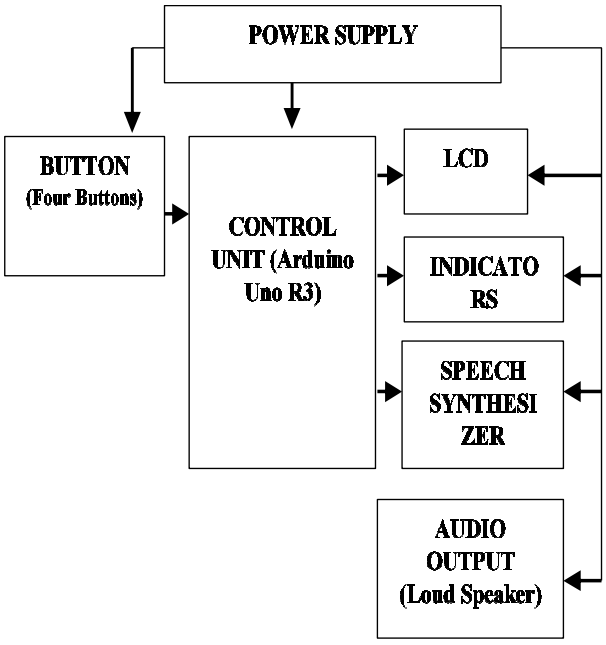
```
void loop() {
Doc1 = digitalRead(doc1);
Doc2 = digitalRead(doc2);
Doc3 = digitalRead(doc3);
Doc4 = digitalRead(doc4);
if(Doc1 == 0){
mp3_play (1);
delay (6000);
count1 ++;
}
if(Doc2 == 0){
mp3_play (2);
delay (6000);
count1 ++;
}
if(Doc3 == 0){
mp3_play (3);
delay (6000);
count1 ++;

}
if(Doc4 == 0){
mp3_play (4);
delay (6000);
count1 ++;
}
```

Arduino Codes for Read RFID Card

D. Hardware Description

The design employs the use of just one microcontroller, which is in the receiver section which combines the decoder, LCD and an audio mp3 recorder. Fig. 3 shows the system block diagram representing the various fundamental units of the system. There is also a transmitter section which comprises of the buttons, when a patient comes in, and scan his/her card on the RFID card reader, after the scanning, a number will be assigned to the patient. Thepatient will proceed to sit in the appropriate sit assigned by the receptionist and wait for a call by the available doctor.



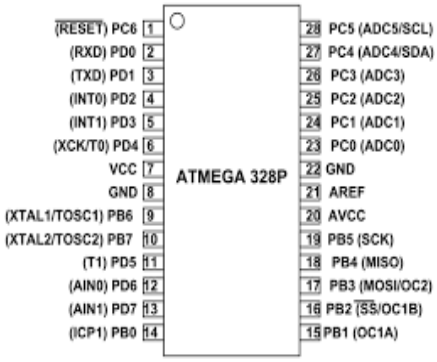
Block diagram of the system

V. Implementation and Testing

Using the algorithmic design, the flowchart, and a programmable microcontroller Integrated Circuit (PIC) (ATMEGA328p), the source code was compiled using an Arduino IDE. Proper consideration was given to the code during compilation in other to avoid any logical errors. The test-run was done using interactive electronic simulation software called Proteus 8.

A. ATmega328

ATmega328 microcontroller shown in fig. 4 is an 8-bit high-performance RISC-based microcontroller that has the combination of 32KB ISP flash memory with read-while-write capabilities, 2KB SRAM, 1KB EEPROM, three flexible timer/counters with compare modes, 23 general purpose working registers, serial programmable USART, internal and external interrupts, a byte-oriented 2-wire serial interface, 6-channel 10-bit A/D converter (8-channels in TQFP and QFN/MLF packages), SPI serial port, programmable watchdog timer with internal oscillator, and five software selectable power saving modes. This device operates between 1.8-5.5 volts.



ATmega328 Chip [17].

B. Schematic Diagram

The hex file was then generated and transferred to the ATMEGA328p microcontroller with the aid of a USB programmer. The design was simulated on a computer to study and analyze the behavior of each stage before physical implementation was carried out. The implementation was finally assembled on Vero board after its initial assembling on Bread board.

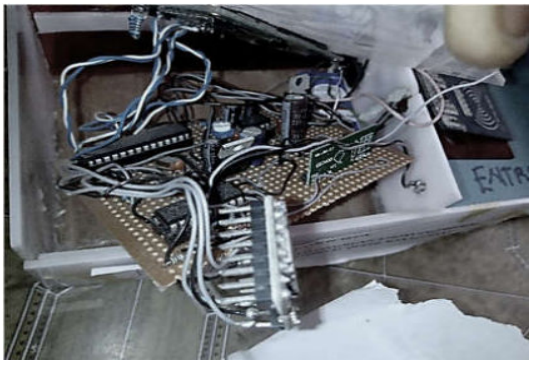
When the system is switched ON it will initialize, then transmitter section will send a signal to the receiver section whereby making the LED in the receiver section to go off. It will wait for further information from the transmitter section.At the entrance into the hospital, each patient will have to scan his/her card on the RFID card reader, after the card has been authenticated it will automatically assign a number to the patient, and the patient will proceed to the appropriate sit waiting to be called to see the doctor available. The receptionist will therefore press the button of the available doctor and this will be led out through an audio player (Speaker) and this will be in the form of first in first out. Then the next patient on the queue will proceed to see the next available doctor.

C. *Physical Connection*

Fig. 5 depicts the connection of the brain of the system that decides which output to give with respect to the available input. This consists of the following components like the Arduino Uno R3 Board, ATmege328 IC microcontroller, LM386 Amplifier, RFID card reader, the Audio output (speaker) connected together to serve as the brain of the system.

Physical connection of the brain of the system

Physical connection of the components that display the connection between the RFID card reader and speaker to the control system. Fig 6 shows the connection of the conponents on the board.



Physical connection of the RFID card reader and the speaker to the control system

VI. RESULT AND TESTING

After the completed design was checked and properly verified, the necessary components for the realization of the design were assembled and arranged as shown in fig. 7. the result shown on the screen represents the result of the consultation room named "Doctor 1 to Doctor 4" which consists of a button and a LED. Upon a simple press, a call to the next patient number is triggered and the patient will therefore go to the doctor number specified in other for the doctor to attend to him/her.

The one-sided Printed Circuit Board (PCB) layout of the simulated circuit diagram was made on Proteus 8.0 virtual software, etched, drilled and components were soldered after tracks continuity test was made using Digital Multi-meter (DMM) in testing the component. The PCB played an important role in determining the overall performance of electronic equipment. It ensures that the noise was introduced as a result of component placement and track layout is held within limits.

Casing of the hospital electronic queuing resolution system

The completed project start-up pace one is shown in fig. 8, shown on the screen is the start-up message "Automatic Hospital Queue Management System". This provide the user with the complete name of the system during start up.

Start-up pace one

The next activityis shown in fig. 9 as the start-up pace two showing the supervisor's name as "Supervised by Engr. U. I. Bature". This is the second start-up message, this serves as the message displaying the name of the system designer.

Start up pace two

The demonstration of the third consultation room named "Doctor 3" is shown in fig. 10, which consists of a button and LED. Upon a simple press, a call to the next patient number is triggered whereby the patient will therefore go to the doctor number 3 in other to get proper attention from the doctor. It is a similar case to that of first and second unit, but note that the doctor can only attend to each and every other patient only when less busy.

This fig. 10 shows "patient=3 and Next: 3", this demonstrate the calling of patient number 3 and Doctor 3. Doctor 3 means consultation room 3. This could be further explained as patient number 3 should go to the consultation room 3.

Activities in the third consultation room

VII. CONCLUSION

The design, construction and testing of an electronic hospital queuing solution system starts with the understanding of the queue system itself, this is very essential to broaden the horizon of understanding. However, the consideration of the control strategy and component to be used plays a vital role in accomplishing this project. Consequently, several articles have been reviewed to investigate the current approaches for queue management system, even though the current approaches have proven to ease and give benefits to service. The system was successfully designed and tested. The four units were successfully tested in such a way that all units can trigger a call which is processed by the Arduino Uno and an output is produced through the LCD and the frequency synthesizer which synthesizes the hex data which is then amplified by an LM386 amplifier which produces a vocal output through a loudspeaker. This project can greatly help in reducing the stress faced in the hospital during consultations. Though the aim and intents of this project has been realized, room for improvement is still open; database system could be incorporated to improve on service delivery. The application of this kind of queuing resolution system should be invigorated in our hospitals to enhance services and eradicate the old-fashioned queuing process.

ACKNOWLEDGMENT

The authors are thankful to the Department of Computer and Communications Engineering, Abubakar Tafawa Balewa University (ATBU), Bauchi, Nigeria for the provision of facilities, materials and enough time that make this work into completion.

REFERENCES

[1] Rosalind Kelly, Erwin Losekoot, Valerie A. Wright-StClair, “hospitality in hospitals: the importance of caring about the patient,” Hospitality & Society, Vol. 6, pp. 113–129, 2016.

[2] Shortle, J.F., Thompson, J.M., Gross, D. and Harris, C.M., Fundamentals of queueing theory, 5th ed, Vol. 399, John Wiley & Sons, 2018. pp. 4-8.

[3] X. Hu, S. Barnes, B. Golden, “Applying queueing theory to the study of emergency department operations: a survey and a discussion of comparable simulation studies,” International Transactions in Operational Research, vol. 25, pp. 7–49, 2017.

[4] Sam Afrane, Alex Appah, “Queuing theory and the management of Waiting-time in Hospitals: The case of Anglo Gold Ashanti Hospital in Ghana,” International Journal of Academic Research in Business and Social Sciences, Vol. 4, pp. 35-44, 2014.

[5] Onoja Anthony “the application of queuing theory on patient satisfaction in ante-natal care clinic,” unpublished.

[6] Ameh N, Sabo B, Oyefabi M O., “Application of queuing theory to patient satisfaction at a tertiary hospital in Nigeria,” Nigerian Medical Journal. Vol. 54, pp. 64-67, 2013

[7] Arun R. and Priyesh P. P., “Smart Queue Management System Using GSM Technology,”Advance in Electronic and Electric Engineering, Vol. 3, pp. 941-950, 2013.

[8] Obulor R. and Eke B. O., “Outpatient Queuing Model Development for Hospital Appointment System” International Journal of Scientific Engineering and Applied Science, Vol. 2, pp. 15-22, 2016.

[9] A. Jaiswal, C. C. Mishra, C. Samuel, “Time-dependent traffic route selection for unsignalized junctions in tandem queueing network,” Proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM), Pretoria / Johannesburg, South Africa, October 29 – November 1, 2018.

[10] Md. Nasir Uddin et al., “Automated Queue Management System,” Global Journal of Management and Business Research: (A) Administration and Management, Vol. 16, pp. 51-58, 2016.

[11] Y. Lin and Y. Lin, "Mobile ticket dispenser system with waiting time prediction," The 16th Asia-Pacific Network Operations and Management Symposium, IEEE Transactions in Hsinchu, Taiwan, 17-19 Sept. 2014, pp. 1-6, 2014.

[12] Sujit Pasalkar, Ankit Kharade, Mangesh Nikam, Prakash Jagade “queue control system using android,” International Engineering Research Journal, Vol. 2 pp. 2162-2164, 2016.

[13] M. Ghazal, R. Hamouda, S. Ali,” A Smart Mobile System for the Real-Time Tracking andManagement of Service Queues” International Journal of Computing and Digital Systems Vol. 5, pp. 305-313 2016.

[14] Z. Cai, M. Xiong, D. Ma and D. Wang, “Traffic design and signal timing of staggered intersections based on a sorting strategy,” Advances in Mechanical Engineering, vol. 8, pp1-9, 2016

[15] A. Jaiswal, C. Samuel, C. C. Mishra, "Minimum carbon dioxide emission based selection of traffic route with unsignalised junctions in tandem network," Management of Environmental Quality, Vol. 30, pp. 657-675, 2019.

[16] A.H. Nor Aziati, N. S. Binti Hamdan, “Application of Queuing Theory Model and Simulation To Patient Flow At The Outpatient Department,” Proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM), Bandung, Indonesia, March 6-8, pp. 3016-3028, 2018.

[17] ATMEGA328P [Online]. Available: http://html.alldatasheet.com/html-pdf/241077/ATMEL/ATMEGA328P/312/2/ATMEGA328P.html [Accessed: 4th August, 2019]

IoT-Based Home Automation, Energy Management and Smart Security System

[1]Abbas Musa Hassan, [1*]U. I. Bature, [2]N. M. Tahir, [1]K. I. Jahun, [2]Kamal Abubakar Abubakar, [1]Shadrach Adamu
[1]*Department of Computer and Communications Engineering, Abubakar Tafawa Balewa University Bauchi (ATBU)*
P.M.B 0248 Bauchi, Nigeria
[2]*Department of Mechatronics and System Engineering, Abubakar Tafawa Balewa University Bauchi (ATBU)*
P.M.B 0248 Bauchi, Nigeria
*biusman@atbu.edu.ng, amhassan@atbu.edu.ng

***Abstract*—The Internet of Things (IoT) controlled home automation allows objects to be sensed and/or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems. Energy consumption can be measured through its environmental impact and usage; the measure of the amount of power consumed by the load side of an electrical circuit is termed energy consumption. Energy consumption is a major issue in the modern world. The high power consumption experienced today was as a result of inefficient power monitoring and controlling techniques in our households, businesses, institutions etc. The proposed study is to showcase the design and implementation of an IoT-Based Home Automation, Energy Management and Smart Security System. The system unit was developed using components such as the universal Arduino microcontroller and esp8266 Wi-Fi module. The proposed system is intended to assist and provide support to enhance in power usage. Also, it will greatly help in reducing the power consumption and the risk of fire incidences. One of the features of the developed unit is the ability to monitor and control electrical and electronics appliance remotely from the internet through a website application. The feature of the system ensures home security and ease of control. It effects in improved efficiency, accuracy and economic benefit in addition to reduced human intervention.**

Keywords— Monitor, Internet of things, Electronics appliance, Home security, Arduino microcontroller.

I. INTRODUCTION

Home Automation System is a channel whereby owners and inhabitants can remotely control various electrical and electronics home appliances. In home automation system, robotics and computer technologies are used on household appliances and such home automation are termed domestics. The most vital feature of home automation is the energy saving offered to its clients, most especially the preoccupied ones, whereby the overall home energy usage can be monitored to ensure effective power usage by turning-off all unnecessary appliances to minimize energy consumption [1, 2]. The most interesting feature of the internet based home automation is its convenience. One does not need to go and Turn – ON the geyser at home and then wait for the water to be warmed. While at work, one can just turn – ON the geyser to make sure that once one is home, the water has already got warmed enough and is convenient for use, thereby ultimately saving time. Although security issues exist, the user can remotely be conveniently monitoring the house via surveillance cameras. The incidents of property intrusion should as well be monitored. The home entertainment section should help a user in controlling the sound distribution all over the house, or the intensity of light from the sofa comfort while still viewing an exciting movie.

Due to an increase in population and consequently an increase in energy demand, an inexorable need for conservation of energy through any possible means arises. The key cause of energy consumption is the failure to monitor as well as control electrical and electronics appliances remotely. There will be well over 26 to 64 billion connected devices by 2020, as prompted by survey Gartner in [3]. IoT is rapidly becoming an increasingly growing topic of discussion both in and out of the workplace. Its concept does not only have the panorama to influence how we live, it also has an impact on how we work. This concept is the basis for linking devices with an on/off switch (s) to the internet. Internet of things has rapidly evolved through the years; this has always been the talk in our workplace, homes and some industries. IoT is termed as the connectivity into physical devices and objects, embedded with electronics, forms of hardware and internet connectivity whereby these devices can interact and communicate with one another over the network and can consequently be controlled remotely and monitored [4, 5, and 6]. This system is applied in areas where human physical presence is not necessarily required all the time [7].

The major objective of this work is the development of a model, for an internet based home automation system. The aim is to form a podium that permits communication between web applications and microcontroller located at a remote position wherever in the universe.

II. REVIEW OF PAST WORK

ZigBee-based Home Automation arrangement was developed by [8], the arrangement was designed and implemented using ZigBee in order to control and monitor the home appliances. It has both simple flexible user interface (UI) with system remote access. A devoted virtual home is further applied, to accommodate both the safety and security challenges of the system. A Wireless Fidelity (Wi-Fi) – Based Home Automation Using Android and Arduino Platform was designed by [9], a system based on Wi-Fi networking technology which uses Arduino based board with a controller was proposed. The system has a graphical user interface (GUI) which is provided by android applications for users to get connected to the hardware and establish interaction with their home. Along with the feature of automation and control, the system also offers a way for the users to know about power usage details by the home appliances.

IoT-based Smart Security & Home Automation System was designed in [10]. A system that has leverage over the related types of existing systems was designed, such that alerts and status sent via Wi-Fi connected microcontroller – managed system can be realized on the user's phone from any length, regardless of whether the phone is connected to internet or not. The microcontroller used in the existing prototype is the TI-CC3200 launch - pad board, which typically comes through an embedded microcontroller and an on – board Wi-Fi shield, via which all the electrical home appliances can be managed and controlled. In [11], integration of smart house and smart meter to key-based and identity management schemes was proposed. And based on the prototype the proposed schemes are feasible for adequate security of the smart house. [7] Proposed low cost and smart home automation device via IoT. This is proved to be effective in the control of all electronics appliances in the house as well can be used for online billing via a web site. In [12], an algorithm was developed to save energy in the house using motion sensors. [13] Presented a home automation using hand recognition for the visually challenged persons. Using this technique, visually challenged can be able to interact with the electronic devices in the house. [14] Proposed a technique which collects and store data using the power of cloud computing services. Thus, neural network models will be generated using this information to generate energy consciousness by guiding the house occupants. [15] Presents an IoT – based monitoring and control of the smart home by eye-tracking, using the system home electronics such as lamps, fan, television, radio etc can be easily controlled by disabled persons.

III. Design and Methodology

This section provides the detail explanation of the design and methodology employed. Starting from the hardware block diagram, system architecture, circuit diagram and software development process, so many materials were needed to realize the proposed system.

A. System Architechture

A microcontroller-based home automation scheme is shown in Fig. 1. In this design, user's mobile software interrelates with the microcontroller by means of the network server via internet protocol. The microcontroller commands come from the user interface (UI). The microcontroller then executes the necessary tasks based on a pre-defined control algorithm that governs the controller. The controller gives the device status and then updates the information on to the server for the mobile application of the user.

The server takes care of the users while ensuring safe communication amongst the mobile application of the user and the control unit. The user once identified, would then be given access to the control interface (i.e. the Web page). The benefits gained by employing microcontroller include; compact circuitry, affordability, and an overall improved flexibility. Microcontroller can completely replace integrated circuits (ICs). It is easily reprogrammable for functionality modification. The Arduino Uno R3 microcontroller was used for this research work, because it uses ATmega16U2 which allows for more memory and faster transfer rate. The major imperfection here was the system's reliance on mobile connectivity to the internet. Once the mobile connectivity is bargained, the user cannot be monitoring and controlling the system remotely. The restricted number of devices solely relies on the I/O boundary of the employed microcontroller. A Programmable Logic Circuit (PLC) can be used as the system controller; the microcontroller is less robust than the PLC. Microcontroller has been chosen over the PLC due to cost consideration. In addition, PLC is not an open source; hence the Arduino microcontroller becomes the finest choice due to system prototyping. Fig.1 shows the system architecture which comprises of the Web section and its components, the communication section, the hardware section etc.

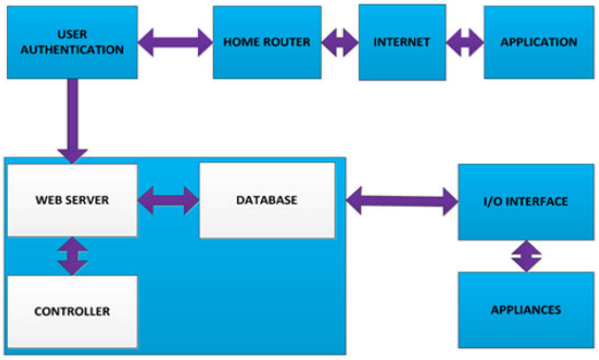


Fig. 1. Proposed Home Automation (Internet Based Home Automation with Microcontroller)

B. Hardware System Diagram

The microcontroller is an autonomous system having a processor, peripherals and memory. It can therefore be deployed as an embedded system. Most modern day programmable microcontrollers are embedded within consumer products and machinery such as mobile phones, home appliances, automobiles and computer systems. This is why a microcontroller is also known as an "embedded controller." An embedded system can be more or less sophisticated, depending on the memory requirement, programming length and software complexity level. I/O devices consist of solenoids, switch buttons, Liquid Crystal Displays (LCDs), digital relays, and sensors such as humidity sensor, temperature sensor, and light intensity sensor or power usage as shown in Fig. 2.

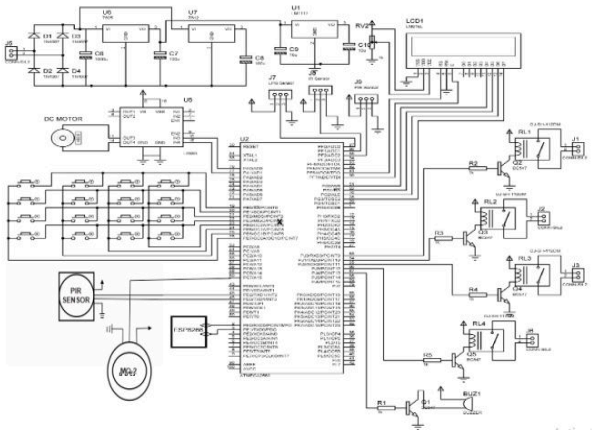


Fig. 2. Hardware System Diagram

C. *Device Construction*

After the completion of the design, it was later verified and confirmed that all the required components for the physical actualization of the design were gathered as required. The block diagram was established into the schematic shown in Fig. 4. The circuit diagram was used for the system hardware components Modeling and Simulation in Proteus Virtual System Modeling Environment (VSM), Version 8.0

D. *Software Development*

ATMega328p Microcontroller must be first programmed before been used in an electronic hardware. Assembly programming language is used in programming the Microcontroller. The coding must be both correct and functional so as to ensure proper working of the Microcontroller. The codes' functionality is ascertained via Keil Micro Vision 4. Proteus Virtual System Modeling (VSM) Version 8.0 was employed for the simulation. Arduino Sketch 1.8.7 software was employed for code burning onto the Microcontroller. The generated hex code was stored in the computer after compilation. This generated hex code is burnt onto the ATMega328p via Arduino Sketch 1.8.7 Software. Fig. 3 shows the software design using a flow chart.

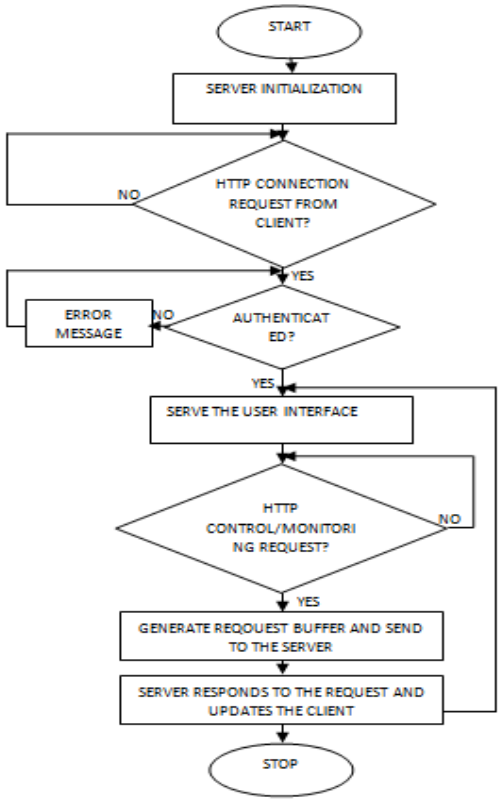


Fig. 3. Flow Chart

The section authentication in the Flow chart is designed to help users to get access to the door system, it controls any unauthorized access to the entrance. The Arduino codes are shown in fig. 4. The first loop in the codes will check the password provided for match if match is success, access will be granted otherwise access will be denied and the second loop will check for appropriate alert and make the beep sound.

```
void check_door() {
  for (i = 0; i < 30; i++) {
Lcd.Clear();
Lcd.SetCursor(0,0);
Lcd.Print("Input password");
Lcd.setCursor(0, 1);
Lcd.SetCursor(5,1); //To Adjust one Whitespace for Operator
Lcd.Print(num 1);
DigitalWrite(buzzer, LOW);
Char Key=myKeypad.getkey();
If(Key!= NO_Key && (Key == '1' || Key == '2' || Key == '3' ||
Key == '4' Key == '5' || Key == '6' || Key == '7' || Key == '8'
Key == '9' || Key == '*' || Key == '0' || Key == '#')) { Num 1 =
Num 1 + Key; int NumLenght = Num 1.lenght();
Lcd.SetCursor (5, 1); // To Adjust One Whitespace For Operator
Lcd.Print(Num 1);
DigitalWrite(buzzer, LOW);
Beep()
If (NumLength == 4)
{ if (Num 1 ==Password) {
Lcd.clear();
Lcd.SetCursor(0, 0);
Lcd.Print("Password Accepted");
Lcd.SetCursor(0, 1);
Lcd.Print("Pls wait…");
DigitalWrite(Buzzer, LOW);
Delay(2000);
Opendoor(); }
else{
Lcd.Clear();
Lcd.SetCursor(0, 0);
Lcd.Print("Wrong Password");
Lcd.SetCursor(0, 1);
DigitalWrite(Buzzer, LOW);
    For (i=0; i<8; i++) {
Beep();
}
    }
```

Fig. 4. Arduino Codes for the Unauthorized Access

IV. RESULT AND DISCUSSION

The simulation was performed using Proteus 8.0. To safeguard the design effectiveness, the design numbers of components were linked in Proteus 8.0 environment

A. *Website*

Web sites are basically generated for several motives; it could be for entertainment or education etc. Nowadays, web

sites can be employed in controlling home appliances. The user interface (UI) via which the user is presented with clickable switches to be able to control home appliances as well as to monitor their power consumption is achieved with an Arduino web server. The web site was formed with Java Scripts, XML and HTML. Once a user is allowed to have an access, the UI becomes available for turning on/off devices and power consumption status check; everyone has access to the web site homepage.

1) Web Login Page

To access the main control and notification page, an authorized user has to login through the login page as shown in the fig. 5. The user will be required to provide his login details; naming style was designed to accommodate alphabets and spaces for security reasons. After a successful login, user can get access to the application platform where he can control the appliances.

The user interface control section consists of buttons which once clicked will send an equivalent request to the server. The server then turns ON/OFF the associated appliances. In addition, the server updates the web user with the matching images of appliance status.

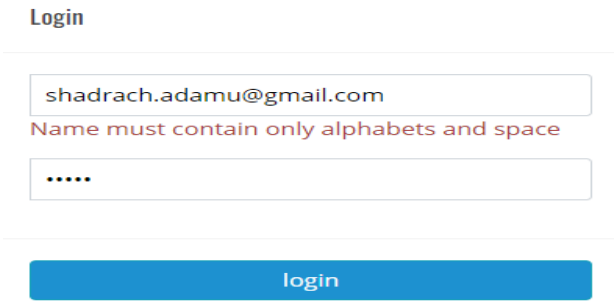


Fig. 5. Web Login Page

2) Control Panel

This is the most important part of the web design section; here control is being carried out on home appliances. Fig. 6 shows the control panel. User can view the appliances and control the appliances. From fig. 6, appliances like Kitchen Bulb, Bedroom Bulb, Main Room Bulb and Main Door control buttons were shown with the status OFF. Therefore, user can conveniently switch ON/OFF appliances of his choice.

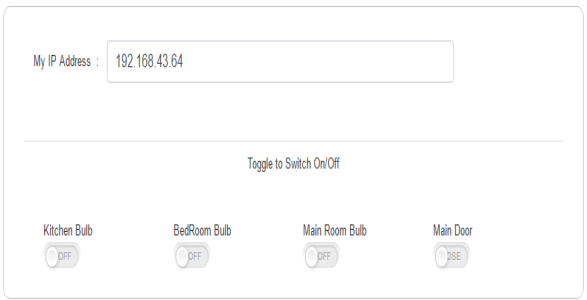


Fig. 6. Web Control Panel

3) Notification Page

On this web page of the application platform, the owner of the platform will receive notifications from the system in the event of any unauthorized access. Fig 7 shows the intruder detected and wrong password detected messages; this was incorporated to enhance the security of the system.

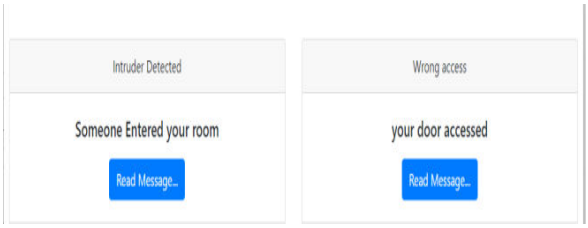


Fig. 7. Web Notification Page

B. The System Unit

The system unit is housed in a Polyvinyl Chloride (PVC) Casing of dimension 15cm x 8cm x 5cm. A little vent window was also created to minimize component overheating as shown in fig. 8. This shows the complete physical components connected, the keypad for login purposes, the display unit for displaying messages etc.

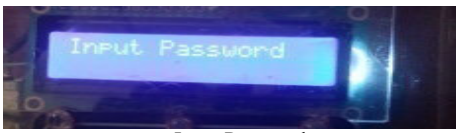
Fig. 8. System Unit

C. The System Test

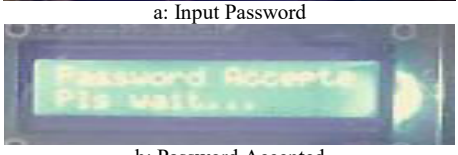
This section presents various system test processes, the door opening/closing via keypad, door opening/closing from the web control panel, switching ON/OFF of the bulbs via the web control panel etc.

1) Opening/Closing of the Door Using Keypad

Keypad serves as a means of accessing the door system; fig. 9 shows how to open the door starting from a-c.



a: Input Password



b: Password Accepted

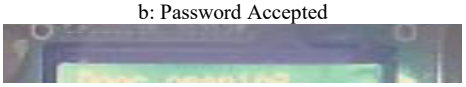


Fig. 9. Door Openning via the System Unit

The door opening can be successfully implemented only if the user has a registered password, user can simply key in his password, after the password has been acknowledged then the door will automatically start opening, user can also view the process via the system display as shown in fig. 9c. The door closing is just a replica of the opening procedures.

2) Opening/Closing of the Door via the Web Control Panel

Alternatively, door can also be accessed from the web control panel, in case of any problem with the keypad system as sown in fig 10.

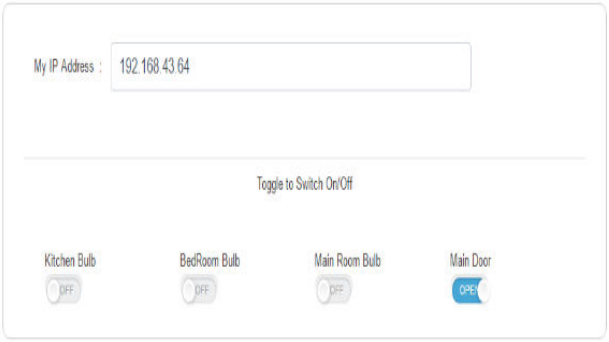


Fig. 10. Door Openning via the Web Control Panel

The opening of door via the web control panel is achieved after a successful login to the web application, main door control button is shown in active state in fig. 10, the blue color ON indicate the main door is now opened.

3) Switching the Bulbs from the Web Control Panel

The web control panel can be used as the alternate way of switching the bulbs; fig. 11 shows how bulbs can be switched ON/OFF from the control panel. Appliances like Kitchen Bulb, Bedroom Bulb and Main Room Bulb control buttons were shown in blue color, this indicate the ON state of the appliances. User of the web application can accessed the web application control panel via a successful login.

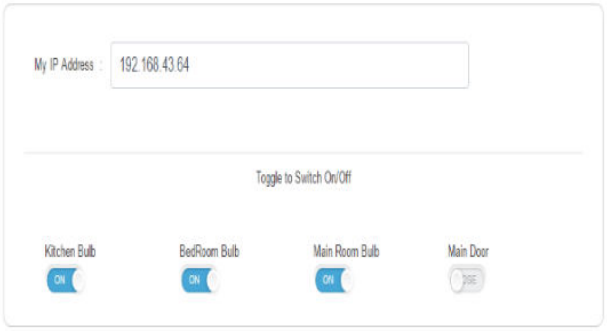


Fig. 11. Door Openning via the Web Control Panel

Likewise, the switching of the bulbs can be done via the web application control panel after login.

V. AREA OF APPLICATION

A. Surveillance and Security

Security systems can be incorporated with this system like the Cameras, luminance sensors etc. this can enable the user to monitor various section of his environment.

B. Management of Energy Consumption

Energy consumption of the appliances can be reduced by turning off the appliance remotely, this indicate that appliances can be turn ON/OFF and monitored in real-time.

C. Entertainment

The system is capable of switching ON/OFF appliances, this indicate that it can be incorporated with other system to like the sound system, lighting system etc.

D. Access Control

System like Digital locking can be integrated via the keypad section of the system; this will enhance security and provide monitoring about the person leaving/entering the house.

VI. CONCLUSION

This paper has summarized a design and development of the web enabled home automation scheme. System for the IoT-controlled Home Automation scheme has an unbounded application in the technology driven marketplace due to the various numbers of devices linked to the internet.

The IOT-Based Home Automation, Energy Management and Smart Security System presented in this paper describes the processes of appliances access and control via the system unit keypad and alternatively via the web application control panel, each component of the system is modularized so as to make it more efficient, hence guaranteeing it to be integrated with an innumerable device range. The rudimentary idea of the scheme is to provide a well secured and convenient system for the client, which will in turn help in the great degree of control, mobility and seurity.

ACKNOWLEDGMENT

The authors are grateful to Tertiary Education Trust Fund (TETFund) for the funding to attend this particular conference.

REFERENCES

[1] Joumaa, H, Ploix, S, Abras, S and De Oliveira, G. "A MAS integrated into Home Automation system, for the resolution of power management problem in smart homes", *Energy Procedia*, Vol. 6, March 2011, pp. 786-794

[2] Mehdi, G and Roshchin, M. "Electricity Consumption Constraints for Smart-home Automation: An Overview of Models and Applications", *Energy Procedia*, Vol 83, December 2015, pp. 60-68

[3] Gartner "Consumer Applications to Represent 63 Percent of Total IoT Applications in 2017", [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016 [Accessed: 6th August 2019]

[4] A. Bissoli, D. Lavino-Junior, M. Sime, L. Encarnação, and T. Bastos-Filho, "A Human–Machine Interface Based on Eye Tracking for Controlling and Monitoring a Smart Home Using the Internet of Things," Sensors, vol. 19, no. 4, p. 859, Feb. 2019 [Online]. Available: http://dx.doi.org/10.3390/s19040859

[5] Kumar, S. Sanjay, Ayushman Khalkho, Sparsh Agarwal, Suraj Prakash, Deepak Prasad, and Vijay Nath. "Design of Smart Security

Systems for Home Automation." Nanoelectronics, Circuits and Communication Systems, pp. 599-604. Springer, Singapore, book series (SCI, volume 771) 2019

[6] Monowar, Maisun Ibn, Shifur Rahman Shakil, Abdulla Hil Kafi, and Md Khalilur Rhaman. "Framework of an Intelligent, Multi Nodal and Secured RF Based Wireless Home Automation System for Multifunctional Devices." Wireless Personal Communications (2019): Vol. 105(1) 1-16.

[7] Mahmud, Sadi, Safayet Ahmed, and Kawshik Shikder. "A Smart Home Automation and Metering System using Internet of Things (IoT)." In 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), pp. 451-454. IEEE, 2019.

[8] Khusvinder Gill, Shuang-Hua Yang, Fang Yao, and Xin Lu. "A ZigBee-Based Home Automation System*" IEEE Transactions on Consumer Electronics*, Vol. 55, No. 2, May 2009.

[9] Sushma P., Roopa M J."Wi-Fi Based Home Automation Using Android & Arduino Platform". International Journal for Research in Applied Science & Engineering Technology (IJRASET). ISSN: 2321-9653, Volume 5 Issue VI, June 2017.

[10] Ravi Kishore Kodali, Vishal Jain, Suvadeep Bose and Lakshmi Boppana. "IoT Based Smart Security and Home Automation System".*IEEE International Conference on Computing, Communication and Automation (ICCCA2016)*, ISBN: 978-1-5090-1666-2. 2016.

[11] Vilmar Abreu, Altair Santin, Alex XavierAlison, LandoAdriano, Witkovski Rafael, RibeiroMaicon Stihler, Voldi Zambenedetti, Ivan Chueiri, A Smart Meter and Smart House Integrated to an IdM and Key-based Scheme for Providing Integral Security for a Smart Grid ICT, *Mobile Networks and Applications,* Vol. 23, Issue 4, pp 967–981 2018.

[12] Lohan, Vibha, and Rishi Pal Singh. "Home Automation using Internet of Things." Advances in Data and Information Sciences, (SCI, volume 771) , pp. 293-301. Springer, Singapore, 2019.

[13] Kumar, Ammannagari Vinay, Adityan Jothi, Adil Hashim, Pulivarthi Narender, and Mayank Kumar Goyal. "Home Automation Using Hand Recognition for the Visually Challenged." In Integrated Intelligent Computing, Communication and Security, pp. 405-413. Springer, Singapore, 2019

[14] Popa, Dan, Florin Pop, Cristina Serbanescu, and Aniello Castiglione. "Deep learning model for home automation and energy reduction in a smart home environment platform." Neural Computing and Applications (2019): pp. 1-21, 2019.

[15] Susheela, K., E. Sri Harshitha, M. Rohitha, and S. Maheswara Reddy. "Home Automation and EMonitoring Over ThingSpeak and Android App." In Innovations in Electronics and Communication Engineering, pp. 127-137. Springer, Singapore, 2019.

A Data Visualization Approach for Enhancing Room Management in University Buildings

Fortune Chuku
School of Information Technology and Computing
American University of Nigeria
Yola, Adamawa State, Nigeria
fortune.chuku@aun.edu.ng

Charles Nche
School of Information Technology and Computing
American University of Nigeria
Yola, Adamawa State, Nigeria
charles.nche@aun.edu.ng

Sandip Rakshit
School of Information Technology and Computing
American University of Nigeria
Yola, Adamawa State, Nigeria
sandip.rakshit@aun.edu.ng

***Abstract*—** Facility management requires that resources are managed effectively. Facility managers lack an efficient method of extracting meaningful knowledge from large datasets containing occupant behavior information. This study explores what insights about occupants' behavior can be obtained from a network data log by applying a visualization approach to analyze and present the data findings, to assist in enhancing facility management. As previous works have explored the phenomenon using models developed from natural sciences, this study uses an organizational learning model developed from a social science theory to guide the conduct of the data analysis to obtain useful insights about what times the academic community use the buildings on campus. Microsoft Excel 2016, a powerful IT tool is used to analyze and visualize the two-week log data from sixteen buildings, comprising academic and residential rooms, which showed pattern of building occupancy during the day time and at night, as well as during different periods in a semester. The results showed that students use the academic facilities, that is the library and classrooms, the most during day time between 12pm and 1pm, and in the evening from 7pm to 11pm, extending into 2am during exam periods. It was also observed that in the week preceding an academic assessment, students use the academic facilities more during night time than during day time. Recommendations were made, based on the knowledge obtained, on how facility management at American University of Nigeria can be improved. Implications and suggestions are made for future research works.

***Keywords*—**facility management, data visualization, WiFi, occupant behavior, organizational learning

I. INTRODUCTION

People interact with building facilities differently. The facilities in this context consists of the rooms in a building. The manner in which patients in a hospital use the building spaces is not the same as how students in a university use the buildings, and it is also different from how people in a corporate building occupy theirs. Buildings are complex to understand because of the changing behaviors of their residents. With increasing developments in urban cities people want better services that will improve their working and living standards. For this reason, building administrators and facility managers require information about how people interact with the buildings that they live and work in daily. The information would improve their decision-making process in planning and designing buildings that would meet the present and future needs of urban dwellers. This has formed facility management research which brings benefits such as proper building maintenance, efficiency in spending, and increased customer safety and satisfaction [34].

The role of a facility manager is to manage an organization's resources as well as the operational and critical support services in the organization [17]. In order to implement good facility management practices, the characteristics of the organization must be understood. People spend on average 20 hours per day indoors [40] which accounts for majority of their lifetime, thus, understanding how humans occupy buildings is of great benefit for engineering, business, and scientific purposes [7]. Understanding the mobility patterns of people has become increasingly important due to the dynamic nature of urban cities. It is very common for people to carry their smartphones wherever they are going thus their movement and behavioral patterns can be captured to enable improved decision making by urban planners and building managers [23]. Understanding the routine behaviors of members of an academic community can help universities to know their needs and how to improve and increase campus facilities [32] and conditions, leading to efficiency and effectiveness of the academic community [5].

In universities, resource management is difficult [7]. Facility management lacks efficient methods for collecting real-time information of building occupation to inform on capacity and usage [37]. Planning and maintenance of building facilities is a challenge faced by facility managers [20]. The inability of facility managers to forecast how buildings would be used by its occupants has led to an overcrowding problem in Nigerian public universities [5]. This paper leverages the occupancy data of American University of Nigeria to answer the following questions: *What social science theory or framework can be used to guide a study on occupancy data visualization?*, and *What insights on building occupation can be obtained from the WiFi logs of AUN?*

This paper is structured as follows: Section II provides a discussion of related works applying the use of WiFi monitoring to provide solutions to the challenges faced by facility management. Also, in this chapter, the theoretical framework used in the study is explained. Section III contains information about the methodology and the research method applied to conduct this study. The process of data collection and the analysis tool that was used is also communicated. Section IV presents the results of the data analysis as well as a discussion of the findings and concludes the study by stating the relevance and significance of the research. Recommendations for future studies are also provided.

II. RELATED WORK AND LITERATURE REVIEW

Studies on user behavior when using a WiFi network have classified the objective of the data analysis according to three classes, namely; network optimization [2, 1, 8, 6, 19, 39, 36], assessment of user mobility [15, 16, 26, 29, 30], and type of users [15, 35, 38]. This thesis can be grouped into the second class involving WiFi monitoring of user mobility. Human indoor mobility tracking can be beneficial in applications such as emergency response, facility management, and healthcare research [25].

Majority of user behavior studies that have examined the usage of wireless networks have focused on bandwidth usage. Redondi *et al.* [27] in a study used a classification technique based on Quadratic Discriminant Analysis (QDA) to find out if there were differences between the WiFi usage of students attending an architectural class and an engineering class. The knowledge obtained from the study was recommended for application in smart cities and smart buildings to provide city administrators with an understanding of how to respond to residents' needs. Redondi *et al.* [27] further suggested an exploitation of Wi-Fi technology in enabling high level services. Thapaliya and Springer [36] performed an analysis of WiFi data collected across buildings in the University of Nevada Reno (UNR) to obtain the spectrum usage statistics of students on campus, in order to be better informed on when and where to deploy Unmanned Aerial Vehicles (UAVs) acting as access points, to provide good connectivity services.

However, other studies have used techniques to analyze and group WiFi usage data for the purpose of understanding human mobility patterns. Calabrese *et al.* [9] used eigendecomposition to study the WiFi network of Massachusetts Institute of Technology (MIT) in order to obtain an assessment of building use across the campus. The use of eigendecomposition technique by the researchers introduced the concept of place-based behaviors in human mobility research. Following this, Sookhanaphibarn *et al*. [32] used an eigendecomposition framework to analyze the one-year WiFi log data of Bangkok University in order to obtain the routine structure of people on a campus. Building on their work, Sookhanaphibarn *et al*. [33] used eigendecomposition and hierarchical clustering techniques to design and develop a visualization system of the routine structure of students on a campus.

In a literature review carried out on occupant behavior, Hong *et al*. [13] suggested that insights should be derived from social science theories to enhance the understanding of the phenomenon. This research study will contribute to the literature by using a social science theory in guiding the understanding of occupant behavior.

Knowledge management is of importance to facility management to enhance the efficiency and effectiveness of organizations [12]. KM processes involve knowledge acquisition, refinement, sharing, and utilization [18]. The relationship between KM and Organizational learning can be conceptualized in various ways. For instance, Easterby-Smith and Lyles [21] viewed OL as a concept that focuses on the process, while KM focuses on the knowledge content which an organization acquires, creates, refines, and uses. Through organizational learning, an organization can improve on its knowledge utilization [18]. Huber, as cited in [31], viewed organizational learning theory as having four constructs that describe the learning process of organizations namely, data acquisition, information sharing, information interpretation, and organizational memory. Data acquisition involves the process of scanning or collecting data. Information sharing is the process by which the data collected is distributed within the organization. Information interpretation is the process of giving meaning to the data. Organizational memory involves storing the data for future use as stated by Huber in [31]. In OL literature, researchers have described other learning processes such as monitoring, understanding, and sense making [11]. However, in [11], these concepts were organized according to three stages that make up the OL process. The first stage is scanning, which involves monitoring the environment to collect data. The second stage involves interpretation where the data is given meaning. The third stage is the learning stage where action is taken based on the interpreted data. The three stages are illustrated in a simplified model as shown in Figure 1. A feedback loop interconnects the three stages to input the new data obtained from the learning stage into the preceding stages for subsequent interpretation. The justification for using OL theory to conduct this study was to explore how the acquisition of new knowledge could inform an organization on the ways to improve their services through better decision making.

III. METHODOLOGY

When employing a method to study human indoor behavior, Petrenko [25] suggested that such techniques should be conducted over an extended period of time, and should bear no biased representation of the sensed individual's age, gender, or origin. This study took this suggestion into consideration in the formulation of its methodology. The model as shown in Figure 2 was used to guide the process of extracting knowledge from building occupancy data.

This research uses a descriptive method to interpret meaning from data obtained from the network department of AUN. A descriptive research is concerned with the acquisition of information about phenomena for the purpose of description and interpretation. To acquire data for this research, a primary data collection method was used. Primary data are data that are collected to answer a research question [14]. The positivist paradigm is based on the assertion that real reality or events can be captured or observed through the use of research instruments and explained by means of logical analysis [4]. Positivism emphasizes objectivity in the conduct of research involving the use of research methods including quantitative analysis and laboratory experiments [4]. A positivist approach involving the use of quantitative analysis will be employed in this study.

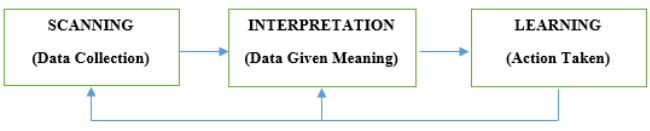


Fig. 1: A simplified model of organizational learning process [11]

Fig. 2: Organizational learning research model

A. Case Study

AUN is a large-scale university located in North East Nigeria with a population size of 2,277 comprising 1,200 staffs and faculties and 1,077 students. The security personnel make up a significant portion of the staff size. The university campus has several academic and residential buildings including, the Robert Pastor library which accommodates several lecture and study rooms; an arts and science building where lectures and laboratory sessions are held and where faculty offices are located; a cafeteria; a law school building accommodating lecture and study rooms as well as a library; a commencement hall housing the university gym; nine male and female dormitories; a clinic and two administrative buildings. Every corner of these buildings is connected to the campus wireless network. There are numerous access points in the buildings providing connectivity to the occupants every day. The APs only cover usage inside the buildings as outside the signal strength is weak. AUN's network engineers use Cisco and Grandstream network applications to monitor WiFi usage on its campus.

By accessing the access points located in several buildings through the network applications, the number of connected users can be determined in real time. The collected data which were in the form of WiFi logs contained usage records of users in sixteen buildings for two weeks, starting from November 23rd to December 6th, 2018. This period of observation was in the Fall 2018 semester, just two weeks before the final exams starting on December 10th, 2018. The choice of this period was to discover whether some factors influence the students' choice to be in a particular building and room at a particular time. Also, the sixteen buildings sampled for data collection and analysis were selected based on the purpose of their usage. The research was concerned with collecting information about the volume of WiFi usage of AUN community at particular times and of certain locations. Hence, no individual data was disclosed to the researcher in the process of data acquisition from the university's network engineers.

The analysis of the collected data in this research was performed using Microsoft Excel 2016, an information technology tool used for statistical and analytical purposes. The data obtained from the WiFi logs was analyzed with Excel to illustrate a visualized pattern of the occupation of the university buildings. The primary objective of visualization is to explore data to gain insights as shown by Owen *et al.* in [10]. This pattern will show how AUN's academic community use the different buildings on campus during different hours and days of the week. Visualizing the data will enable patterns to be identified so that whatever information is obtained can be easily understood [28]. The knowledge derived from the analysis can then be fed back for further processing of the data to obtain subsequent interpretations and meanings as shown in Figure 2.

IV. ANALYSIS AND RECOMMENDATIONS

WiFi can be used to provide information about building occupancy. However, it does not give a count of the number of people in the building, rather it measures the number of connected devices which are attributed to people in the building. The number of connected devices give an indication of the presence of people at a particular location and at a certain time based on an average number of connected devices per person [37]. The data to be analyzed in this study include device connectivity data of students and staffs connected to the AUN WiFi network. In order to be connected a user account is required, thus, devices of guests, or those connected to a different network, or that are not connected at all are not included. Table 1 shows the statistics on the average and maximum occupation for each of the sixteen buildings over the two-week observation period.

In order to infer the graphics presented below, some details about the data need to be taken into account:

i. Between 9pm on December 2nd, 2018 and 6am on December 3rd, 2018 no occupancy data was recorded in the following dormitories; CC, DD, AISHA KANDE, and R-VOLPI as shown in Table I.1 in the appendix. This was as a result of a problem with the fiber cable supplying internet connectivity to these buildings.

To allow for a detailed analysis of the dataset and to prevent repetition, the sixteen buildings were grouped according to the purpose of their usage. The grouping is as follows:

i. Academic Buildings – Robert Pastor Library, Arts and Science, and Law Library
ii. Residential Buildings – AA, BB, CC, DD, EE, FF, Aisha Kande, G-Volpi and R-Volpi
iii. Offices – Admin 1 and Admin 2
iv. Leisure – Cafeteria and Commencement

Table 1: Statistics on the number of occupation per week

BUILDING	Average Occupation		Maximum Occupation	
	Week 1	Week 2	Week 1	Week 2
AA	77.38	81.70	112	113
BB	93.66	99.13	116	146
CC	77.61	75.85	115	112
DD	52.68	50.14	91	99
EE	132.02	139.27	169	185
FF	114.09	115.42	153	158
AISHA KANDE	87.39	74.55	141	126
G-VOLPI	146.74	152.15	187	191
R-VOLPI	113.07	111.22	150	169

CAFETERIA	27.32	30.77	85	99
RP LIBRARY	116.01	123.03	408	346
COMMENCEMENT	10.95	12.86	56	43
ARTS & SCIENCE	63.52	63.04	226	210
LAW LIBRARY	36.92	30.32	144	144
ADMIN 1	31.17	33.39	115	115
ADMIN 2	26.24	23.78	127	117

The dataset was then analyzed and similarities in the pattern of occupancy were found among the buildings in the categories. A building was thus selected from each category to communicate the pattern of occupation in them. They are library, G-Volpi, admin 1, and cafeteria.

As observed in Figure 4, occupancy in the library during the day time (between 7am and 7pm) in the week from Monday to Thursday decreased in the second week when compared to the first week (shown in Figure 3), as the final exams drew closer. This is because it is common practice for classes to stop holding a week to the final exams to allow students enough time to revise for their exams. On the other hand, occupancy in the library during night time (between 7pm and 3am) increased in the week preceding the final exams (shown in Figure 4) when compared to the previous week (shown in Figure 3). This is as a result of students coming to the library at night time to study.

To illustrate the building usage volume, pie charts (Figure 5 and Figure 6) were plotted for the two weeks under observation using maximum occupation values listed in Table 1. It was observed that in the second week, there was increased occupancy in almost all dormitories when compared with the previous week as students had fewer classes. As explained above, this result is attributed to students spending more time in the dormitories as classes generally stop holding a week to the final exam.

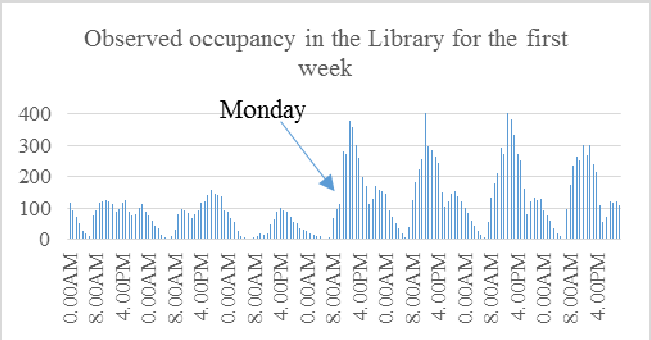


Fig. 3: Observed occupancy in the RP library for the first week

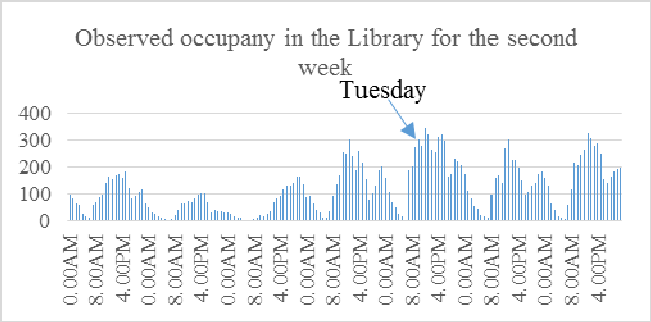


Fig. 4: Observed occupancy in the RP library for the second week

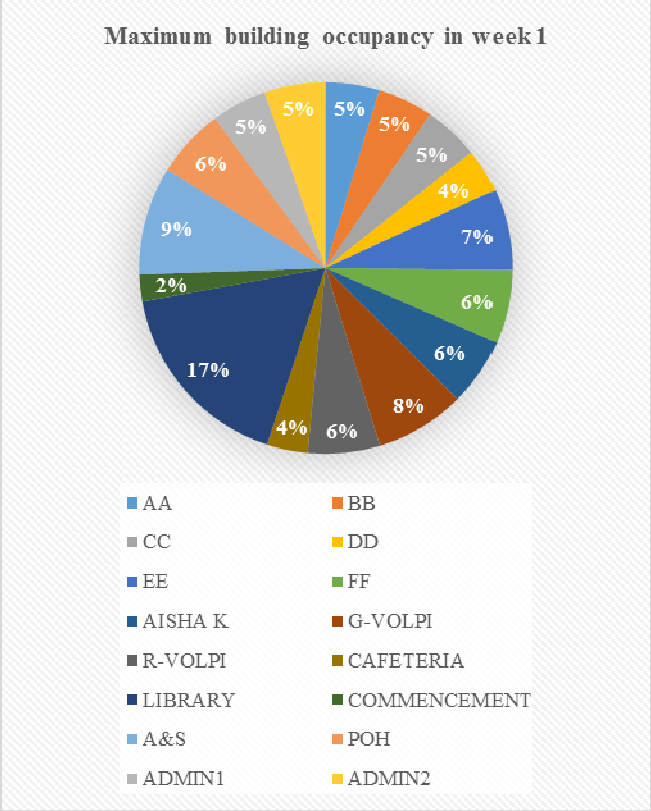


Fig. 5: Maximum occupancy values from week 1 for all buildings

The exceptions were CC, and Aisha Kande, and this could have been as a result of the network failure that occurred in the second week as noted above which resulted in no occupancy values being recorded. The average occupancy in the residential buildings over the two weeks' period also increased in all the dormitories except CC, DD, AISHA KANDE, and R-VOLPI were no records were obtained due to the network failure that occurred. Two of the academic buildings (RP library and arts and science) experienced decreased maximum occupancy in the second week when compared with the first week. Again, this is because classes generally stop holding a week before the final exams.

Average occupancy in the arts and science building decreased in the second week when compared with the first

week, however, the reverse was the case for the RP library. What this implies is that the average number of students that have their classes in the arts and science building is less than the average number of students that spend their time studying in the building during revision week. This is probably because the RP library building is where students usually spend their time studying which explains the increase in average occupancy in the second week when compared with the first week.

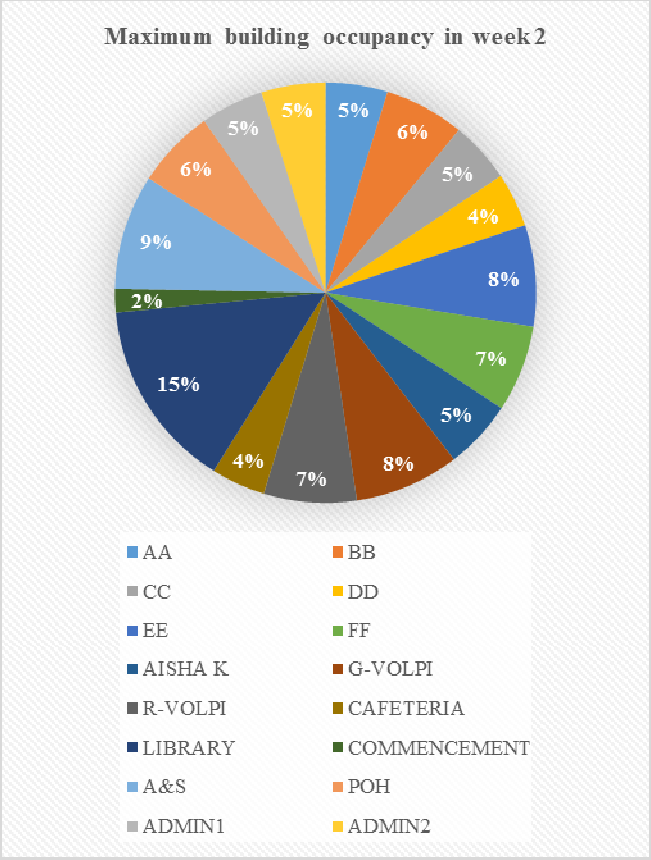


Fig. 6: Maximum occupancy values from week 2 for all buildings

To describe the pattern of occupancy in the buildings, graphs were plotted using minimum and maximum occupancy values over the two-week period. As shown in Figure 7, there is decreased occupancy in the RP library during the weekend, however, occupancy on Saturday is higher than on Friday and Sunday. This is because classes are held on Saturdays for graduate students in the RP library. In the second week (the week before exams), occupancy on Saturday is lower than on Friday and Sunday as classes have stopped holding.

As also observed, there is decreased occupancy in the second week in the RP library when compared to the first week, during the weekday from Monday to Thursday, as classes have generally stopped holding in preparation for the final exams. This results in a graphical form identical to that of a sine wave. With reference to time as illustrated in Figure 8, as expected over the two weeks there was higher occupancy in the day than at night with peak occupancy in the afternoon between 12pm and 1pm. This result is consistent with other works such as [39]. There is a gradual fall afterwards until after 7pm when the library experiences a slight increase in occupancy as a result of students coming for night studies. This peaks at 9pm after which it gradually reduces.

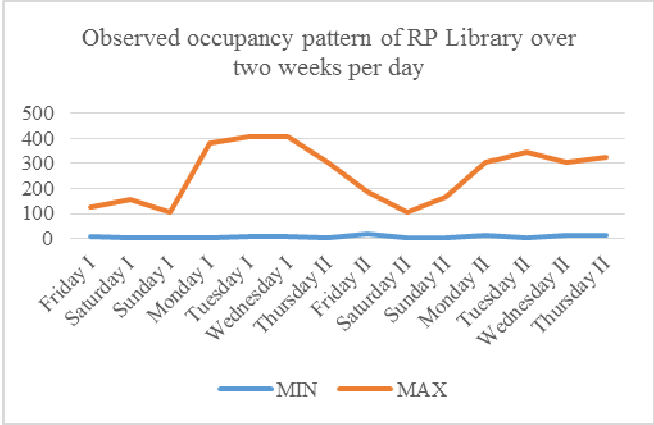


Fig. 7: Occupancy pattern of Library over two weeks per day

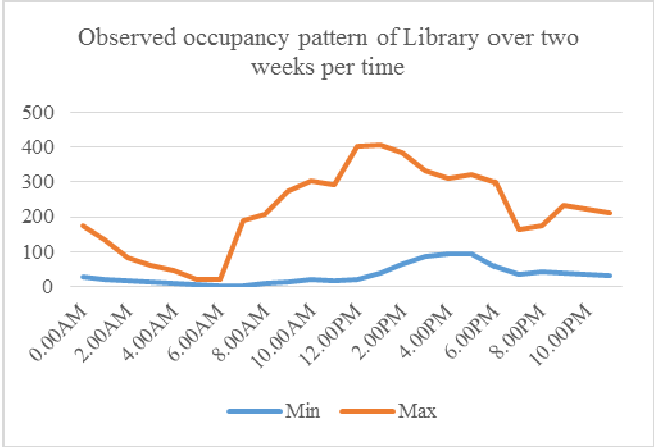


Fig. 8: Occupancy pattern of Library over two weeks per time

A similar pattern as the library occupancy was observed in the administrative buildings (as shown in Figure. 9) with an analysis on admin 1 showing increased occupancy during the weekdays and a near zero occupancy in the weekend over the two-week period. This is expected as the working days of the university are Mondays to Fridays while no work is generally done in the academic offices on Saturdays and Sundays. During the day, admin 1 experiences its highest occupation at 12pm after which it declines at 2pm around lunch time. Again, it increases slightly at 3pm and then drops at 4pm (Figure 10), an hour before closing work time.

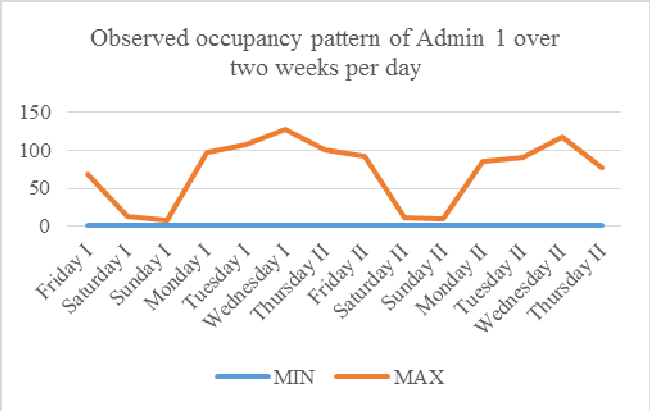


Fig. 9: Occupancy pattern of Admin 1 over two weeks per day

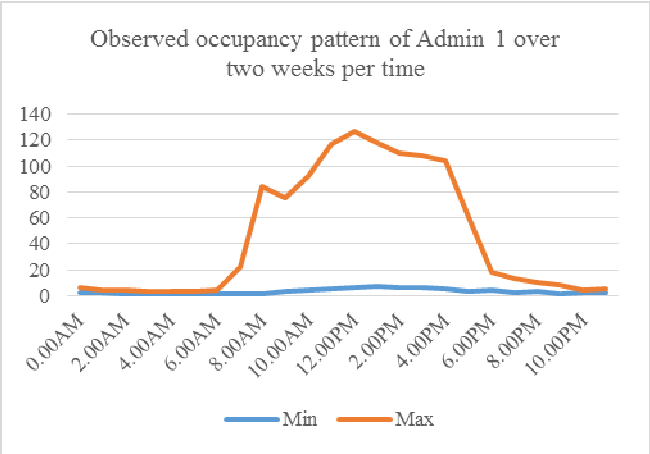


Fig. 10: Occupancy pattern of Admin 1 over two weeks per time

The pattern of occupancy observed in the dormitories showed a uniform trend for the two-week duration with Sunday experiencing the highest occupancy for a weekday as shown in Figure 11. This showed this result because no classes are generally held on Sundays, hence the students spend most of the day in their dormitories engaging in one leisure activity or the other.

With reference to time (Figure 11), the observed pattern revealed that the dormitories experience the highest occupation at night time between 10pm and 1am. The fall in occupancy after this time could be as a result of reduced activity and mobility among students as they go to bed. It is assumed that some students shut down some of their devices like their laptops when they go to bed. Also, everyone returns to their various rooms from the common areas as night falls and they go off to bed.

From the analysis of the dataset, it was observed that the cafeteria experienced peak occupancy during the day time at 1pm and during evening between 6pm and 7pm. Lunch and dinner are served in these periods respectively. This is illustrated in Figure 13.

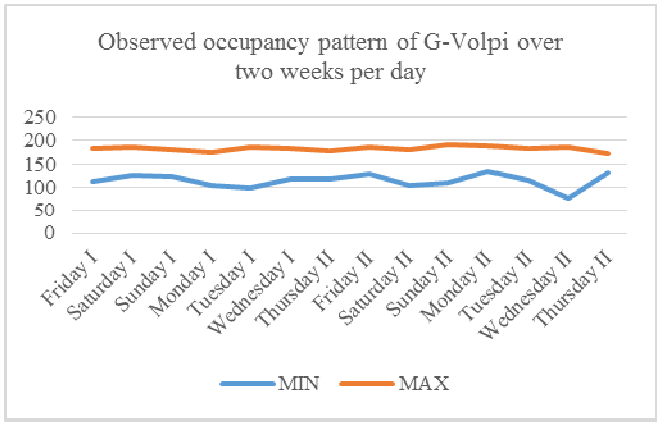


Fig. 11: Occupancy pattern of G-Volpi over two weeks per day

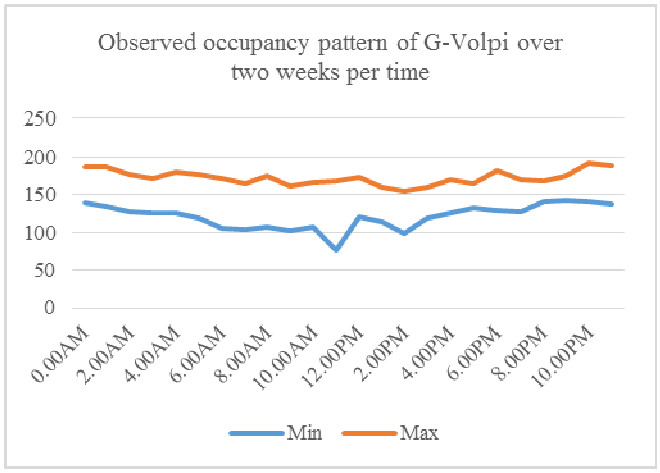


Fig. 12: Occupancy pattern of G-Volpi over two weeks per time

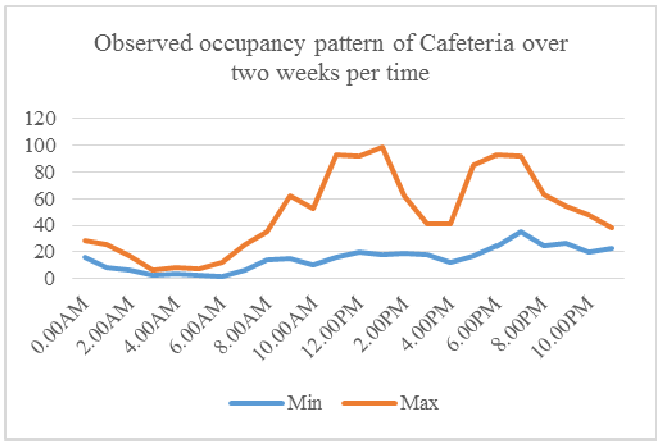


Fig. 13: Occupancy pattern of Cafeteria over two weeks per time

The aim of this thesis was to explore what insights could be obtained from network log data through the use of a visualization approach. In trying to achieve this aim, this study proposed a research model based on the organizational learning model which was used to guide the conduct of data interpretation by providing meaning to the collected data. In

the process of analyzing the data, new knowledge was obtained about occupants' behavior and this knowledge was used to perform further analysis. The results obtained in this study are consistent with similar works in user behavior literature. As the first study on WiFi usage trend in a Nigerian university and also the first study applying a model based on a social science theory, this thesis provides vital implications.

Nutt [22] notes that the primary function of facility management is to effectively manage resources. Also, Petrenko [24] asserts that insights from understanding occupant behavior are applicable in improving safety controls, emergency operations, and building management. Based on these claims, recommendations will be made to inform AUN management on how to improve its facility management. The findings from this study can help the management:

i. To enhance disaster management preparation and response

One important area that the insights from this study can be beneficial to AUN is in disaster management preparation. According to Akter and Wamba [3], disaster preparedness involves two main objectives; to allocate resources and to formulate emergency procedures to save lives, properties, and the environment. This study has revealed that students use the academic facilities the most during the day time between 12pm and 1pm, and also in the evening from 7pm to 11pm, extending into 2am during exam periods. The university's security team could take this into consideration when conducting vulnerability assessments to determine when the university community is most vulnerable. Also, decisions such as allocating security personnel for patrol duty could be better coordinated to ensure that they are properly staffed during these peak periods to promptly respond to any incidents that may occur. These peak periods could more so be used to conduct disaster preparedness activities such as a fire evacuation drill to ensure maximum participation from staffs and students. The insights obtained from this study could further help the university with providing timely response in the event of an incident on campus such as a building collapse or fire incident. Having information, backed by data, about the estimated usage of a building at a particular time and during a particular period could inform the disaster management team on where to focus its scarce resources for maximum efficiency.

ii. To improve facility management

With regards to enhancing building management, having information about less frequently visited buildings on campus such as commencement, cafeteria, admin 1 and admin 2 could inform decisions that would ensure efficiency in energy consumption to be taken. Such decisions could include the use of energy-saving bulbs that use motion sensors to detect the presence of people in a place, so that they go off automatically whenever a building is unoccupied for a certain period [40]. To ensure proper facility maintenance, periods when the buildings experience peak occupancy should be provided with adequate toiletries and sanitation to prevent the spread of bacteria, resulting in poor health conditions. Unlike many Nigerian public institutions, overcrowding is not a problem that is experienced at AUN. Nevertheless, close to exam periods the academic facilities can become much occupied resulting in difficulties in finding available study rooms. Although there are often times empty spaces, they usually have faulty switches or none at all. The FM team can ensure that as exams approach, all faulty switches undergo repairs to ensure increased ease of using the academic facilities.

iii. To automate the process of obtaining occupants count in the library

Every day, mostly during the evenings and at nighttime, students doing work study in the library walk around to take a count of the total number of people on each of the three floors. This is then collated and documented in a spreadsheet. This study has shown that this process can be automated through the use of occupancy monitoring resulting in efficiency.

iv. To guide facility management team's decision-making

Currently, when the FM team at AUN schedules repairs or power shutdowns, they do so without any information that guides their decision-making process to ensure that the activities of the university community is minimally disrupted. This study has found that the academic and office buildings are the least occupied in the morning hours between 3am and 6am. Thus, this would be the most suitable time for the FM team to carry out repairs or maintenance that would affect these buildings. Repairs at the dormitories could be scheduled between 11am and 1pm on weekdays when the majority of students are found to be in the classroom.

v. To enable increased student satisfaction

This study has shown that during the weekends, particularly on Sundays, and when exams are not in proximity, students spend most of their time in their hostel accommodations. For increased student satisfaction in AUN, activities could be organized during these periods to encourage learning and socializing amongst students.

vi. To ensure network optimization

Network optimization is a popular area where insights from occupancy information can be applied. Leading to exams, students use internet facilities more than usual to conduct research work and to study online materials such as educational videos. Increased bandwidth could be allocated during this time to ensure seamless flow of traffic.

V. CONCLUSION

This thesis obtained insights about the pattern of building occupancy in AUN for two weeks, showing that students and staffs frequented the academic and administrative facilities during the day time with peak occupancy witnessed between 12pm and 1pm. The graphics presented also showed that with proximity to exams, the academic buildings had reduced occupancy during the day time, with increased occupancy at night time between 7pm and 2am. The residential buildings had an even distribution over time, but experienced maximum occupancy on Sundays. Based on these findings, recommendations have

been made on ways that facility management can be improved in AUN.

Some limitations observed in this study could serve as justification for future research work. One limitation was that the period of observation was for two weeks, towards the end of the semester. Hence, it might not be adequate to make generalizations about the behavior of occupants. Another limitation was that this study did not use any measuring technique, such as the use of questionnaires, or tool, such as the use of a camera, to confirm or verify the results obtained.

This study analyzed network log data containing information about building occupancy by a university community for two weeks. Future work could explore occupancy behavior over a longer period of time to obtain more generalizable insights. Secondly, to verify the results, a technique involving the use of questionnaires or interviews could be used. Finally, future work could explore the application of insights from occupancy behavior to provide smart building and smart campus services.

REFERENCES

[1] M. Afanasyev, T. Chen, G. M. Voelker, A. C. Snoeren, "Analysis of a mixed-use urban wifi network: When metropolitan becomes neapolitan," In *Proceedings of the 8th ACM SIGCOMM Conf. on Internet Measurement*, ser. IMC '08. New York, NY, USA: ACM, 2008, pp. 85–98.

[2] M. Afanasyev, T. Chen, G. Voelker, and A. Snoeren, "Usage patterns in an urban wifi network," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 5, pp. 1359–1372, Oct 2010.

[3] S. Akter, S. F. Wamba, "Big data and disaster management: a systematic review and agenda for future research," *Annals of Operations Research*, pp. 1–21, 2017.

[4] A. A. Aliyu, M. U. Bello, R. Kasim, D. Martin, "Positivist and non-positivist paradigm in social science research: Conflicting paradigms or perfect partners," *J. Mgmt. & Sustainability*, vol. 4, pp. 79-95, 2014.

[5] I. P. Asiabaka, "The need for effective facility management in schools in Nigeria," *New York science journal*, vol. 1, no. 2, pp. 10-21, 2008.

[6] M. Balazinska, P. Castro, "Characterizing mobility and networkusage in a corporate wireless local-area network," In *Proceedings of the 1st Intl. Conf. on Mobile Systems, Applications and Services, MobiSys* '03, New York, NY, USA: ACM, 2003, pp. 303–316.

[7] G. Biczok, S. D. Martínez, T. Jelle, J. Krogstie, "Navigating MazeMap: indoor human mobility, spatio-logical ties and future potential," In *IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS),* 2014*,* pp. 266-271.

[8] D. P. Blinn, T. Henderson, D. Kotz, "Analysis of a Wi-Fi hotspot network," In *Proceedings of the Intl. Workshop on Wireless Traffic Measurements and Modeling (WiTMeMo '05),* USENIX Association, June 2005, pp. 1–6.

[9] F. Calabrese, J. Reades, C. Ratti, "Eigenplaces: Segmenting space through digital signatures," *Pervasive Computing, IEEE*, vol. 9, no. 1, pp. 78–84, Jan 2010.

[10] Chen, M., Ebert, D., Hagen, H., Laramee, R. S., van Liere, R., Ma, K. L., ... & Silver, D, "Visualization Viewpoints-Data, Information, and Knowledge in Visualization," *IEEE Computer Graphics and Applications*, vol. 29*,* no. 1, pp. 12, 2009.

[11] R. L. Daft, K. E. Weick, "Toward a model of organizations as interpretation systems," *Academy of management review*, vol. 9*,* no. 2, pp. 284-295, 1984.

[12] P Hebert, S. Chaney, "Enhancing facilities management through generational awareness," *Journal of facilities management*, vol. 9*,* no. 2, pp. 145-152, 2011.

[13] T. Hong, D. Yan, S. D'Oca, C. F. Chen, "Ten questions concerning occupant behavior in buildings: The big picture," *Building and Environment*, vol 114, pp. 518-530, 2017.

[14] J. J. Hox, H. R. Boeije, "Data collection, primary versus secondary," 2005.

[15] J. Kim, A. Helmy, "Analysing the mobility, predictability and evolution of WLAN users," In *International Journal of Autonomous and Adaptive Communications Systems*, vol. 7, no. 1-2, pp. 169-191, 2014.

[16] M. Kim and D. Kotz, "Modeling users' mobility among wifi access points," In *Papers Presented at the 2005 Workshop on Wireless Traffic Measurements and Modeling, WiTMeMo '05*. Berkeley, CA, USA: USENIX Association, 2005, pp. 19–24.

[17] D. Kincaid, "Integrated facility management," *Facilities*, vol. 12, no. 8, pp. 20-23, 1994.

[18] W. R. King, "Knowledge management and organizational learning," In *Knowledge management and organizational learning*, Boston, MA. 2009, pp. 3-13.

[19] D. Kotz, K. Essien, "Analysis of a campus-wide wireless network," *Wireless Networks*, vol. 11, no. 1-2, pp. 115-133, 2005.

[20] S. Lavy, "Facility management practices in higher education buildings: A case study," *Journal of Facilities Management*, vol. 6, no. 4, pp. 303-315, 2008.

[21] M. Lyles, M. P. Easterby-Smith, *The Blackwell handbook of organizational learning and knowledge management*. Blackwell, 2003.

[22] B. Nutt, "Four competing futures for facility management," *Facilities*, vol. 18, no. 3-4, pp. 124-132. 2000.

[23] M. Papandrea, K. K. Jahromi, M. Zignani, S. Gaito, S. Giordano, G. P. Rossi, "On the properties of human mobility," *Computer Communications*, vol. 87, pp. 19-36, 2016.

[24] A. Petrenko, S. Bell, K. Stanley, W. Qian, A. Sizo, D. Knowles, "Human spatial behavior, sensor informatics, and disaggregate data," In *International Conference on Spatial Information Theory*, Springer, Cham, 2013, pp. 224-242.

[25] A. Petrenko, A. Sizo, W. Qian, A. D. Knowles, A. Tavassolian, K. Stanley, S. Bell, "Exploring mobility indoors: an application of sensor-based and GIS systems," *Transactions in GIS*, vol. 18, no. 3, pp. 351-369, 2014.

[26] T. S. Prentow, A. J. Ruiz-Ruiz, H. Blunck, A. Stisen, M. B. Kjærgaard, "Spatio-temporal facility utilization analysis from exhaustive wifi monitoring," *Pervasive and Mobile Computing*, vol. 16, Part B, pp. 305–316, 2015.

[27] A. E. Redondi, M. Cesana, D. M. Weibel, E. Fitzgerald, "Understanding the WiFi usage of university students," In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC),* 2016, pp. 44-49.

[28] F. M. Rojas, K. Kloeckl, C. Ratti, "Dynamic City: Investigations into the sensing, analysis and application of real-time, location-based data," 2008.

[29] A. J. Ruiz-Ruiz, H. Blunck, T. S. Prentow, A. Stisen, M. B. Kjærgaard, "Analysis methods for extracting knowledge from large-scale wifi monitoring to inform building facility planning," In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2014, pp. 130-138.

[30] T. Scheytt, K. Soin, K. Sahlin-Andersson, M. Power, "Introduction: Organizations, risk and regulation," *Journal of Management Studies*, vol. 43, no. 6, pp. 1331-1337, 2006.

[31] J. M. Sinkula, "Market information processing and organizational learning," *Journal of marketing*, vol. 58, no. 1, pp. 35-45, 1994.

[32] K. Sookhanaphibarn, E. Kanyanucharat, "Empirical study of routine structure in university campus," In *International Conference on Online Communities and Social Computing,* Springer, Berlin, Heidelberg, 2013, pp. 201-209.

[33] K. Sookhanaphibarn, E. Kanyanucharat, W. Paireekreng, "A system for visualizing routine structure from WiFi access," *Procedia Manufacturing*, vol. 3, pp. 4828-4835. 2015.

[34] D. Steenhuizen, I. Flores-Colen, A. G. Reitsma, P. Branco Ló, "The road to facility management," *Facilities*, vol. 32, no. 1-2, pp. 46-57, 2014.

[35] D. Tang, M. Baker, "Analysis of a local-area wireless network," In *Proceedings of the 6th annual international conference on Mobile computing and networking,* New York, NY, USA: ACM, 2000, pp. 1-10.

[36] A. Thapaliya, S. Sengupta, J. Springer, "Wi-Fi spectrum usage analytics in University of Nevada, Reno," In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* 2018, pp. 1-6.

[37] S. C. van der Spek, E. Verbree, B. M. Meijers, F. J. Bot, H. H. Braaksma, R. C. Braggaar, ... B. R. Staats, "In_sight: Using Existing Wi-Fi networks to Provide Information on Occupancy and Exploitation of Educational Facilities using at Delft University of Technology," In *Proceedings of the 13th International Conference on Location Based Services: LBS,* 2016, pp. 151-156.

[38] X. Wei, N. C. Valler, H. V. Madhyastha, I. Neamtiu, M. Faloutsos, "A behavior-aware profiling of handheld devices," In *2015 IEEE Conference on Computer Communications (INFOCOM),* 2015, pp. 846-854.

[39] E. Zola, F. Barcelo-Arroyo, "A comparative analysis of the user behavior in academic WiFi networks," In *Proceedings of the 6th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks,* ACM, 2011, pp. 59-66.

[40] H. Zou, Y. Zhou, H. Jiang, S. C. Chien, L. Xie, C. J. Spanos, "WinLight: A WiFi-based occupancy-driven lighting control system for smart building," *Energy and Buildings*, vol. 158, pp. 924-938, 2018

Temporal Variation of Duty Cycle in the GSM Bands

[1]Ganiyu A. Ishola, [1]Balogun V. Segun, [1]Nasir Faruk, [2]Kayode S. Adewole, [1]Abdulkarim A. Oloyede, [1]Lukman A. Olawoyin and [2]Rasheed.G. Jimoh

[1]Department of Telecommunication Science, University of Ilorin
[2]Department of Computer Science, University of Ilorin, Nigeria

Abstract*—in a quest to efficiently manage the scarce spectrum, researchers, governments and regulatory bodies have been conducting spectrum occupancy measurements in various countries across the world. This paper discusses the analysis inferred and the results arrived at from the research conducted on a long-term spectrum occupancy measurement in the GSM bands at a fixed location in the Faculty of Communication and Information sciences, University of Ilorin, Nigeria. The measurements data were collected for 4 weeks continuously for a time span of 24hours. The data gathered were resourceful for the calculation of the temporal variation of duty cycle which resulted into average spectrum occupancy of 10% and 9% for both the GSM 900 and 1800 respectively. The results showed some form of temporal variation due to the varying degree of activities in days and nights, weekdays and weekends. This puts up an effort to develop spectrum usage repository for big data analytics. This would help researchers and policymakers to understand the activities going on in the spectrum bands for them to devise intelligent spectrum models for effective utilization.

Keywords: duty cycle; spectrum occupancy; big data; GSM; long term

I. INTRODUCTION

The advent of wireless communication coined with an avalanche of multimedia services and wireless gadgets majorly contribute to the radio spectrum Several campaigns [4 - 6] on spectrum usage have been performed and it is concluded that the radio spectrum is underutilized due to the fact that the primary users are not always active using the band allocated to them. The scarcity. Unlike other natural and precious resources such as land, water, gas etc., the spectrum is, however, reusable [1] [2]. To avoid pollution in the radio spectrum and increase its benefits, regulatory bodies such as the Nigerian Communications Commission (NCC) are assigned to manage the spectrum. The Command and Control Model (CCM) used by most regulators allows the licensed users to explore the bandwidth exclusively, with no interference. Unfortunately, the static nature of allocation of spectrum rendered the spectrum to be underutilized, thereby increasing the level of its scarcity. In oppose to the current approach, flexible spectrum sharing paradigm, regarded as Dynamic Spectrum Access (DSA) was introduced to improve on spectrum usage. This paradigm has paved ways for increased utilization by allowing not only the users with licensed but also the opportunistic users to use the temporarily vacant radio spectrum in a perfectly controlled manner. In this approach, the unlicensed users, sometimes regarded as secondary users, scan the spectrum in a search for white spaces (i.e. unused spectral spaces), also regarded as spectrum holes, in order to opportunistically enjoy the radio spectrum with no interference suffered by the primary users. The opportunistic users must, however, be capable of sensing when a band is busy in order to vacate it or idle in order to explore it [3]

idea of cognitive radio is a promising attempt to bridge the space between underutilization and spectral scarcity [2]. CRs are designed to be intelligent with the ability to sense the busy period and the idle period of the

spectrum. It's imperative that one measures the occupancy level of some spectrum to arrive at a particular pattern of usage by primary users. The degree (i.e., short-term or long-term) of spectrum occupancy measurement can be a trade-off between fast result and accuracy as a long-term measurement (e.g., a month or more than) will be enough measurement to determine temporal spectrum gaps unoccupied by the primary users for the implementation of cognitive radios.

Some measurement had been taken over a short period of time say a day, 3 days, 7 days and even more [4], but in this research, the measurements were conducted for over three weeks so that we can have a large stream of data for decision making. This would help researchers to understand the activities going on in the spectrum band so that they can device intelligent spectrum models for effective spectrum utilization. Alongside this, it will also help policymakers in determining which band have low occupancy so that adequate strategies can be put in place for its effective harnessing [7].

II. RESEARCH METHODOLOGY

A. Measurement Setup

In setting up the measurement for the purpose of this campaign, an Agilent N93242C Handheld Spectrum Analyzer (HAS), which spans through the frequency 100 kHz and 7 GHz (tunable to 9 kHz), played a vital role. A retractable whip antenna with a range of 70 MHz – 1000 MHz is connected to the Spectrum Analyzer. The antenna is an omnidirectional antenna which is capable of receiving signals from every angle of the site. The data received from the spectrum analyzer is directly measured in dBm. The pre-amplifier feature of the spectrum analyzer is turned on to enhance the overall sensitivity. The attenuation level of the spectrum analyzer is set at 0 dB in order to detect weak signals. A functioning GPS (Global Positioning System) is connected to the Agilent device to determine the location features of the site. With the help of an external hard drive (1 Terabyte), the log files of the spectral activity which is generated in real-time are stored on the external drive which will be further saved on a laptop for subsequent processing and analysis. Same measurement configurations were maintained for all the bands considered throughout the period of the measurements. Although, some configurations used were based on default setting such as Resolution bandwidth (RBW), number of points, Sweep Time etc. The backup of the spectrum analyser must not be exhausted to ensure the battery is constantly up because any configurations performed on the spectrum analyser will be lost once the spectrum analyser goes off.

Fig. 1. Agilent N9342C Spectrum Analyser Setup in the TCS Laboratory

B. Measurement Location

The measurements took place in the Telecommunication Science Laboratory located at the top floor of Faculty of Communication and Information Sciences (FCIS) building, University of Ilorin, Nigeria with coordinates (Longitude: 4° 40' 28''E Latitude: 8° 29' 19''N). These measurements were conducted for 4 weeks for a period of 24 hours continuously. The service bands considered are GSM 900 MHz and GSM 1800 MHz as shown in Table 1. Only downlink transmissions were captured. These bands are expected to experience enough spectral activity as owing to the fact that these UHF bands are vastly used almost everywhere in the world.

Table 1: Service Bands Measured

Service bands	Frequency Range (MH	Bandwidth (MHz)
GSM 900	925 – 960	35
GSM 1800	1805 – 1880	75

C. Data Collection and Processing

Irrespective of the bands currently being measured, at the initial stage, the raw data collected were represented in matrix forms using the elements of received signal powers. The signal powers are denoted with $P(t_i, f_j)$ with the unit of dBm. Where f_j represents the channel and t_j represents the time frame which is usually 461 points (m) as used in the measurements. A three-stage process is used in evaluating the occupancy statistics and they are as follow: inputting of raw data, adaptive detection threshold, and computation of average duty cycle of every channel. Raw data inputs are unstructured received power levels at the antenna end. All the sets of data are in a matrix of $i \times j$ with the elements of $P_{(ti,fj)}$ = Y (i.e. Y is a matrix of sampled power) as:

$$Y = \begin{bmatrix} P_{(ti,fj)} \dots & \dots P_{(ti,fj)} \\ \vdots & \vdots \\ P_{(ti,fj)} \dots & \dots P_{(ti,fj)} \end{bmatrix} \quad (1)$$

Equation (1) can also be written as:

$$Y = y(n) \quad 2)$$

Where y (n) is a matrix of received signal power at every point n. Every channel has a different noise power level, sometimes referred to as background noise, which varies from each band. An adaptive threshold is set for every band for every location. It is commonly known that when the received signal power is higher than the threshold set the frequency band is said to be occupied, otherwise, the band is considered idle.

D. Spectrum Sensing and Duty Cycle

In this experiment, energy detection method was used as the spectrum sensing method. This is because it is the most widely used sensing technique and it is arguably the simplest since it operates by requiring no prior knowledge of the primary user. It performs its operations by setting a threshold which is compared with the received signal power [9]. The chosen threshold relies on noise power. This method is sometimes referred to as period gram or radiometry [9]. The resolution on the occupancy band can be attained by equating the decision metric, M, to a fixed detection threshold λ_E [9]. This is corresponding to distinguishing the two following hypotheses [10];

$$H_1: x(n) = s(n) + w(n) \quad (3)$$
$$H_0: x(n) = w(n) \quad (4)$$

Where *s(n)* is the signal sent by the licensed users, *x(n)* is the signal received by the opportunistic users, *w(n)* is known as the additive white Gaussian noise. Hypothesis H_0 represents the absence of a licensed user. Meanwhile, hypothesis H_1 depicts the presence of the licensed user. The calculation of energy is mathematically expressed as shown below [9][10]:

$$\sum_{n=0}^{N} |x(n)| \quad (5)$$

The importance of duty cycle can never be overlooked in this research as it is majorly used in characterizing the spectrum utilization. Duty cycle is always based on the frequency calibrated on the Agilent device (Spectrum Analyser) [11]. Basically, it is just how often a channel is busy or idle during the sample period. The duty cycle is usually represented in percentage (%) to indicate the occupancy rate of a channel or frequency band [12].

$$\text{Duty Cycle} = \frac{\text{Signal Occupation Period (n)}}{\text{Total Observation Period (m)}} \times 100\% \quad (6)$$

Where m is the period of observed measurement and n is the duration of the signal. With a given time series of channel power measurements, the average duty cycle can be computed using equation (5). Temporal variation can be defined as the time evolution of duty cycle at a certain frequency point as it varies with time. In this experiment, the temporal variation was evaluated in two ways: the short-term which is based on hourly-wise spectrum occupancy measurement and the long-term which is based on minute-wise spectrum occupancy measurement was neglected.

III. RESULTS AND DISCUSSION

The spectrum occupancy rate was determined through the computation of average duty cycle value. Figures 2 and 3 provide the spectrogram of the spectrum usage for the GSM 900 MHz and GSM 1800 MHz bands. The spectrograms show the time variance of the duty cycle.

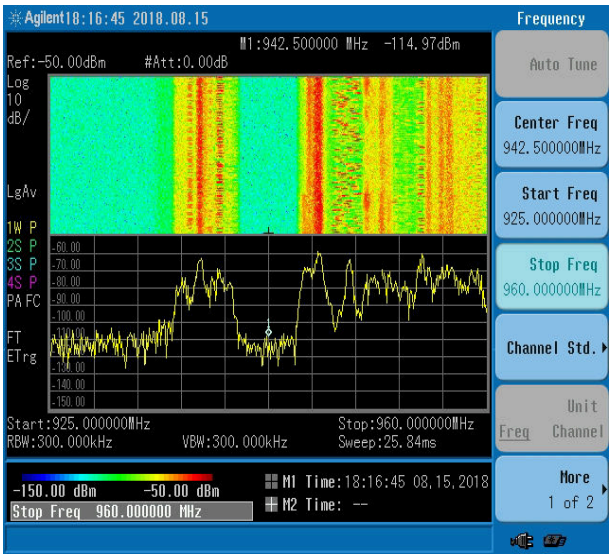


Fig. 2: Downlink Configuration & Spectrogram of GSM 900

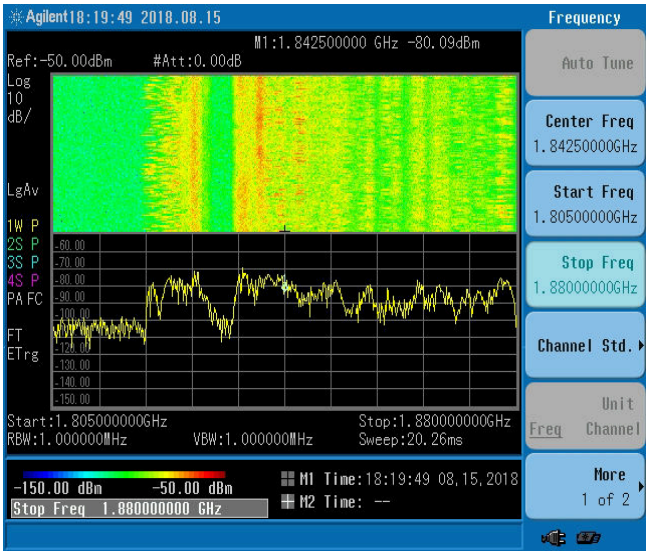


Fig 3. : Downlink Configuration & Spectrogram of GSM 1800

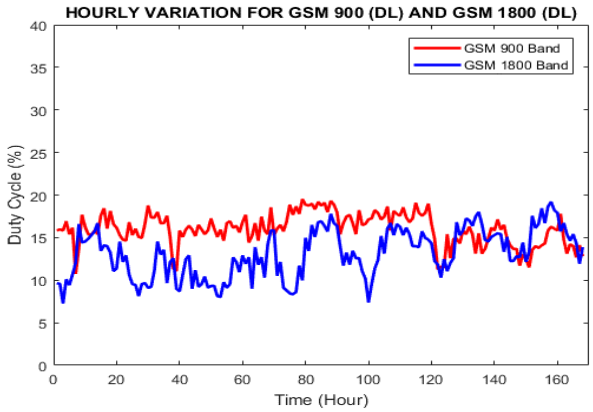


Fig.4: Hourly variation of GSM 900 and GSM 1800 (Downlink)

In Figure 4, hourly duty cycle of a 7-day measurement data is depicted in the graph for both GSM 900 and GSM 1800 bands. It can be observed from the graph that both the GSM 900 and GSM 1800 follow almost the same pattern. This is due to the fact that both bands are used by same technology, 2G, and they both perform the same operation. The major difference between the two is the frequencies at which they transmit and receive. According to the graph, GSM 900 appears to have higher occupancy rate than GSM 1800. This is the hourly-wise temporal variation of duty cycle and it is short-time because the computation is based on hourly data rather than minute-wise data. We arrived at the average duty cycle of 15% for the GSM 900 band and 12% for the GSM 1800 band.

IV. CONCLUSIONS

This research has discussed spectrum occupancy measurements conducted in the GSM bands and has determined the temporal variation of duty cycle in the GSM bands. It was observed that both the GSM 900 and GSM 1800 follow almost the same pattern in terms of spectrum usage with occupancy values of 15% and 12% respectively. Even though the hourly variation provides a clear variance of the occupancy rate, the minute-wise assessment is still desirable to adequately generate more data set for the development of predictive models such as the neural network.

REFERENCES

[1] K. Patil, K. Skouby, A. Chandra and R. Prasad, “Spectrum occupancy statistics in the context of cognitive radio,” 2011 The 14th International Symposium on Wireless Personal Multimedia Communications (WMPC), Brest, 2011, pp. 1-5.

[2] M. López-Benítez and F. Casadevall "Spectrum Occupancy in Realistic Scenarios and Duty Cycle Model for Cognitive Radio" Journal of Advances in Electronics and Telecommunications, April 2010, 1(1) pp. 26-34.

[3] B. Najashi, F. Wenjiang, and M. Almustapha, “Spectrum Hole Prediction Based on Historical Data: A Neural Network Approach,” in College of Communication Engineering, Chongqing University, Chongqing, China, Jan. 2014, pp. 1–8.

[4] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, “Chicago Spectrum Occupancy Measurements & Analysis and a Long-term Studies Proposal,” In Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum, Boston, USA, August 2006, p. 1. ACM

[5] M. Wellens, J. Wu, and P. Mähönen, “Evaluation of Spectrum Occupancy in Indoor and Outdoor Scenario in the Context of Cognitive Radio,” 2007 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communication, Orlando, FL, USA, 2007, pp. 420 - 427.

[6] T. Harrold, R. Cepeda and M. Beach. "Long-term measurements of spectrum occupancy characteristics." 2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Aechen, 2011, pp. 83-89.

[7] O. D. Babalola, E. Garba, I. T. Oladimeji, A. S. Bamiduro, N. Faruk, O. A. Sowande, O. W. Bello, A. A. Ayeni and M. Y. Muhammed, “Spectrum occupancy measurements in the TV and CDMA bands,” 2015 International Conference on Cyberspace (CYBER-Abuja), Abuja, 2015 pp. 192-196

[8] F. Wenjiang, and C. Kadri, “An Insight into Spectrum Occupancy in Nigeria,” International Journal of Computer Science, 2013, 10(1) pp. 394.

[9] I.A. Sikiru, et al., “Effects of detection threshold and frame size on duty cycle in GSM bands,” 2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON), Owerri, 2017, pp. 343-346

[10] M. Subhedar, and G. Birajdar, “Spectrum sensing techniques in cognitive radio networks: A survey,” International Journal of Next-Generation Networks, 2015, 3(2), pp.37-51.

[11] M. Wellens, A. De Baynast, and P. Mähönen, “Performance of dynamic spectrum access based on spectrum occupancy statistics,” IET communications, 2008, 2(6), pp.772-782.

[12] J. J. Lehtomaki, R. Vuohtoniemi, and K. Umebayashi, “On the measurement of duty cycle and channel occupancy rate,” IEEE Journal on Selected Areas in Communications, 2013, 31(11), pp.2555-2565.

Evaluating Ensemble Techniques for Predicting Students' Grades Based on their Interaction with Learning Management System

Umar Isa Usman
Department of Computer Science
Ahmadu Bello University
Zaria, Nigeria
umarusmanrambido@gmail.com

Salisu Aliyu
Department of Computer Science
Ahmadu Bello University
Zaria, Nigeria
aliyusalisu@abu.edu.ng

Barroon Isma'eel Ahmad
Department of Computer Science
Ahmadu Bello University
Zaria, Nigeria
barroonia@yahoo.co.uk

Yusuf Abubakar
Department of Computer Science
Nuhu Bamalli Polytechnic
Zaria, Nigeria
yusufabukausar@gmail.com

Abstract*—Ensemble classifiers have been found to be better and more accurate compared to individual classifiers that make them up. Data mining techniques have been used by many researchers to predict students' performance, but very few studies used ensemble techniques on datasets from a learning management system to predict students' performance. Furthermore, the study of the effect of maximizing the size of partitioned data subset on the accuracy of ensemble models is also not common in the literature. In this paper, both homogeneous and heterogeneous ensemble techniques were used to predict students' final grade on a real world dataset. The base classifiers used are Decision Tree(DT), Naïve Bayes(NB) and K-Nearest Neighbor(KNN) while the ensemble methods used are: Random Forest, Bagging, Boosting and Voting. Waikato Environment for Knowledge Analysis (WEKA) was used for data preprocessing, attribute selection and 10-fold cross validation. The result obtained shows that, Random Forest ensemble method using decision tree outperforms other ensemble methods with an average accuracy of 90.3%, then the bagging method using DT with an average accuracy of 85.3%, then the boosting method using DT, then bagging/boosting method using NB and finally, the voting ensemble method. The bagging/boosting method using KNN was found to be the least method in terms of accuracy and other measures. It was also observed that in the bagging method, maximizing the size of partitioned data subsets does not improve the ensemble accuracy.

***Keywords*—Learning Management System, MOODLE, Ensemble, Classifiers**

I. INTRODUCTION

Learning Management System (LMS) can offer a great variety of channels and workspaces to facilitate information sharing and communication among participants in a course. They let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning with forums, chats, file storage areas, news services among others. Examples of commercial learning management systems are Blackboard and TopClass, while free learning management systems include: MOODLE, Ilias and Claroline [1]. Performance of students may be influenced by several factors such as gender, age, parents socio economic status, area of resident, nature of school, medium of teaching, daily study hours or nature of accommodation [2].

Developing an effective student performance prediction model is very important but rather difficult task for educational institutions. The aim of students' performance prediction models is to predict whether or not a student will perform well or not. If the prediction models could not provide a certain high level of prediction accuracy rate, it will lead to making incorrect decisions and hence, making it difficult for education counsellors to properly counsel students.

Related literature and studies have shown that data mining techniques can be used by education data miners to trim down the students failing ratio by providing recommendations to educational system stakeholders e.g. students, teachers, researchers and administrators where these recommendations might have a significant impact in improving learning process [1-3]. It is also known from several studies that combining multiple data mining classification techniques or ensemble classifiers usually outperforms single classifiers[5].

Ensemble classifiers refers to a group of base classifiers that are cooperatively trained on a dataset in a supervised classification problem in order to improve the performance of classification models. The main discovery is that ensemble methods are often much more accurate than the individual classifiers that make them up [5] [6].

Although, several literatures have demonstrated the superiority of ensemble classifiers over the single base classifiers, but most of them constructed ensembles of specific types of base classifiers for example, ensembles on decision trees [7, 8] or only on homogenous ensembles [9] or solely on heterogeneous classifiers [10, 11].

Despite some previous works on comparing ensemble classifiers techniques, where results show that boosting method outperforms other ensemble techniques, they maintained that the performance of ensemble techniques is usually domain dependent [5, 7, 9, 10, 12]. So, in the domain of student performance prediction a comparative study of

student predictive models built using different ensemble techniques is not common in the literatures. This fact raises our research question concerning which ensemble classifiers is the best.

Furthermore, the study of the effect of maximizing the size of partitioned data subset on the accuracy of the bagging ensemble technique in predicting students' performance is also not common in the literature. Finally, how do we identify the most significant attribute that are major indicators of high performance from our dataset?

This work answers the above questions by developing and comparing different ensemble classifiers for students' performance prediction based on their interaction with MOODLE. Also, the effect of maximizing the size of partitioned data subsets size on ensemble accuracy was also investigated. The significant attributes which are indicators of high performance were also determined using the attributes selection evaluator in the WEKA tool environment.

II. Ensemble Techniques

Ensemble methods are categorized into homogeneous (the same classifiers) and heterogeneous (different classifiers) [13]. Voting and stacking are examples of heterogeneous ensemble technique while Bagging, Boosting and Random Forest are examples of homogeneous ensemble technique. The Homogenous ensemble technique is divided also, into dependent and independent methods. In a dependent method, the output of a base classifier is used in the creation of the next classifier. Boosting is an example of dependent method. In an independent method, each classifier performs independently and their outputs are obtained using majority vote for classification problems or averaging for regression problems. Fig. 1 shows the types of ensemble techniques.

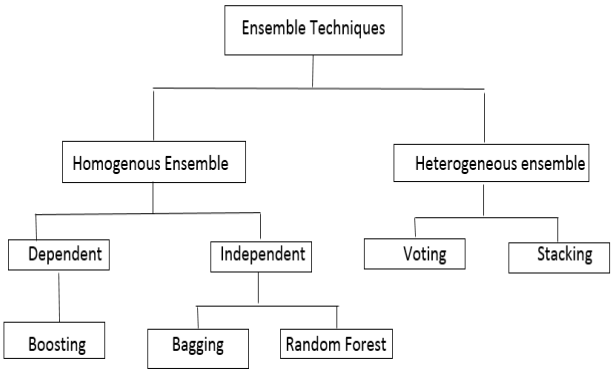


Fig. 1. Types of ensemble techniques

Bagging and random forest are examples of independent methods. These techniques resample the original data into samples of data, and then, each sample are then trained by a different classifier, individual classifiers results are then combined through a voting process, the class chosen by most number of classifiers is the ensemble decision [6].

A. Base Classifiers

Here, we introduce briefly the base classifiers that are used for constructing ensemble classifiers using any of the above approaches.

1) Decision Tree: Decision Tree builds classification model in the form of a tree structure and generate rules (conditional statements) that can easily be understood by humans and easily used within a database to identify a set of records. Most commonly used DT algorithms by researchers in EDM are: Iterative Dichotomiser3 (ID3), Classification and Regression Trees (CART) and C4.5 Algorithms.

ID3 is a recursive algorithm that employs a top down greedy search through the space of possible branches with no backtracking. ID3 simply uses a fixed set of examples to build a Decision Tree, and later the developed DT will be employed to classify new future samples [14]. The ID3 constructs a decision tree based on information gain/entropy measures [15]. According to [16] when datasets are split to grow decision trees, it is done to reduce impurity and entropy is one way to measure the degree of impurity given by the equation 1.

$$Entropy = \sum_j P_j \log_2 P_j \quad (1)$$

Where P_j is the probability of each class j in the dataset. The Information gain on splitting on an attribute say A, is the difference between the degree of impurity of the parent dataset say S and the weighted summation of impurity degrees of the subset dataset Si split on attribute A with values v mathematically described as:

$$Information\ Gain(S, A) = H(S) - \sum_v \frac{|Sv|}{|S|} H(Sv) \quad (2)$$

Where H(S) is the entropy of the parent dataset or node S, $H(S_v)$ is the entropy of the subset split base on values v of attribute A and |S| is the total number of entries in dataset S.

CART was introduced with reference to classification and regression trees, so that it can respectively handle categorical and continuous variables in building decision trees. Classification trees are used to identify the class that a categorical target variable would likely fall into. On the other hand, regression trees handle prediction of continuous target variables. CART works through the recursive partitioning of the training set in order to obtain subsets that are as pure as possible to give the target class. Gini Index is the default impurity measure used in CART for categorical target variables. It is essentially a measure of how well the splitting rule separates the classes contained in the parent node [17]. Given a training dataset S and that the target attribute takes on j different values, then the Gini index of S is defined as:

$$Gini(S) = 1 - \sum_{i=1}^{j} (P_i)^2 \quad (3)$$

Where Pi is the probability of S belonging to the class i. A Gini split measures the divergences between the probability

distribution of the target attributes values which is achieved by selecting the attribute with the maximum gain. The gain by a Gini split on a dataset S and attribute A is given as:

$$Gini\ split(S, A) = Gini(S) - \sum_{i=1}^{j} (|S_i|/|S|)\ Gini(S_i) \quad (4)$$

Quinlan [4] developed C4.5, as an extension of ID3 algorithm in order to handle problems associated with ID3. C4.5 generates decision trees using an approach whereby each node splits the classes based on the gain of information. C4.5 accepts both continuous and discrete features. It can also handle incomplete data points, as well as, over-fitting problem by a clever bottom-up technique usually known as "pruning".

2) Naïve Bayes: The NB algorithm is a machine learning algorithm that is used for solving classification problems. It is called "Naïve' because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. It is a probabilistic classifier that is based on the Bayes Theorem which states that: "The probability of the event A given the event B is equal to the probability of the event B given A multiplied by the probability of A upon the probability of B" [2]. The theorem is mathematically expressed as:

$$P(A \mid B) = (P(B \mid A) \times P(A))/P(B) \quad (5)$$

In any classification problem, there are multiple features and classes. According to [3] the aim of the Naïve Bayes classifier is to compute the conditional probability of an object with a feature vector say $x_1, x_2, \cdots, x_n$ belonging to a particular class C_i, for $1 \leq i \leq k$ expressed mathematically as:

$$P(C_i|x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_n|C_i) \times P(C_i)/P(x_1, x_2, \ldots, x_n) \quad (6)$$

Equation 6 can be reduced to:

$$P(C_i|x_1, x_2, \ldots, x_n) = \left(\prod_{j=1}^{j=n} P(x_j|C_i)\right) \times P(C_i)/P(x_1, x_2, \ldots, x_n) \quad (7)$$

So, Naïve Bayes classifiers are easy to build and useful particularly for very large dataset, along with simplicity.

3) K-Nearest Neighbor: KNN is a non-parametric classification algorithm used to predict a target label by finding the nearest neighbor classes [5]. K in KNN refers to the number of nearest neighbors the classifier will use to make its predictions. KNN stores all available cases and classifies new cases based on a similarity measure of the nearest neighbors. The nearest neighbors are computed using any of the distance measures like Euclidean distance, Manhattan distance, Hamming distance etc.

So, given a dataset and an unknown data point, we compute the distance of the unknown data point from all the points in the dataset. The class of the majority neighbors is then selected for the unknown data.

KNN is usually used when there are non-linear decision boundaries between classes.

B. Ensemble (Combination) Methods

1) Bagging: Bagging stands for bootstrap aggregation (a technique for re-sampling data into subsets). The idea behind this method is to increase the accuracy of unstable classifiers by creating a composite classifier, and then combine the results of the learned classifiers into a single prediction [6]. Breiman [7] showed that bagging is effective on "unstable" learning algorithms where small changes in the training set result in large changes in predictions. It means bagged ensembles tend to improve upon their base models more if the base model learning algorithms are unstable and differences in their training sets tend to induce significant differences in the models.

2) Boosting: Boosting belongs to a family of algorithms that are capable of converting weak learners to strong learners. The general boosting procedure is simple. It trains a set of learners sequentially and combines them for prediction, then focuses more on the errors of the previous learner by editing the weights of the weak learner. A specific limitation of boosting is used only to solve binary classification problems. This limitation is eliminated with the AdaBoost algorithm. AdaBoost is an example of boosting algorithm, which stands for adaptive boost. The idea behind this algorithm is to pay more attention to patterns that are hard to classify. The amount of attention is measured by a weight that is assigned to every subset in the training set. All the subsets are assigned equal weights (uniform probability) initially, in each iteration the weight of wrongly identified instances is increased while the weights of truly identified instances are decreased. Then, the AdaBoost ensemble combines the learners to generate a strong learner from weaker classifiers through a voting process [6].

In boosting, as contrary to bagging, each classifier is influenced by the performance of the previous classifier. In bagging, each sample of data is formed using uniform probability, while in boosting, instances are chosen with a probability that is proportional to their weight. Furthermore, bagging works best with high variance models which produce variance generalization behavior with small changes to the training data. Decision trees and neural networks are examples of high variance models [8].

3) Random Forest: Random Forest proposed by [9] is a popular and powerful ensemble supervised algorithm which is capable of performing classification and regression tasks. The algorithm creates a forest of random trees (DT) taking different sub-samples (bootstrap) from the original data at

training time and outputting the class, that is, the mode of the classes output by individual trees, in order to overcome over-fitting problem of weak DT. Over-fitting means a model fits well to training dataset while fails to the validation dataset, i.e. the model memorizes the features in the training data set instead of learning the patterns which prevent it from being able to generalize to the test data. Random forest uses best set of predictors when splitting the data set rather than considering all predictors. RF is different from bagging that uses full set of predictors while splitting the data or building the models. This strategy make the RF turns out to perform very well compare to many other classifiers including discriminant analysis, support vector machines and neural networks [9].

III. RELATED WORK

Table 1 lists related works that had developed classifier ensembles using several base classifiers. The types of base classifiers used, the ensemble techniques used and the purpose of developing such ensemble classifier are used as the basis for comparing the related works.

TABLE 1. COMPARISONS OF RELATED WORKS

Work	Classification Techniques	Ensemble Technique
Amrieh et al.[10]	Artificial Neural Network, Naïve Bayes, Decision Tree	Bagging, Boosting, Random Forest
Satyanarayana and Nuckowski[11]	J48, Naïve Bayes, DT Ensembles	Ensemble Filtering
Abubakar and Ahmad[12]	DT ensemble	Random Forest
Salini and Jeyapriya[13]	Logistic Regression, Support Vector Machine, DT Ensemble	Majority Voting
Ashraf et al.[14]	J48, Random Forest	Stacking
Mythili et al. [15]	DT Ensemble, Decision Table, Multi-Layer Perceptron, J48	Random Forest

According to Table 1, it is clear that constructing ensembles and comparing them with individual base classifiers in the area of predicting students' performance is an active research. Related studies using ensemble classifier techniques have shown that they are superior to many single base classifiers. [15] for example, compared the performance of Random forest, J48 and decision table algorithms and results obtained showed that Random forest algorithm is more accurate and the time taken to develop the model is less when compared to the other classifiers.

On the other hand, [10] used ensemble technique on an e-learning educational dataset that was collected from the Kalboard 360, a Learning Management System, to develop a model of performance of students with a new category of features called the behavioral features. Their model was used to evaluate the impact of student's behavioral features on their student academic performance using bagging, boosting and random forest with Naïve Bayes, Artificial Neural Network and Decision Trees as the base classifiers.

Multiple classifiers: J48, Naïve Bayes and Random Forest was also combined by [11] to improve the quality of student data by eliminating noisy instances and improving the predictive accuracy. They identified association rules that influence student outcomes using a combination of rule based techniques. They compare their technique with single model based techniques and concluded that using ensemble models not only gives better predictive accuracies on student performance, but also provides better rules for understanding the factors that influence student performance.

Random Forest was used by [12] to predict the academic performance of students based on their interaction with LMS, assessment and prerequisite knowledge. Result from their research show that Naïve Bayes outperforms Random Forest, however Random Forest, an ensemble of Decision Trees outperforms a single Decision Tree. In addition to the performance prediction, Random Forest was also used to identify the significant attributes that influence students' performance, which was validated by a statistical test using Pearson correlation. Their research revealed that lab task, assignments, midterm and prerequisite knowledge are significant indicators of students' performance predictions.

Majority Voting Ensemble on Random Forest, Logistic Regression and SVM classifiers was studied by [13] where they deduce that Majority Voting ensemble method results to a better performance accuracy.

[14] used ensemble stacking method with base classifiers: J48, RF and Random Tree on pedagogical (academic data) datasets to improve the prediction accuracy of student's performance using under sampling and over sampling filtering technique resulting to accuracies of 95.96% and 96.11% respectively.

It can be seen from the related work that none of the researchers extensively used both the homogeneous and heterogeneous ensemble methods on a Learning Management System datasets in an attempt to providing clues for constructing the best students' performance

predicting ensemble model. Therefore, there is no clear answer to the question about which ensemble classifier can provide the highest students' performance prediction accuracy. More specifically, very few studies examine the performances of different ensemble techniques on predicting students' performance.

IV. EXPERIMENTS

A) *Experimental Setup*

In the next few sections, we describe the data collection, the proposed model, the evaluation metrics we used, the size of partitioned data subsets and how we intend to determine the most significant attributes.

1) Data Collection and Description: The dataset used in this work was obtained from the Institute of Computing and Information Communication Technology (ICICT), Ahmadu Bello University (A.B.U) Zaria, Nigeria. The dataset contains 515 records of students' MOODLE activities which include: Course view, Resource view, Assignment submit, Assignment view and Forum view which are going to be compared with students' final grades to find out factors that indicate high performance. A total of six (6) attributes were selected including the class attribute. The tasks of data collection and data pre-processing includes: data cleaning, feature selection, data reduction and data transformation processes was done to improve the quality of the dataset. Table 2 gives the descriptions of the selected features and Table 3. illustrate a brief description of the data collected.

TABLE 2. ATTRIBUTES DESCRIPTION

S/N	Features	Description
1.	*Course_View*	Number of course views during semester
2.	*Assign_View*	Number of assignment views during semester
3.	*Assign_Submit*	Number of assignment uploads and updates during semester
4.	*Resource_View*	Number of resource views during semester
5.	*Forum_View*	Number of forum views during semester
6.	*Performance*	The students overall score/grade (pass or fail)

TABLE 3. DATA DESCRIPTION

	Course_ View	***Resource _View***	***Assignt _Submit***	***Assignt_ View***	***Forum _View***
count	511	505	513	513	500
mean	12.912	7.402	5.159	31.910	0.412
std	18.334	10.688	2.650	25.317	1.1246
min	0	0	0	0	0
max	106	68	14	181	14

2) Proposed Model: The preprocessed data was exported to WEKA where the individual classifiers: DT, NB and KNN was combined separately in the case of homogeneous and together in the case of heterogeneous to form ensemble models. Fig. 2. Illustrates the steps required to construct the ensemble models.

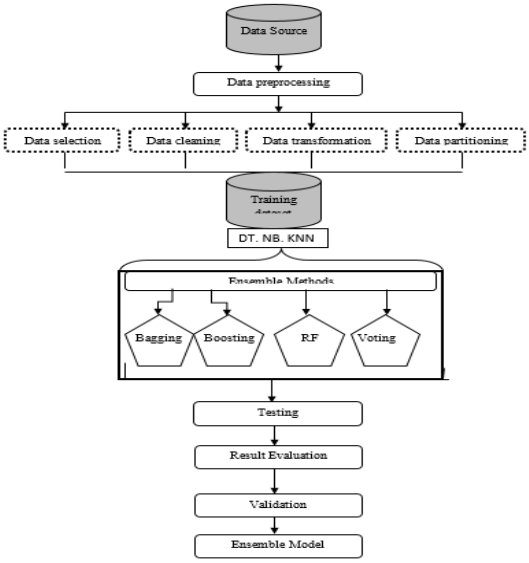


Fig. 2. Ensemble Model Construction

3) Evaluation Metrics: The training and testing was done using 10-folds cross validation. The performance of the models was evaluated using the confusion matrix metrics. The confusion matrix contains entries which include True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) which were used to compute the Accuracy, Precision, Recall and F-measure of each predicted and actual classes of the model as described by equations 8-11. Table 4 shows the confusion matrix.

TABLE 4: CONFUSION MATRIX

		PREDICTED CLASS	
		Pass	*Fail*
ACTUAL CLASS	*Pass*	*TP*	*FP*
	Fail	*FN*	*TN*

i. *Accuracy*: this is the proportion of total number of predictions that were correct. The higher prediction accuracy is, the better the model.

$$\text{Accuracy (A\%)} = (TP + TN)/ (TP + TN + FP + FN) * 100 \quad (8)$$

ii. *Precision*: this is percentage of instances that the classifiers marked and classified in the class and are actually in the class.

Precision (P %) = TP/(TP+FP) * 100 (9)

iii. *Recall or Sensitivity*: this is the proportion of actual positive cases which are identified correctly.

Recall (R %) = TP/(TP+FN) * 100 (10)

iv. *F-Measure* is the accuracy of harmonic mean of precision and recall that is the weighted average of the class.

F-Measure(F%)= (2*Precision*Recall)/ (Precision + Recall) *100 (11)

4) Size of Partitioned Data Subsets

Bagging method as a technique for re-sampling data into subsets was used to investigate the effect of maximizing the size of partitioned data subset on ensemble accuracy. Different data subset sizes were used by increasing and decreasing the number automatically in order to find its effect on the performance.

5) Determination of the most significant attributes

A function in WEKA called Correlation Attribute Evaluator which evaluates the worth of an attribute by measuring the Pearson's correlation between it and the target class was used to determine the most significant attribute in predicting students' performance. The search method used is the Rank method.

B) Experimental results

In the next few sections, we describe the results obtained from ensemble classifiers and non-ensemble classifiers and also the determination of the most significant attribute.

1) *Results on ensemble classifiers*

In this work the ensemble techniques used are Bagging, Boosting, Random Forest and Voting Ensemble methods. The evaluation measures are as shown in Table 5.

TABLE 5. ENSEMBLE EVALUATION MEASURE

Evaluation Measur	Bagging Method			Boosting Method			Voting Method	Random Forest (RF)
Classifiers Type	DT	NB	KNN	DT	NB	KNN	(DT, NB, KNN)	DT
Accuracy	85.3	84.9	77.3	85.0	84.9	76.7	83.7	90.3
Precision	84.7	84.5	84.7	84.6	84.6	85.1	84.5	90.3
Recall	96.6	99.1	89.2	94.5	98.9	87.8	98.9	100
F-Measure	90.2	91.2	86.9	89.3	91.2	86.4	91.1	94.9

Table 5 presents the result of the experiment using ensemble methods: Bagging, Boosting, Voting and Random Forest. It was observed that, good results were obtained when ensemble methods with single classifiers are used. Random Forest method using DT outperform other ensemble methods with an accuracy of 90.3% which means 51 students were identified accurately while 11 students were wrongly identified. Precision, Recall and F-Measure results are 90.3%, 100% and 94.9% respectively. Bagging and Boosting using DT are the second and third methods with accuracies of 85.3% and 85% respectively. Voting ensemble was the fourth method with an accuracy of 83.7% using DT, NB and KNN combine together. KNN using Bagging and Boosting methods are the least classifiers with accuracies of 77.3% and 76.7% respectively. Precision, Recall and F-Measure results for these classifiers are as shown in Table 3. The performance of each ensemble technique is as shown in Fig. 3.

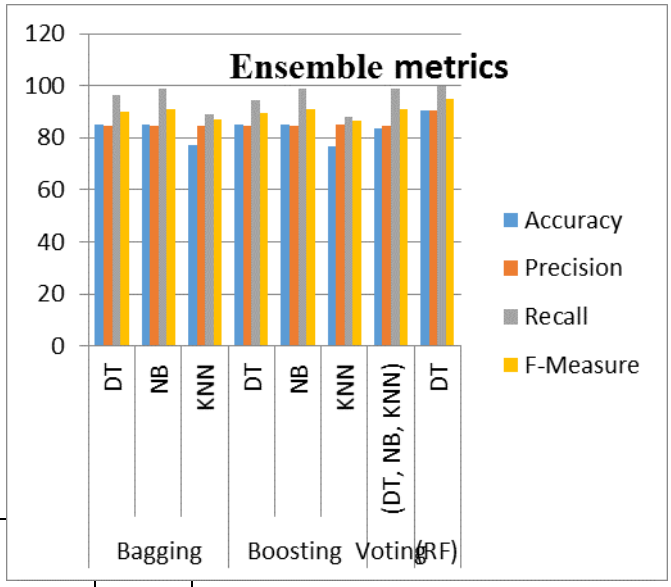


Fig. 3. Chart of ensemble model performance measure

The result shows that when running bagging method with different partitioned data subsets, as the partitioning is increasing or decreasing, there will be no significant changes in the result. This shows that maximizing the size of partitioned data subsets does not improve the ensemble accuracy.

2) *Comparison between ensemble classifiers and non-ensemble classifiers*

Table 4 shows the classification results using the following techniques: NB, KNN and DT. It indicates that DT model outperforms other data mining techniques with 84.1% accuracy, followed by the NB with 83.7% accuracy and then KNN with 76.7 % accuracy. The precision, Recall and F-Measure for the three classifiers are as shown in Table 6.

TABLE 6. EVALUATION MEASURES FOR BASE CLASSIFIERS

Evaluation Measure	Naïve Bayes (NB)	Decision Tree (DT)	Nearest Neighbor (KNN)
Accuracy	83.7	84.1	76.7
Precision	84.5	84.8	85.1
Recall	98.9	98.9	87.8
F-Measure	91.1	91.3	86.4

TABLE 7. COMPARISONS BETWEEN SINGLE CLASSIFIERS AND THEIR ENSEMBLES.

Classifiers	Evaluation Metrics			
	Accuracy	Precision	Recall	F-Measure
NB	83.7	84.5	98.9	91.1
NB ensembles (Boosting)	84.9	84.6	98.9	91.2
NB ensemble (Bagging)	84.9	84.5	99.1	91.2
DT	84.1	84.8	98.9	91.3
DT ensemble (Boosting)	85.0	84.6	94.5	89.3
DT ensemble (Bagging)	85.3	84.7	96.6	90.2
DT ensemble (Random Forest)	90.3	90.3	100	94.9
KNN	76.7	85.1	87.8	86.4
KNN ensemble (Boosting)	76.7	85.1	87.8	86.4
KNN ensemble (Bagging)	77.3	84.7	89.2	86.9

Table 7 shows the performance difference between the single classifiers and their ensembles in terms of accuracy, precision, recall and f-measure. This result indicates that ensemble classifiers are more superior to their individual base classifiers and the DT ensemble (Random Forest) is the best classifier.

3) *Determining the most significant Attributes*

In this section the relationship between students' activities on the MOODLE with students' final grade and the determination of significant important attributes were discussed. Table 8 shows the results.

TABLE 8. SIGNIFICANT ATTRIBUTES

Attributes	Correlation	Rank
Assign_Submit	0.57	1
Assign_View	0.44	2
Course_View	0.38	3
Resource_View	0.16	4
Forum_View	0.05	5

From Table 8, it can be observed that, the attributes *Assign_Submit* has a strong positive correlation with the score, with a correlation coefficient of 0.57. Hence, it's significant in determining students' final grade. *Assign_View* and *Course_view* have medium correlations with the score with correlation coefficient values of 0.45 and 0.38 respectively. Finally attributes *Resource_View* and *Forum_View* both have weak correlations with the score, hence, not significant factors for determining students' final grade in a particular course. Fig. 5. is the attributes correlation chart.

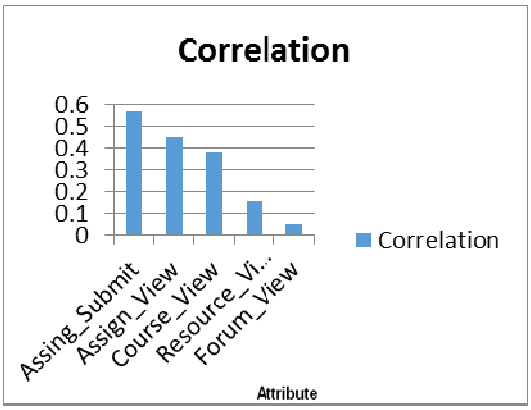


Fig. 5. Attributes Correlation Chart

Fig. 5. shows the most significant attributes in determining students' performance while interacting with MOODLE. The graph revealed that there is a positive correlation between assignment submission and course view with the final grade.

V. CONCLUSION

In this paper, we compared the performance of ensemble classifiers in predicting students' performance based on their interaction with a learning management system. More ensemble classifiers over the non-ensemble classifiers.

Our experimental results based real-world dataset collected from the Institute of Computing, Ahmadu Bello University, Zaria showed that Random Forest method using DT outperform other ensemble methods with accuracy of 90.3% and also the single classifiers. This work therefore, concludes that the DT using ensemble methods is the best classifier compare to other ensemble classifiers used in predicting students' final grade. The attributes that have the most significant impact on students' performance prediction was identified to include: *Assign_Submit, Assign_View*, and *Course_view*. Also, our result showed that maximizing the size of partitioned data subsets in bagging ensemble technique does not necessarily improve the accuracy of ensemble models.

For future work, some other approach may be considered. First, other ensemble methods such as stacking and Gradient Boosting Method (GBM) can be compared with the current study for more elaborate performance comparison. Secondly, more number of attributes can be used to conduct similar experiment in order to obtained better results. Thirdly, feature selection process should be used to select appropriate features by reducing dimensionality and redundant data, as a result of large amount of attributes/features used. Finally, a more elaborate use of different data sources may be required for an elaborate comparative study of both homogeneous and heterogeneous ensemble techniques in predicting students' performance base on their interaction with a learning management system.

REFERENCES

1. Usman, U.I., et al., *A Comparative Study of Base Classifiers in Predicting Students' Performance Based on Interaction With LMS Platform.* FUDMA JOURNAL OF SCIENCES-ISSN: 2616-1370, 2019. **3**(1): p. 231-239.
2. Raschka, S., *Naive bayes and text classification i-introduction and theory.* arXiv preprint arXiv:1410.5329, 2014.
3. Gao, H., X. Zeng, and C. Yao, *Application of improved distributed naive Bayesian algorithms in text classification.* The Journal of Supercomputing, 2019: p. 1-17.
4. Quinlan, J.R., *Induction over Large Data Bases*. 1979, Stanford Univ Calif Dept Of Computer Science.
5. Wang, W., et al., *Detecting Android malicious apps and categorizing benign apps with ensemble of classifiers.* Future Generation Computer Systems, 2018. **78**: p. 987-994.
6. Zhou, Z.-H., *Ensemble methods: foundations and algorithms*. 2012: Chapman and Hall/CRC.
7. Breiman, L., *Bagging predictors.* Machine learning, 1996. **24**(2): p. 123-140.
8. Dietterich, T.G. *Ensemble methods in machine learning*. in *International workshop on multiple classifier systems*. 2000. Springer.
9. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
10. Amrieh, E.A., T. Hamtini, and I. Aljarah, *Mining educational data to predict Student's academic performance using ensemble methods.* International Journal of Database Theory and Application, 2016. **9**(8): p. 119-136.
11. Satyanarayana, A. and M. Nuckowski, *Data mining using ensemble classifiers for improved*

specifically, several studies have demonstrated the superiority of ensemble classifiers over the single base classifiers, but most of them constructed these ensemble classifiers either solely on homogenous ensembles or on heterogeneous classifiers. Although also, existing comparative studies on ensemble classifiers techniques show that boosting method outperforms other ensemble techniques, they maintained that the performance of ensemble techniques is usually domain dependent. This claim is not clearly established in the domain of predicting students' performance base on their interaction on a learning management system.

To establish this claim, three widely used classification techniques including the decision tree (DT), naïve bayes (NB) and K-Nearest Neighbor (KNN) are used for constructing both the homogeneous and heterogeneous ensemble classifiers. NB ensembles, KNN ensembles and DT ensembles by bagging, boosting and random forest were constructed for performance comparison.

In addition, the ensembles of DT, NB and KNN was also constructed by voting and compared with the homogenous ensembles. Furthermore, an experiment using only the base classifiers was conducted to further confirm the superiority of

prediction of student academic performance. 2016.
12. Abubakar, Y. and N.B.H. Ahmad, *Prediction of Students' Performance in E-Learning Environment Using Random Forest.* International Journal of Innovative Computing, 2017. **7**(2).
13. Salini, A. and U. Jeyapriya, *A Majority Vote Based Ensemble Classifier for Predicting Students Academic Performance.* International Journal of Pure and Applied Mathematics, 2018. **118**(24).
14. Ashraf, M., M. Zaman, and M. Ahmed, *Using Ensemble StackingC Method and Base Classifiers to Ameliorate Prediction Accuracy of Pedagogical Data.* Procedia computer science, 2018. **132**: p. 1021-1040.
15. Mythili, M. and A.M. Shanavas, *An Analysis of students' performance using classification algorithms.* IOSR Journal of Computer Engineering (IOSR-JCE), 2014. **16**(1): p. 63-69.

A Conceptual Framework for Detection of Learning Style from Facial Expressions using Convolutional Neural Network

[1,2]Farouk Lawan Gambo*
[1]*Department of Computer Science*
Federal University Duste
Jigawa, Nigeria
farouk4142@gmail.com
farouk.gambo@fud.edu.ng

[2]Gregory Maksha Wajiga and [2]Etemi Joshua Garba
[2]*Department of Computer Science*
Moddibo Adama University of technology,
Yola , *Nigeria*
{gwajiga, e.j.garba}@mautech.edu.ng

***Abstract* - There are millions of learning materials over the internet that students can use to assimilate new information. But once their preferred learning style is known, they can be provided with a responsive recommendation so that can focus more on representations that will foster their understanding. Providing students with preferred learning object no doubt increase their motivation and hence their learning outcome. Identifying student's learning styles allows them to learn better and faster through several means. Traditionally, a test (use of questionnaire) is usually conducted for automatic detection and prediction of student's learning preferences particularly in e-learning. This approach though valid and reliable in detection of learning styles, but it is also associated with many challenges; learner self-report bias, individual learning styles may vary over time, Students not aware of the importance or the future uses of the questionnaire. To this end, this paper proposed a conceptual framework for detection of learning style from facial expression using Convolution neural network.**

Index Terms - Learning preference; Emotions; Predictive models; Convolution Neural Network (CNN).

I. INTRODUCTION

Educational technologies such as multimedia learning have become an integral part of our modern life. Easy access to information and comprehensive learning resources were provided by these educational technologies. Additionally, they allow instructors in learning environments to present information through different means, including animations, text, audio files, pictures, etc. These ample opportunities help learners with a responsive recommendation based on their wishes and preferences, which in turn allow them to focus on those representations that ease their understanding [1-4].

People differ in so many ways including what type of instruction is assumed to be most suitable for them known as learning style. Providing students with preferred learning style will no doubt motivate them and in return increase their learning outcome. Individuals have their own unique learning methods. e.g. some assimilate faster with a visual way of learning (e.g.diagrams, graphs pictures), while some prefer verbal way (reading or listening) etc. These methods of learning are what is known as preferred learning styles [8] and the assumption has been popular in the research areas of multimedia learning and e-learning in practice [5-7].

Identifying a student's learning style help both students and instructors; students assimilate more when their learning preferences are considered throughout their learning process while instructors provide students with possible intervention to foster the learning process. Traditionally, a test (use of questionnaire) is usually conducted for automatic detection and prediction of student's learning preferences particularly in e-learning. Even though this approach is valid and reliable in detection of learning styles, it is also associated with many challenges; [9,10]. Amongst which are: learner self-report bias and other drawbacks like individual learning styles may vary over time, Students choosing answers arbitrarily because they may be unaware of the importance or the future uses of the questionnaire and e.t.c.

The major motivation of this research is that humans emote while they are interacting with computers and are affected by their emotions [11]. Humans facial expressions (Emotions) are also indicators of how learners interact with, perceive, and respond to the learning object in learning environment. Thus, the prediction of learning style needs to be able to capture such a phenomenon in order to be free from the use of questionnaire which in turn would lead to an effective learning and academic achievement.

There has been a growing interest in improving the interaction between humans and computers particularly in e-learning. It was argued that "to achieve effective human-computer intelligent interaction, there is a need for the computer to interact naturally with the user, similar to the way humans interact" [12]. Humans interact not only through speech but also through certain body language to stress more on some part of speech. Emotions are displayed through physiological means; facial expressions, vocal and other means. Researcher in [13] emphasized that "There are more and more evidences appearing that shows emotional skills are part of what is called 'intelligence' " and one of

the most important ways for humans to display emotions is through facial expressions "[12].

This research proposes a framework capable of recognising student's affective states and infer learning styles from them through;

i. Developing of an algorithm for efficient recognition and classification of emotion based on facial expression.
ii. Identifying and mapping emotional classes onto specific learning style that positively correlates with different learning style
iii. Developing learning style predictive model from feature extracted in the learning style emotive database.
iv. Evaluation of the predictive model for recognition and prediction of the learning style using square mean error.

So far to the best of our knowledge, no system has proposed a scheme for the integration of such recognition ensembles in order to deduce the learning preferences of human subjects from facial expressions. Thus, this research focuses on the impact of facial expression to predict learning style of a student. Section II give related literature and section III explain our proposed framework for learning style detection. Finally the last section concludes and expected contribution.

II. Related Work

A. Learning Style

Literature shows that some approaches usually use algorithms from various fields; machine learning, data mining, artificial intelligence and/or computational intelligence to build a model from existing student behaviours to determine or classify their learning style. These approaches so far yield an average precision of 77% [10]. Identifying student learning style helped in knowing strength and weakness of a student in learning process. According to [14] there are two approaches to automatic detection of learning style namely: literature-based and data-driven. A data-driven approach aims at constructing a classifier that imitates a learning style instrument. The automatic detection of learning styles in data-driven approaches usually carried out by an AI classification algorithm which takes the user model as input, pass it on to predictive model and returns the students' learning style preferences as output [9]. While Literature-based approaches do not just construct a model from the student's data but use rules from literature to build a respective model [10]. Several researchers have used different machine learning algorithms such as Naïve Bayes Classifier and ANN to develop their model. Researcher in [15] proposed the use of a feed-forward ANN (a 3-layer perceptron) with back propagation under a supervised learning model to identify learning styles. Ten behavior patterns such as what kind of reading material did the student prefer, does the student revise their answers on exams prior to submission and e.t.c were used as inputs. While the neural network produces three values, representing the learning styles on three of the four learning style dimensions of the FSLSM as output. Bayesian network (BN) was used by [16] in order to detect students' learning styles. In their paper, they identified various behaviors that may be relevant to identifying learning styles in a given learning system. Then, a BN was trained with data from 50 students, using initial probabilities based on expert knowledge. [17] Proposed an automatic detection approach for learning styles capable of adapting to the learner's wishes to provide learning object that suit their learning style. [10] Used four computational intelligence algorithms (artificial neural network, genetic algorithm, ant colony system and particle swarm optimization) to investigate with respect to their potential to improve the precision of automatic learning style identification. Each algorithm was evaluated with data from 75 students. The artificial neural network shows to be the most promising results with an average precision of 80.7 per cent, followed by particle swarm optimization with an average precision of 79 per cent. Another automatic approach found in [18], mapped learning styles from the identified emotional tendencies of students, with believe that emotion influenced the student's learning style during the learning process. The research was conducted in Bahasa Indonesia which classifies learning style of the students based on sentiment analysis from student's tweet. The research concluded that there is a relationship between emotional tendencies on Twitter and the student's learning style.

In view of the same reasoning with [17] that emotional tendencies are related to learning styles based on student's tweet as well as the ability of a student to communicate emotionally and cognitively in HCI and education [19], a new approach can be used to automatically detect student learning preference based on their facial expression during the learning process. So that a proper intervention (i.e learning object) can be offered to maintain the student motivation.

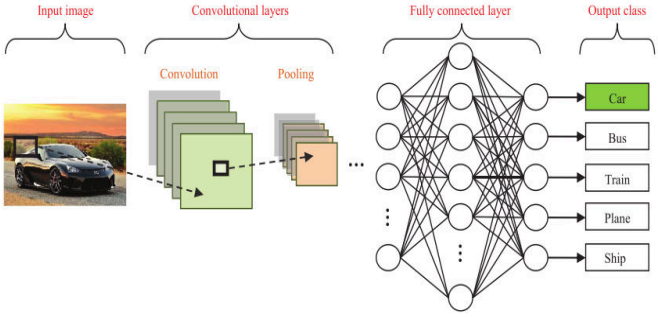
B. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) sometimes called ConvNets are described as deep, feed-forward Artificial Neural and which specialized mainly at analysing image data [20]. CNN models are mainly built from three (3) types of hidden layers; convolution layers, pooling layers and fully connected layers. This class of ANN have demonstrated exemplary performance on complex learning problems. CNN models are very powerful and have achieved impressive results on solving lot of complex problems [21]. Deep convolutional networks have been used by [22] for the recognition of emotions in human faces, the research augments five CNN input layer with a feature selection method, and tests result on video based CNU database of facial expression yielded a recognition rate of 90.5 per cent. Another research by [23] also applied deep neural networks for identification of facial expressions; they equally used a set of features based on Haar like features as inputs for a 7 layer convolutional neural network. Analysis on 327 images taken from the Cohn-Kanade dataset yielded an average recognition accuracy of 72 per cent.

While these researches have attempted to recognise human emotions, no system has proposed a scheme for the

integration of such recognition ensembles in order to deduce the learning preferences of human subjects

As the name implies, the main thing that differentiates CNN from ANN is its architecture that comprised of large amount of layers in the network. Due to the number of layers in the CNN, the original input is transformed many times than shallow networks. This way the network is able to 'learn' harder tasks than shallow networks, for example more complicated features can be extracted in image recognition [21]. There are four major types of deep learning architectures which include deep neural networks (DNN) [24], convolutional neural networks (CNN) [20], recurrent neural networks (RNN) [25], and emergent architectures [26]. The CNN architecture is motivated by the visual cortex of the human brain [27]. CNN consist of layers namely; convolutional, polling and fully connected layers. Figure 1 shows the architecture of a CNN for an image classification task.



CNN image classification pipeline.[30]

Convolutional Layers: This layer is responsible for feature extraction from input images. Input images are convolved with weights learned from the images in order to calculate a new feature map. The feature maps come from the arrangement of neurons in a network.

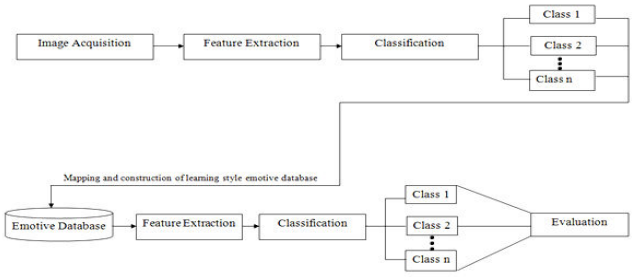
Polling Layers: This layer is responsible for reducing spatial resolution of feature maps [31]. Earlier studies use average pooling for aggregating all input values while recent studies [32] use max polling for aggregating maximum value for receptive fields.

Fully Connected layers: This is the high level in the network responsible for extracting more advanced abstract features. In [31], FC layer is composed of a sigmoidal neuron which sums the outputs of the last preceding convolutional layer while in some recent image classification tasks, softmax fuction at the last layer of the network which essentially converts the output of the last layer into a probability distribution (Krizhevsky *et al.* 2012).

This research use deep learning architecture (CNN) to classify emotional classes based on facial expression to infer learning styles and this is expected to increase the detection accuracy of our system.

III. SYSTEM OVERVIEW

This research aimed to design a framework for detection of learning style from facial expressions. The proposed model of our system composed of two parts; emotion recognition and learning style detection based on facial expressions, here is where we adapt the use of deep neural network architecture (CNN) which has the ability to efficiently classify emotion and learning style based on emotional classes.



An Overview of the Prediction Framework.

The predictive model that we propose in our system follows classification technique procedure which divide the steps into image acquisition, feature extraction and classification I, Emotive database, feature extraction and classification II, and evaluation

A. Image Acquisition

This is the stage where sample of images containing different emotion can be obtain from the subjects by instructing them to make facial expressions [28]. Alternatively, it can be done by showing subjects emotionally loaded pictures for them to mimic [29]. In order to allow facial expressions come naturally, key frames are to be extracted from the video sequences collected on students' behaviour during different learning situations.

B. Feature Extraction and Classification I

This phase uses the configured deep learning architecture CNN to automatically extract input features and classify accordingly. The deep neural network architecture (CNN) used here has the ability to classify different emotions. The architecture configured here is a variant of the Lecun network which harnesses two convolutional layers, two pooling layers and one output layer. The first layer takes an image reduced to 32X32 for computational efficiency and convolves around a singular span of 5X5 filter. This first convolution yields four sets of 28X28 feature maps. These feature maps are the subjected to pooling over a 2X2 window, resulting in six sets of 14X14 feature maps. These maps are then pooled a second time over a 2X2 fully mapped window, the result is then convolved around six 5X5 filters and then pooled to give the network's final output.

The network is trained using published affective datasets and sample image we sourced from image acquisition for bias adjustment. This step is responsible for data curation, which is responsible for dividing the dataset into two subsets. The CNN model uses the first subset of the dataset for actual training and the other for testing

C. Construction of Emotive Database

A new database of emotive learning styles would have to be constructed in order to accurately determine which emotional classes correlate with the different learning styles. The construction of the emotive dataset is done by extracting frames from video of student's learning scenarios while learning different learning object.

D. Feature Extraction and Classification II

After the emotive database is constructed, our proposed algorithm would then map features selected from emotional classes to the different learning styles in the corresponding degree of accuracy already determined while constructing the learning style database. In this part, we proposed the use of transfer learning where transference of neural weights from the layers of one network to another can be achieved in order to optimize the system. This will entail training the network for classification II by transferring the weights from the network of the classification I. Furthermore, video frames collected from different learning situations divided into two parts; one part of the data is extracted as a training set, while the other part used for testing.

E. Evaluation

The performance of the system would have to be evaluated by computing the mean square error performance on the system on the test data and then on a target data that has been suitably configured for cross validation purposes. This would require a partitioning of the dataset into a training data, a test data and a validation data. To ensure that these subsets are complimentary, multiple cross validations would be performed to keep variation between them minimum as possible. The mean of the predictive error recorded for these multiple validation trails would then be computed in order to arrive at an accurate assessment of the systems performance.

IV. Conclusion

In this study, we present a conceptual framework that will use student's facial expression, extract feature using configured CNN to recognise emotion and used them to classify student's learning style. If this novel approach is fully implemented; it is hoped to provide a better, more accurate and relevant studies in detection of learning style. The future work hopes to develop an application that will classify student learning style from user's facial expression and this will benefit research areas like e-learning, m-learning and affective education.

References

[1] Moser S. & Zumbach J. (2018). Exploring the development and impact of learning styles: An empirical investigation based on explicit and implicit measures, *Computers & Education*. doi: 10.1016/j.compedu.2018.05.003.

[2] Ainsworth, S. (2008). How should we evaluate multimedia learning environments? In Rouet, J.F., Lowe, R., & Schnotz, W. (Eds.), *Understanding Multimedia Comprehension* (pp. 249-265). New York: Springer.

[3] Sadler-Smith, E., & Smith, P. (2004). Strategies for accommodating individuals' styles and preferences in flexible learning programs. *British Journal of Educational Technology*, 35(4), 395-412.

[4] Richardson, J.T.E. (2010). Approaches to studying, concepts of learning and learning styles in higher education. *Learning and Individual Differences*, 21(3), 288–293

[5] Cassidy, S. (2004). Learning Styles: An overview of theories, models, and measures. *Educational Psychology*, 24(4), 419-444.

[6] Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and ped agogy in post-16 learning: A systematic and critical review*. London: Learning & Skills Research Centre

[7] Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, 9 (3), 105-119.

[8] John, A., Shahzadi, G., and Khan, K.,(2016) Students' Preferred Learning Styles &Academic Performance. *Sci.Int. (Lahore)*, 28(4), 337-341.

[9] Feldman,J.,Monteserin,A.,&Amandi,A.(2014).Detectingstudents'perc eptionstylebyusinggames.*Computers&Education,71*,14–22.

[10] Bernard, J., Chang, T.W., Popescu, E., Graf, S. (2017). Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms. *Expert Systems with Applications* 75, 94-108

[11] Reeves, B., & Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, Cambridge

[12] Azcarate, A., Hageloh, F., van de Sande, K., & Valenti, R. (2005).Automatic facial emotion recognition, *Universiteit van Amsterdam*

[13] Goleman,D.(1995) *Emotional Intelligence*. New York: Bantam Books.

[14] Graf, S. (2007). *Adaptivity in learning management systems focussing on learning styles*. Vienna University of Technology

[15] Villaverde, J.E., Godoy, D., & Amandi, A. (2006). Learning styles' recognition in e-learning environments with feed-forward neural networks. *Journal of Computer Assisted Learning, 22* (3), 197–206.

[16] García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks precision for detecting students' learning styles. *Computers & Education, 49* (3), 794–808.

[17] Hasibuan, M. S., Nugroho L.E., Santosa P.I., & Kusumawardani S.S.(2016), A Proposed Model for Detecting Learning Styles Based on Agent Learning. Retrieved August, 20th 2018 from https://doi.org/10.3991/ijet.v11i10.5781

[18] Habibi, R., Setyohadi, D.B., & Santoso, K.I. (2017). Student learning styles and emotional tendencies detection based on Twitter. In 4th International Conference on *Information Technology, Computer, and Electrical Engineering (ICITACEE)*, (pp. 239-243).

[19] Ghasemaghaei, R.(2017). Multimodal Software For Affective Education: User Interaction Design And Evaluation. (Unpublished Doctoral Dissertation Carleton University.Ottawa, Ontario)

[20] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097-1105

[21] Sorgedrager, R. (2018). Automated malaria diagnosis using convolutional neural networks in an on-field setting: The analysis of low quality smartphone based microscope images. (Unpublished Masters thesis. Delft University of Technology)

[22] Byeon, Y., & Kwak, K. (2014). Facial expression recognition using 3D convolutional neural network.International journal of Advanced computer Science and Application, 5(12).

[23] Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., (2015). Development of deep learning-based facial expression recognition system. *Frontiers of Computer Vision (FCV),* pp. 1-4.

[24] LeCun, Y. and Ranzato, M. (2013). Deep Learning Tutorial, In: *Tutorials in International Conference on Machine Learning (ICML'13)*, 2013. Citeseer

[25] Williams, R. J. and Zipser, D. (1989). A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Comput*,1(2), 270-80.

[26] Lena, P.D., Nagata, K., Baldi, P.F. (2012). Deep Spatio-temporal Architectures and Learning for Protein Structure Prediction. *Advances in Neural Information Processing Systems, 2012. p. 512–20.*

[27] Hubel, D. H., & Wiesel, T. N. (1959). Receptive Fields of Single Neurones in the Cat's Striate Cortex. *Journal of Physiology*, 148(1), 574–591.

[28] Levenson, R. W. , Ekman, P., and Friesen. W. V. (1990). Voluntary Facial Action Generates Emotion Specific Autonomic Nervous System Activity. *Psychophysiology*, 363-384.

[29] Collet, C., Vernet-Maury, E., Delhomme, G., and Dittmar, A. (1997). Autonomic Nervous System Response Patterns Specificity to Basic Emotions. *Journal of the Autonomic Nervous System*, 62(1-2) 45-57.

[30] Rawat, W and Wang, Z,(2017). Deep Convolution Nueural Networks for Image Classification: A Comprehensive Review. *Neural Computation,* 29(9), 2352-2449.

[31] LeCun, Y. (1989). Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman, & L. Steels (Eds.), *Connections in perspective* (pp. 143–155). Zurich, Switzerland: Elsevier.

[32] Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv 1505.00853v2.*

A New Concept of an Intelligent Protection System Based on a Discrete Wavelet Transform and Neural Network Method for Smart Grids

*Ali Hadi Abdulwahid
Department of Electrical Engineering Techniques
Engineering Technical College
Southern Technical University
* E-mail: dr.hajjiali@stu.edu.iq , *dr.alhajji_ali@yahoo.com

***Abstract*—In most countries, renewable energy is utilised for the generation of power in consideration of the increasing pollution and technology development. The main concern in the protecting energy systems in utility firms is spotting and identifying the basic cause of electrical faults in order to prevent power outages. This article recommends Back-Propagation Neural Network (BPNN) and Discrete Wavelet Transform (DWT), which are new techniques for the classification and detection of faults in the microgrids' transmission lines. The neutral network utilizes MATLAB in the completion of training and simulation techniques. These signals contain high frequencies that are disintegrated through Daubechies4 mother wavelet 'Db4'. In neural network training, Wavelet Energy Coefficients (WECs) and Wavelet Transform Coefficients (WTCs) are applied as back propagation inputs for the detection of patterns and classification of faults. The information obtained for detection and classification of fault is fed to the neural network. This article suggests Wavelet Transform (WT) for the detection of fault and technique to recognition of disturbance. Faulty voltage signals are processed and gathered through the WT for the detection of faults. The simulation confirms the accuracy and reliability of the new algorithm.**

Keywords- Smart Grid; artificial intelligence; Renewable Energy; fault detection; wavelet transform

I. INTRODUCTION

In order to attain highest performance level, one should ensure that throughout the process the energy is reliable and of the best quality. Therefore, it is extremely important to identify fault in transmission lines via distinct techniques and correct them fast so that they can maintain its reliability and accuracy. Due to errors and other noises, a power system's current and voltage may contain brief components, random noise, and high harmonics. Hence, accurate and quick detection of transmission line and fault interference classification plays an essential in power systems.

The power plant's electrical energy is transmitted to the consumption place, in the traditional. Normally, power stations are situated far from industrial and urban areas as it needs a special environmental condition. Electrical energy is produced under medium voltage, and over long distance it is delivered in high voltage. After which, they use a multi-stage step-down transformer to step-down the voltage [1].

Smart grids are the future of power generation companies. This grid applies digital technology for the provision and distribution of electricity to consumer according their requirements; it reduces the energy consumption, and applies numerous renewable energy sources for the generation of power [2].

The objectives of a smart grid include:

- Diversification of the group network through inclution of people as a central part not just as consumers but as electricity providers as well;
- Increased utility in renewable energy;
- Lowering the power plants' burden through lessening the dependence on electricity generation;
- Lessening in complete power cut;
- Developing the network capacity in supply of electricity;
- Lessening the time needed for restoration of electricity during a fault;
- Reduction of the curve of peak load; ensuring maximum present exploitation.

In the recent years, different techniques of classifying fault shave have been suggested. Such methods include: combinations of both technologies [7, 8], transformed wavelets [5, 6], and artificial neural networks [3, 4].

The main components of voltage and current's determined data parameters should be examined fast. In evaluation, each time there is a fault, they have to quickly review the power content of fault signal. In the extraction of the energy content for Level 1 and Level 5 they apply the signals energy content and review of simulation Wavelet transform (WT). The energy content points transformation to Level 5 is shown, for fault data which assists in much accurate evaluation. Wavelet Db is seen as the most suitable technique as the fault signal details are review much accurately through it.

In the protection of the energy system, specifically its peripheral devices, and enhancing its performance in any specific operating situations; defective or normal, they require an effective protection system for the installation of modern energy. Modern systems prefer digital relays rather than the electromechanical relays as digital relays are faster,

reliable and accurate. Besides, for execution in real-time it is very crucial to detect faults via error detectors and digital relays. This article suggests a fault-detection algorithm through wavelet transform and usage of independent components analysis.

Installations of modern energy need an effective protection system to guard the system itself and the associated devices and to enhance the performance in defective and normal defective operating conditions. Currently, in modern systems, they replace the electromechanical relays with the digital relays because of their properties like, reliability fast operation, and accuracy. In real-time execution, error detection via error detectors and digital relays is vital. This article, proposes a fault-detection algorithm by wavelet transform and uses it for the analysis of independent factors. Abrupt changes should be detected through detection techniques in order to adapt suitable solutions for protecting the power system from any disturbances. The proposed method outputs were reviewed in different operating scenarios, like the harmonics, frequency variations, and the presence of disturbance. In detecting of faults, WT is appropriate because of its effectiveness in time-frequency resolutions and its reliability for attaining best feature for identification functions [3–5].

This article proposes a new technique for the classification of transmission line faults into an intelligent network through a Back-Propagation Neural Network (BPNN) and Discrete Wavelet Transform (DWT). The main concept of the technique is the creation of a simple multi-layer perceptual network (MLP) applying the wave energy coefficients and the wave coefficients details of the current as input patterns. The article outlines the new decision-making algorithms development for protection relays of the fault classification and detection. PowerSim is applied for the simulation the fault condition, in order to verify the method.

The fault detection analysis technique is described in Section 2, after which the algorithm is executed in Section 3. Section 4 then explains the result analysis, then finally Section 5 provides the conclusions.

II. General Description of Wavelet Transforms and Artificial Neural Networks (ANNs)

A. Wavelet Transform

Conversion of wavelet has gained increased attention in fault analysis because to its better capability in analysing mobile waves in comparison to other techniques. Then, transient wavelets are transformed into wavelets series which are similar to a time zone signal casing a certain octave frequency range which has a more comprehensive information. For comprehensive calculation of the coefficients it applies a 1D multi-level wavelet decomposition scheme. The syntax is provided below.

WT represents an effective device for time-frequency domain signal analysis and a waveform which contains an amount of zero and a finite period of time. Progressive WT is described as:

$$c(a,b) = \int_R s(\mathrm{t}) \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) dt \tag{1}$$

Whereby (t) represents the parent wavelet while s represents the main signal.

$$a = 2^j, b = k2^j, (j,k) \in \mathrm{z}^2 \tag{2}$$

Whereby the wavelet level is represented by *j*, while time, discretely is represented by *k*. *C*(*a,b*) represents a coefficient which is the magnitude of similarity between the main signal and the scaling mother wavelet. At all the level *a* of processing, the simulation of coefficients of wavelet *Aj* and *Dj* is as follows [6–8]:

$$D_j(t) = \sum_{K \in z} c(j,k) \Psi_{j,k}(t) \tag{3}$$

$$\Psi_{j,k}(t) = 2^{-j/2}(2^{-j} - k), j \in z, k \in z \tag{4}$$

$$A_j = \sum_{j>J} D_j \tag{5}$$

A_J represents the main signal's approximation, and D_j provides the signal's details. As Figure 1 shows, the signals passing via the two filters complement each other and are known as high-pass decomposition g[n] and low-pass decomposition h[n] whilst the signal convolution is explored using the filter's coefficients [9].

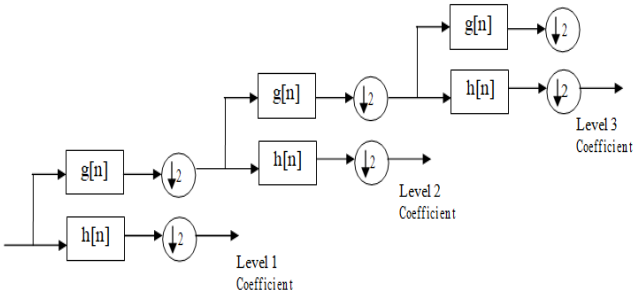


Fig 1. The Discrete Wavelet Transform decomposition process

The high-frequency (HF) wavelets and low-frequency (LF) coefficients are processed and then measured through down-sampling of the outcome of the analysis of the wavelet

multilevel mode, echoing the forementioned single procedure, and the low-frequency coefficients disintegration at all stages, in order for the signal to decompose through a number of phases in the low part-compostable. Below is the formula for determining the discrete *x(n)* of the signal's energy [10]:

$$E = \sum_{i=1}^{n} X^2[i] \tag{6}$$

whereby *n* represents the number of the signal samples.

B. Back-Propagation Neural Network (BPNN)

Naturally, the neural network is partitioned into two sections: training and testing. Training refers to the process of learning in the neural network system, that regulates the mapping of the outputs and input values in order to attain the suitable model, while the testing refers to the evaluation process for the accuracy of the training process model. A BPNN represents a well-trained network which ensures a balance between its ability to give an accurate response to the input patterns and its ability to identify the patterns applied for training, that are identical to the patterns applied during the training. Back-propagation training comprises of the three steps below [11-13]:

- FEED FORWARD. In forward propagation, insert the value (*xi*). Then each cell output (*zj*) of the unseen layer will spread to the unseen layer, through the activation function description, and hence, producing the network output value (*yk*). Following, is the comparison of the network output value (*yk*) against the target value(*tk*). The difference $tk - yk$ is the error which has taken place. When this error is below the tolerance limit, the iteration is prevented. Nevertheless, when the error exceeds the tolerance limit, each line's weight in the network shall be changed to lower the error.
- BACK-PROPAGATION. According to the error $tk - yk$, the formulation factor δk (k = 1, 2, ..., m) is applied in order to assign the error in unit *yk* to each unseen units, that will be linked directly to the *yk*.In addition, δk is applied to modify the line weight, add directly associated to the output unit. Similarly, the unseen layer is determined per cell within the δk component for the underlying layer the primary weight of all the modification, until the determination of the input unit for every factor δ.
- WEIGHT CHANGE. Following the determination, the factor δ, with respect to the weight change for the upper neurons depending on the line's weight factor δ, all the lines' weight shall be modified as well. For specific fault classification and detection models, one applies the Mean Absolute Error (MAE). The suitable model should have the least MAE values.

$$MAE = \frac{\sum_{i=1}^{n} |e_k|}{n} \tag{7}$$

whereby MAE represents the Mean Absolute Error, *ek* is $tk - ty$, *n* represents the number of data, *tk* represents the target, and *yk* represents the output.

III. Proposed Algorithm

This part outlines the fault detection techniques for wind systems linked to the network, as:

Microgrid Simulation whereby we formed distinct faults types. In the defective PCC state, we extracted the voltage signal. Then, the signal was passed via a WT to achieve a frequency analysis time for the identification of the fault condition.

To find the decorrelation and mean we processed the voltage signal. This represents the initial processing level. Then data was minimised and filtered to acquire an appropriate data redundancy. The key input data components were obtained. The independent factor was determined with respect to an iterative basis till we attained a fixed point.

From the matrix, we determined and generated the independent components, s, and correlation matrix solution W, of the symbol signal. The fault signal showed the use of performance index, and when the index was more than the threshold value, a fault was detected.

Figure 2 shows the suggested single-wire transmission line fault classification and detection of the algorithm design process.

In order to obtain information of the transient signal in the frequency and time domains, DWT analyses the input signal. In the recommended algorithm, a 'Db4' mother wavelet is applied to extract DWT coefficients for varying fault classification types.

The BPNN training input comprises of information on the wavelet energy and wavelet coefficients. We should bear in mind the composition of varying fault situations and the importance of generating a training pattern via simulation of different fault types for a single circuit transmission line. Thus, we can identify the fault location, fault types, initial failure, and fault resistance.

For a specific application one should select the suitable BPNN structure and topology. In the recommended scheme, varying architectures have been put into consideration. In this research, the input set applied was 12 output current signal samples for one circuit transmission line. Considering two hidden layers, the number of neurons differs with the

hidden layers(1)=10, the hidden (2)=20 findings and the output set 4, as Figure 2 shows.

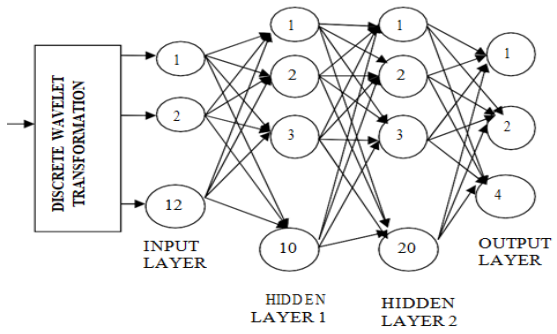


Fig 2. The proposed DWT-BPNN architecture

In this research, we applied MATLAB/Simulink for the detection of the internal faults. The subsequent power supply system has a three-phase, 25-kV, 50-Hz power supply and 10-km transmission line as well as a 47-MVA transformer linked to the "Y-Y" and an inductive load as Figure 3 shows.

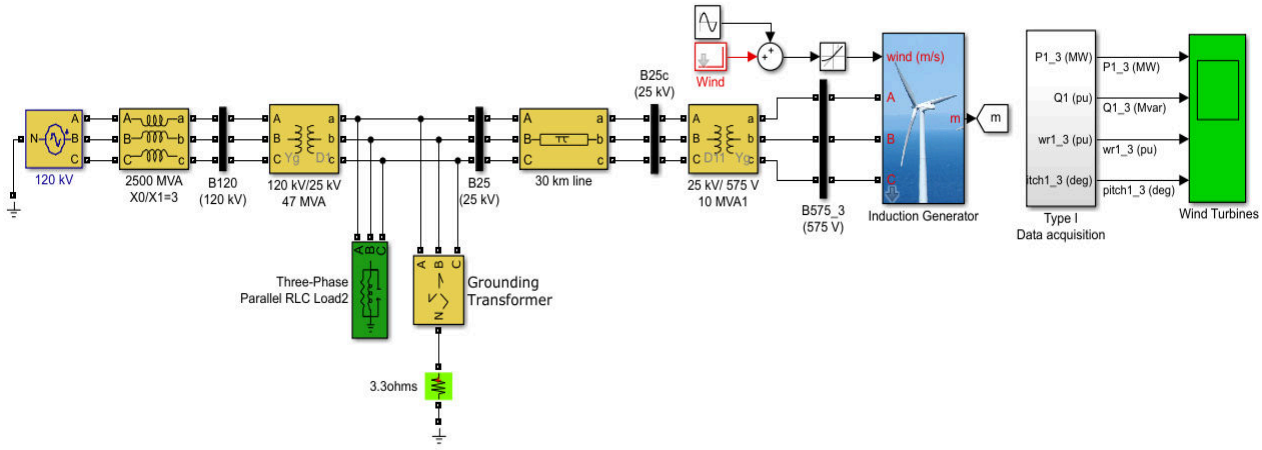


Fig 3. Single line diagram for a smart grid connected to the utility grid

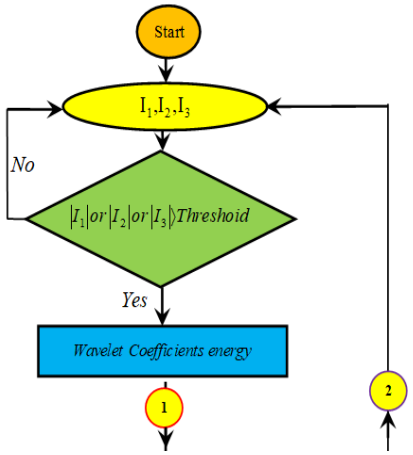
This relay's varying performance diagram is provided with respect to the flowchart. The WT process starts with the input current, which exceeds the threshold value, as indicated in Figure 4.

Figure 4 represents the manner in which a person can use the correlation factor *CFm* to identify the fault type, whereby the element m = 1, while 2 is the phase number. *Fnm* represents a filter, whereby n = 1, while 2 represents the length per filter and *Min* represents the least value of every filter *Fnm*. Computation of each phase is as shown below:

$$A_{n1} = \min(CF_{n1}) \quad (7)$$

$$A_{n2} = \min(CF_{n2}) \quad (8)$$

$$A_{n3} = \min(CF_{n3}) \quad (9)$$



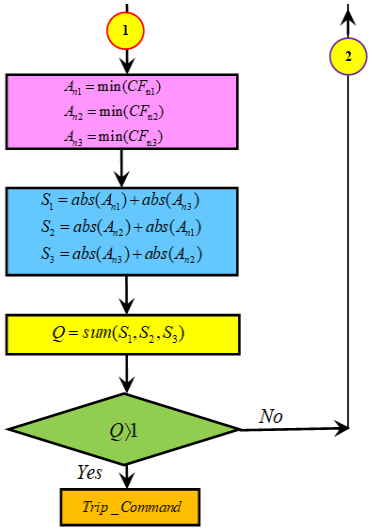


Fig 4. Relay option flowchart

The sum signal energy is described as:

$$S_1 = abs(A_{n1}) + abs(A_{n3}) \quad (10)$$

$$S_2 = abs(A_{n2}) + abs(A_{n1}) \quad (11)$$

$$S_3 = abs(A_{n3}) + abs(A_{n2}) \quad (12)$$

The Q parameter is described as:

$$Q = sum(S_1, S_2, S_3) \quad (13)$$

When the value of Q exceeds 1, then internal fault is detected.

IV. Simulation Results

This part outlines the findings of detection and the respective discussion of the system of grid-connected wind farms. The simulation of the electrical system was conducted in MATLAB/Simulink, with a 2 kHz sampling frequency. For LG faults case, at the common coupling point the fault was eliminated from the network through the WT signal. WT was used to process the signal, in accordance to the instant of the fault utilising a high-pass and low-pass series of filters to detect faults. Figures 5–7 represents the findings for voltage signals with detection. These outcomes confirmed clearly that WT correctly detected the occurrence of prompt fault.

In an external fault situation, the differential relay was not applicable, and in the external fault condition the relay output was zero. In the case of internal fault, at t = 0.25 s, the relay accurately opened the breaker in less than 0.05 s. The fault close to the coupling point matches with the voltage signal in the frequency changes and operating conditions. The yield index simulation acquired showed that the failure started at arround 0.05 s and its detection was through the instantaneous rise in index values. Prior to the fault, it was identified that the value was nearly zero. Once the failure takes place, the value rises immediately, and on setting the threshold value, it then can be detected. However, in such a situation, the threshold value was not set instead, we conducted an evaluation of the defective situation with respect to immediate growth. Figure 1. Represent the action taken to identify the fault through the application of a correlation component, after which the opening command was provided to the breaker, Figure 8 shows the relay's performance.

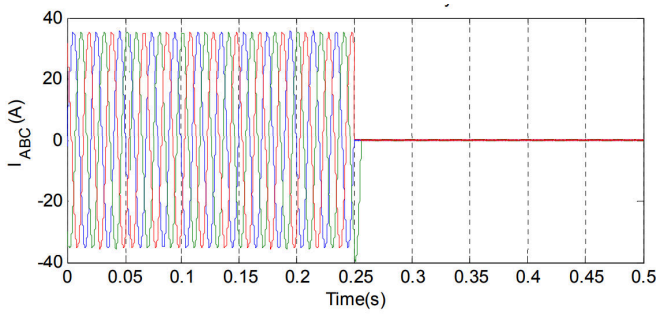


Fig 5. Internal fault current

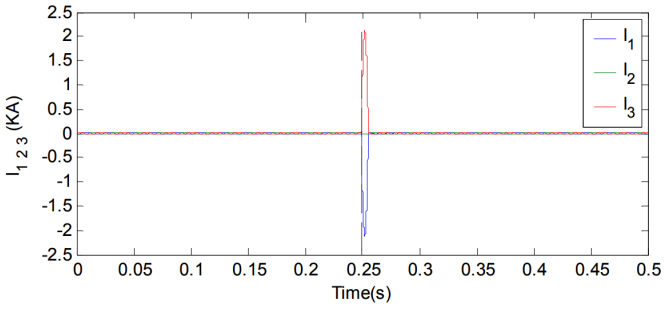


Fig 6. Differential fault current

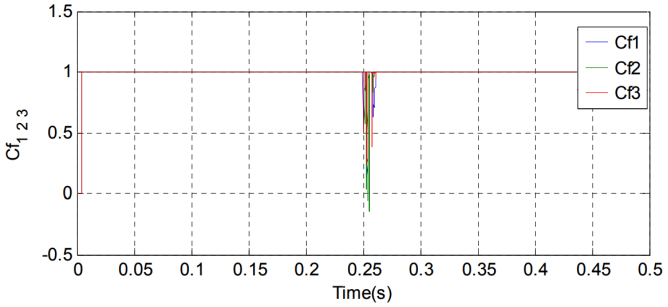


Fig. 7. Correlation factor

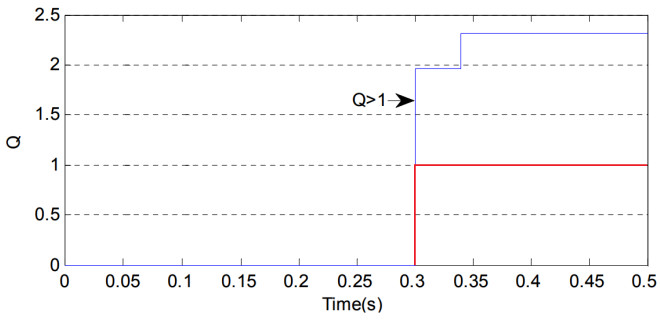


Fig. 8. Relay output and Q status. Delay time = 0.05 s; Fault time = 0.25 s

V. CONCLUSIONS

Currently, a better part of the world has resorted to the utilization of renewable energies for the electricity generation because of the growth in pollution problems, technological advancements and such renewable energies should be linked to the electricity network in what is now referred to as a smart grid. The smart grids represent the future of the electricity transmission and distribution in this new century. They depend greatly on utilising renewable energy resources, optimizing the application of the produced electricity, lowering the energy production cost, offering alternatives in purchasing of electricity, and involving consumers as part of the process of electricity generation. This research shows a new technique of evading malfunctioning of differential protection relays through transient fault currents. Moreover, it can be applied as an appropriate relay for smart grid applications in relays which are believed to operate properly in MATLAB/Simulink within the test systems. The designed relays display good performance speed for the protection of the bus-bars, hence, they are essential. The maximum fault recognition time is < 10 ms and applies for distinct fault types.

This article outlines the application of fault detection through WT and also independent components analysis. The voltage signals considered in the PCC were processed in varying operation scenarios and faults under the above methods. It was confirmed that the WT had the ability to detect a fault suddenly with high accuracy. Nevertheless, in some conditions, we could not differentiate between the cases because of the frequency shifts. Thus, through the application of all operating programs, we can detect the fault accurately.

ACKNOWLEDGEMENTS

I would like to pass my sincere gratitude to Ms Janan Abd Ali Al-Hajji, who supervised my research.

REFERENCES

[1] Wavelet-based transformer differential protection with differential current transformer saturation and cross-country fault detection," IEEE Trans. Power Del., vol. 33, no. 2, pp. 789–799, 2018.

[2] A. H. Abdulwahid and S. Wang, "A busbar differential protection based on fuzzy reasoning system and Rogowski-coil current sensor for microgrid," 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conf. (APPEEC), 2016, pp. 194–199.

[3] A. H. Abdulwahid and S. Wang, "Application of differential protection technique of domestic solar photovoltaic based microgrid," Int. J. Control Autom., vol. 9, no. 371–386, 2016.

[4] K. De Kerf, K. Srivastava, M. Reza, D. Bekaert, S. Cole, D. Van Hertem, and R. Belmans, "Wavelet-based protection strategy for DC faults in multi-terminal VSC HVDC systems", IET Generation Transmission Distribution, vol. 5, no. 4, pp. 496–503, 2011.

[5] M. M. Eissa, "New differential busbar characteristic based on high frequencies extracted from faulted signal during current transformer saturation", IET Gen. Transm. Distrib., vol. 8, no. 4, pp. 619–628, Apr. 2014.

[6] S. AsghariGovar and H. Seyedi, "Adaptive CWT-based transmission line differential protection scheme considering cross-country faults and CT saturation", IET Gen. Transm. Distrib., vol. 10, no. 9, pp. 2035–2041, 2016.

[7] R. P. Medeiros, F. B. Costa, and K. M. Silva, "Power transformer differential protection using the boundary discrete wavelet transform", IEEE Trans. Power Del., vol. 31, no. 5, pp. 2083–2095, 2016.

[8] F. B. Costa, "Boundary wavelet coefficients for real-time detection of transients induced by faults and power-quality disturbances", IEEE Trans. Power Del., vol. 29, no. 6, pp. 2674–2687, 2014.

[9] F. B. Costa, "Fault-induced transient detection based on real-time analysis of the wavelet coefficient energy", IEEE Trans. Power Del., vol. 29, no. 1, pp. 140–153, 2014.

[10] A. Abdulwahid and S. Wang, "A novel approach for microgrid protection based upon combined ANFIS and Hilbert space-based power setting," Energies, vol. 9, p. 1042, 2016, doi:10.3390/en9121042.

[11] R. P. Medeiros, F. B. Costa, and K. M. Silva, "Power transformer differential protection using the boundary discrete wavelet transform," IEEE Trans. Power Del., vol. 31, no. 5, pp. 2083–2095, 2016.

[12] S. Asghari Govar and H. Seyedi, "Adaptive CWT-based transmission line differential protection scheme considering cross-country faults and CT saturation," IET Gen. Transm. Distrib., vol. 10, no. 9, pp. 2035–2041, 2016.

[13] H. Monsef and S. Lotfifard, "Internal fault current identification based on wavelet transform in power transformers," Electr. Power Syst. Res., vol. 77, no. 12, pp. 1637–1645, 2007.

Automatic Generation of Design Class Model from Use Case Specifications

Muhammad Shaheed Abdullahi
Dept. of Computer Science
Ahmadu Bello University
Zaria, Nigeria
shaheedabdullahi@hotmail.com

Talal Shaikh
School of Mathematical and Computer Sciences
Heriot-Watt University
United Kingdom
t.a.g.shaikh@hw.ac.uk

John Kanyaru
School of Mathematical and Computer Sciences
University of Wolverhampton
United Kingdom
jmkanyaru@gmail.com

Muhammad Aminu Umar
Dept. of Computer Science
Ahmadu Bello University
Zaria, Nigeria
umaralameen@yahoo.com

***Abstract*—Object oriented design mainly involves construction of structural and behavioral models of a system informed by the specification. Mature work exists, example, Model Driven Development (MDD) and Domain Specific Language (DSL) to demonstrate the efficacy of deriving application code from designs. However, automated derivation of designs from specifications remains a challenge. Two key motivations for a seamless derivation of designs from specifications is traceability between phases and reduced development effort. Works in this area have mainly revolved around construction of formal specifications that are amenable to execution by a machine. In this paper, we present an intermediate weaving model which provides a means for incorporating design thinking and semi-automatic construction of a design model from the specification. Our work also involves the use of Natural Language Processing (NPL) to perform textual analysis of the specification, the output of which forms the basis of the weaving model.**

Keywords—Analysis class model, design class model, Natural Language Processing, Requirements specification, Unified Modeling Language, Use Case Description.

I. INTRODUCTION

Software development is a complex task that is carried out in different phases. Hence, there is high possibility of missing information while transiting from one phase to another [1]. The phases are Requirements analysis, design, and implementation [2]. While developing software, models of the system are created and each of these models are expressed, or placed in certain information space; in which an output of each model may be an input to the next model or set of models. Hence, there is need to fill any gap between these transitions to avoid information loss.

Analyzing software requirements, constructing analysis and design models from the requirements could be complex which often requires tool support. Errors that occurred during this process when discovered in later phases of development could be expensive to fix [3]. In the analysis phase, systems analyst gather requirements from client, which are documented in natural language format. The document is used for communication between different stakeholders of the system being developed [4]. The use of Natural Language (NL) can be said to be useful at this stage. However, requirements specification in NL are often ambiguous, lack consistency, and can be incomplete [5]. In addition, other factors such as social, psychological, or geographical can influence understanding and interpretation of the requirements specified in Natural Language [5].

Therefore, in this paper, an intermediate weaving model is presented that provides a means for incorporating design thinking and semi-automatic construction of a design model from the specification.

II. BACKGROUND

In our previous paper [11] we described a method for visual prototyping of specifications as a means to validation, a state-based method for enacting use case specifications was described. We argued that, enactable specifications, due to their visual nature provide cognition cues to stakeholders that enable the teasing out tacit issues about the problem domain. We have since started considering the efficacy of deriving useful design models from the validated specification. In our initial work, we considered the Rational Unified Process (RUP) proposition that it is feasible to derive structural models (especially, class models) from specifications. The proposition argues that textual analysis in which nouns and verbs are obtained from the specification can lead to classes (nouns) and methods (verbs) from the specification. However, this direct derivation results in a trivial class model. This paper proposed an approach to support the construction of a useful design model from the specification. Our work involves the use of natural language processing coupled with an intermediate weaving model between the specification and design phase.

III. PROBLEM DESCRIPTION

Requirement specifications are often used to identify objects, attributes, behaviors, and relationship among objects. These specifications if not adequately captured may always lead to missing information. Furthermore, proper communication among stakeholders is a key to success of any software development project [4]. In the specifications phase, software requirements are specified and documented in NL, and these requirements serve as reference point for the rest of object-oriented software engineering activities. The next phase after specifications is the design phase, in which initial system models are developed from the textual specifications document. As easy as it may sound, this

transition is however a very complex yet the most important task, as error made during this transition is costly and will affect the entire development process. Additionally, this transition is error prone due to the inconsistent and ambiguous nature of NL, and due to the varying skills of software analysts and designers, their understanding of domain knowledge as well as mapping process; different analysts and designers tend to derive different design models from same requirements document [1]. This problem may even get worse when the design models needs to go along with the changing requirements. Hence, a tool support for automating this transition process is highly desirable to fill this gap.

IV. RELATED WORK

Several attempts have been made to automate the process of generating different design models from requirements written in NL, for instance Agile Modeling (AM) is a tool developed to support business architecture, domain object modeling, and Object-Oriented Design (OOD) [6]. It provides tools for automatic Class-Responsibility Collaboration (CRC) cards approach, which follows stakeholder participation approach. A drawback with this, there is need for breakdown of the specifications to a more atomic statement. This approach proposes that statements should be made appropriate for specific requirements such as business rule, use cases, or user story which makes it domain dependent.

A tool is proposed in [5] that support developers in semi-automatic generation of UML models from normalized natural language requirements. The approach focuses on generating use case diagram, followed by conceptual model and collaboration diagram for each use case. The approach then generates collaboration diagram for each use case from which consolidated design class model was generated. This approach is quite similar to the proposed approach.

A tool named Natural Language Object-Oriented Production System (NL-OOPS) is proposed in [7]. NL-OOPS is aimed at generating Object-Oriented Analysis (OOA) model from NL requirement documents. Their approach is based on core Natural Language Processing System (NPLS), Large-scale Object-based Linguistic Interactor Translator Analyzer (LOLITA). LOLITA has a knowledge base named Semantic Net (SemNet), which stores knowledge that can be accessed, extended, or modified. NL-OOPS follows Noun-Phrase approach in identifying classes, attributes and associate classes using links, which is used to construct OOA model. Although the approach is quite essential, some major drawbacks are identified.

CM-builder [1] is another tool that support OO analysis stage of software development. It uses NLP techniques in analyzing requirements document written in English language. The approach involves several stages of text processing by the NLP module that includes; lexical preprocessing, morphological analysis, parsing, as well as discourse analysis, which produce the final discourse model. CM-Builder generates the analysis model in CASE Data Interchange Format (CDIF), which can be visualized in Unified Modelling Language (UML) tool.

TRAM [8] is a tool proposed to automate construction of analysis model from natural language requirements. The approach uses conceptual patterns known as Semantic Object Models (SOMs) to capture meanings of commonly used concepts and relationships in software requirements. TRAM uses NL parser to process the requirements document and build intermediary SOMs as means of filling a gap between natural language requirements and analysis class model, which are mapped into corresponding class diagrams. The approach uses action type to uniquely identify each SOM by categorizing types into nine categories, which are; Change, possession, motion, creation, selling, giving, owning, buying, and issuing.

A tool named RM2PT is proposed in [9] that is aimed at automatic generation of prototype from requirements model. Their approach supports requirements modeling and analysis, automated prototype generation, requirements validation and evolution.

An approach is proposed in [10] to address inconsistency problem that is found in between process models and natural language requirements. The proposed method consists of process model-based procedure for capturing execution-related data in the requirements models, then uses algorithms that take these models as input for generating natural language requirements.

Another approach in [11] uses ontology for constructing business process model in semi-formal language (BPMN). Their study deals with auto-generated requirements specification document from semi-formal modeling. Requirements ontology are used as input, and mapped using a rule-based method to BPMN elements, from which XML format of the BPMN are automatically generated.

The reviewed literatures above have made attempts to automate the process of generating design models from requirements. However, a number of limitations were observed which make room for improvements. For example, the domain object modeling and object-Oriented Design (OOD) [6] may require further breakdown of the specifications to more atomic statement for easy translation. While, a tool proposed in [5] depend massively on human intervention for refinement of UML elements. Similarly, the generated CDIF out by CM-builder [1] requires further refinement by another CASE tool, as there is high possibility of omission and errors in the model. NL-OOPS [7] on the other hand, has essential approach, however, it lacks accuracy contrasting between objects and attributes, the approach also has issues addressing ambiguity and inconsistencies that can be found in requirements document.

The method proposed in this paper provides syntactic construction rules for writing use case specifications, which are analyzed using strong NLP tool, which is a tool developed by Stanford NLP Group [12]. This will be used for the extraction of part-of-speech in each sentence within the use case scenario. While, a weaving model is derived from the scenario following some proposed predefined rules that provides relevant elements for constructing design class model for each use case. A weaving model is an intermediary model to avoid direct derivation of class model from requirements specification. This will address the ambiguity and inconsistency of natural language. The

proposed method will also make clear separation between object classes, methods of a class and attributes, which will address intensive user intervention in identifying these elements.

V. OVERVIEW OF THE APPROACH

The approach is based on predefined rules for constructing use case specifications, which will be used as an input to the prototype system. The use case specifications are analyzed using strong NLP tools, namely Stanford parser, and following certain rules in extracting part-of-speech for each sentence in the flow of events. For identification of design elements from the use case model, an interaction model, sequence diagram is generated for each use case as a weaving model that is used to capture the dynamic behavior of the system, from which design class model can be generated.

Use Case Description Syntactic Rule

Considering the ambiguous nature of requirements specifications written in natural language format, it can be challenging to extract relevant information for construction of class model. Some of the complexities are mapping class elements to the appropriate class, identifying which class relates to the other and the kind of relationship. These challenges can also be due to limitations of NLP tools, their inability of handling complex sentences and performing syntactic reconstruction of the requirements written in NL format. To overcome these complexities, the following syntactic construction rules are proposed for writing the use case description, which will be used as an input to the system for textual analysis using NLP techniques.

1. For all flow of events, a simple sentence should be used in the format,

   Subject: Predicate
   OR

   Subject: Predicate: Object
   For example,
   The Worker resumes.
   The Customer receives parcel.

2. The sentence should be in active voice form rather than passive voice. This will enable the parser identify *verb* between *Subject* and *Object* as message sent.
   For example,
   The Worker gave parcel to Customer.
   Where *gave parcel* can be identified as message passed between Worker (Subject) and Customer (Object).
3. Every object in the flow of events should be written by its name rather than pronoun.
4. Same verb phrase should be used for same action in different sentences in the flow of events. For example,
   System displays information of processing customer.
   AND
   System shows allocation information.

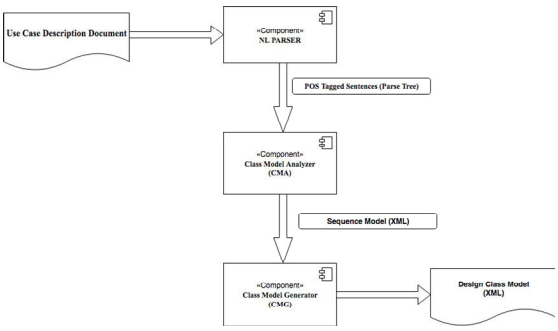


Fig. 1 Overall System Archicteture

The 2 sentences above are different, but perform same action although with different verb phrase. Since the 2 verbs means same thing, one should be used in the two sentences throughout the flow of events.

Overall Description

The process starts with taking use case description document as an input, which goes through several processing stages, by several components to produce design class model as depicted in the architecture in Fig. 1 and described in subsections (i to iii).

i. NLP Parsing
In this processing stage, the use case description document written with guidelines proposed in section A are parsed using Stanford parser. The parser assigned part-of-speech (POS) tags to each word and token in the use case description document, which is a key to identification of different class model elements. The POS tags being assigned includes singular noun denoted as *NN,* plural noun denoted as *NNS,* verb denoted as *VB,* and adjective denoted as *JJ.*

ii. POS Processing and Actor Extraction
for identification of candidate classes, methods, and variables, a textual analysis is performed on the tagged sentences produced by previous stage. A noun-phrase approach is followed using the following set of rules that initially generates static aspect of class model:

R1: For each sentence in the flow of events;

If tagged word is found in a glossary of system-specific keywords such as System, Computer, and Windows, the word is discarded.

R2: For each sentence in in the flow of events;

If frequency of occurrence for a tagged word is less than 2, the word is disregarded.

R3: For each sentence in the flow of events;

If tagged word is found in a glossary of attributes-like words, such as name, address, *information, then the word is regarded as attribute of a class.*

R4: For each use case;

Both primary and secondary actors extracted from actors' field are candidate classes.

R5: For each sentence in the flow of events;

If a word is tagged as a singular noun (NN) or noun plural (NNS), and does not satisfy R1 to R3, it is likely a candidate class.

R6: For each sentence in both basic and alternative flow of event,

If a word is tagged as a verb, in its base form (VB) or a verb, in past tense (VBD), it is likely a candidate method of a class.

R7: For each sentence in the flow of events

If a word is tagged as an adjective (JJ), it is likely a candidate attribute of a class.

The rules proposed above produces analysis class model, which represents static structure of a system, from which a designer can further refine to reduce irrelevant classes and class elements. However, this will only produce list of classes, methods and attributes. It does not indicate how classes relate to each other, which methods and attributes belong to which class. The reason is that, use cases only describe partial behavior of a system but does not reveal internal structure of the system. Rather, it is used to capture and document external requirements [13]. Hence, direct derivation from use cases to class model is insufficient for constructing reasonable software. To deviate from this problem, a sequence diagram is constructed as a weaving model for identification of design class elements from the use cases. This is used to devise interactions between objects of classes, described in section (iii).

iii. Dynamic Behavior with Sequence Diagram as Weaving Model

In this stage of the process, each use case description is parsed for extraction of actors, objects that receives the action and action performed from which sequence diagram can be generated as a weaving model. The reason for generating weaving model is to fill the gap between analysis model, which mainly represents static structure of a system and design model, which represents dynamic behavior of objects for all classes in the class model. Another reason is to identify relationships between classes; these are achieved by identifying message passed between interacting objects.

Sequence diagram is one of the UML diagrams that are used to model interactive behavior of a system. Another diagram is collaboration diagram, which basically has similar objective with sequence diagram of capturing dynamic behavior of a system, but from different perspective. In addition to capturing dynamic behavior of a system, interaction diagrams are used to describe message flow in a system; it is used to capture structural organization of system objects, and to capture interaction among these objects. The major difference between sequence and collaboration models is that, collaboration diagram emphasizes on structural organization between objects while sequence diagram emphasizes on sequence of messages and time ordering. The aim of generating sequence diagram, as a weaving model is to capture events between interacting objects by identifying initiator of the action as subject or actor, the receiver, and the message passed between them as action. By converting use cases into sequence diagram, objects in the analysis model are identified together with their internal operations, which represent internal implementation of behavioral model from which design class model can be generated.

In general, each use case should have a sequence diagram that can be generated with the following proposed syntactic rules for identification of subject, object and action or message passed.

Rule A: for each sentence in the flow of events of every use case;

Rule A-1: If syntactic structure is in the form

Subject + Predicate + Object

If Subject is a Noun Phrase "NP", Object is Noun Phrase "NP" and Predicate is Verb Phrase "VP" in the form,

Noun0 + Verb + Noun1

Then, Noun0 is the Sender Object; Noun1 is the Receiver Object, and Verb is considered as Message passed between Noun0 and Noun1.

Rule A-2: If syntactic structure is in the form

Subject: Predicate (without any object)

If Subject is a Noun Phrase (NP), and Predicate is a Verb Phrase (VP), denoted as

Noun0 + Verb

Then, Noun0 is considered as Receiver Object, and Verb is considered as message passed.

Rule B: For each sentence in the flow of events;

If syntactic structure is in the form

Noun0 + Verb + Noun1 + "of" +Noun2

Where Verb is transitive verb,

Then, there is interaction between Noun0 and Noun2, in which Noun0 is the Sender Object; Noun2 is the Receiver Object, and Verb+Noun1 “concatenated” is considered as Message passed between them.

Also, there is interaction between Noun0 and Noun1, where Noun0 is the Sender and Noun1 is the Receiver Object with Verb as the message passed.

There is “Has A” relationship between Noun1 and Noun2.

Rule C: For each sentence in the flow of events;

If syntactic structure is in the form

Noun0+Verb+Noun0+ “to/for/from” + Noun2

Where Verb is transitive verb,

Then, there is interaction between Noun0 and Noun2, in which Noun0 is the Sender Object; Noun2 is the Receiver Object, and the message passed between Sender and Receiver is Verb+Noun1 “concatenated”.

Also, there is interaction between Noun0 and Noun1, with Noun0 as Sender and Noun1 as Receiver, where Verb is the message passed between sender and receiver.

The proposed rules will be applied to all use cases, and each use case will be realized by constructing its sequence diagram (generated as xml) from which class model is generated. NL parser generates parse tree for each sentence in the use case scenario, which is used for identification of objects and the message passed, in additional to the actors extracted in previous stage.

VI. CONCLUSION

This paper provides a further understanding of Object-Oriented Software Engineering (OOSE). The aim of the paper was to propose a way to automate the process of transition between analysis and design phases of Object-Oriented Software Engineering, by automating design class model generation from use case model. The paper proposed syntactic construction rules for writing use case descriptions, which can be processed using several NLP techniques. The approach applies some proposed rules on pre-processed use case description to generate sequence diagram as a weaver between use case model and the design class model. The weaving model will be use in identifying interacting objects and how the objects relate to each other, as well as messages passed between interacting objects. This helps in mapping methods to its appropriate object’s class and relationships among object classes. The future direction of this work is the implementation of the proposed architecture. The implementation will address practically the issues relating to efficiency such as the redundancy in generating class elements, Incompleteness, as well as ambiguity of natural language in developing OOSE.

References

[1] H. H. a. R. Gaizauskas, "CM-Builder: an automated NL-based CASE tool," in Proceedings ASE 2000. Fifteenth IEEE International Conference on Automated Software Engineering, 2000.

[2] R. Pressman, Software engineering, Boston: McGraw-Hill, 2001.

[3] M. Elbendak, P. Vickers and N. Rossiter, "Parsed use case descriptions as a basis for object-oriented class model generation," Journal of Systems and Software, vol. 84, no. 7, pp. 1209-1223, 2011.

[4] I. Jacobson, Object-oriented software engineering, New York: ACM Press, 1992.

[5] D. K. a. R. S. Deeptimahanti, "Semi-automatic generation of UML models from natural language requirements," in Proceedings of the 4th India Software Engineering Conference, New York, 2011.

[6] Agilemodeling.com, "Class Responsibility Collaborator (CRC) Models: An Agile Introduction," 2015. [Online]. Available: http://agilemodeling.com/artifacts/crcModel.htm. [Accessed 16 March 2015].

[7] L. MICH, "NL-OOPS: from natural language to object oriented requirements using the natural language processing system LOLITA," Natural Language Engineering, vol. 2, no. 2, pp. 161-187, 1996.

[8] K. J. Letsholo, Z. Liping and E.-V. Chioasca, "TRAM: A tool for transforming textual requirements into analysis models," in 2013 28th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2013.

[9] Y. Yang, X. Li, Z. Liu and W. Ke, "RM2PT: a tool for automated prototype generation from requirements model", in ICSE '19 Proceedings of the 41st International Conference on Software Engineering: Companion Proceedings, Montreal, Quebec, 2019, pp. 59-62.

[10] B. Aysolmaz, H. Leopold, H. Reijers and O. Demirörs, "A semi-automated approach for generating natural language requirements documents based on business process models", Information and Software Technology, vol. 93, pp. 14-29, 2018. Available: 10.1016/j.infsof.2017.08.009 [Accessed 26 August 2019].

[11] A. Yanuarifiani, F. Chua and G. Chan, "Automating Business Process Model Generation from Ontology-based Requirements", in ICSCA '19 Proceedings of the 2019 8th International Conference on Software and Computer Applications, 2019, pp. 205-209.

[12] 3 August 2015. [Online]. Available: http://nlp.stanford.edu/software/index.shtml.

[13] S. S. Stephane, "Supporting use case based requirements engineering," Information and Software Technology, vol. 48, no. 1, pp. 43-58, 2006.

[14] J. Kanyaru and K. Phalp, "Validating software requirements with enactable use case descriptions," Requirements Engineering, vol. 14, no. 1, pp. 1-14, 2008.

Control of a Double Pendulum Crane System Using PSO-Tuned LQR

Ibrahim Bako Abdulhamid
Department of Electrical Engineering
Bayero University, Kano
700241 Kano, Nigeria
iabdulhamidbako@yahoo.com

Mustapha Muhammad
Department of Mechatronic Engineering,
Bayero University, Kano
700241 Kano, Nigeria
mmuhammad.mct@buk.edu.ng

Amina Ibrahim Khaleel
Department of Mechatronic Engineering
Bayero University, Kano
700241 Kano, Nigeria
aikhaleel.mct@buk.edu.ng

***Abstract*—A Gantry crane system is commonly used for point to point transportation as well as lowering and lifting of payload. The use of cables for hosting of payload can lead to natural swaying which makes it difficult to perform alignment, fine positioning as well as detrimental to safe and efficient operations. This paper presents a particle swarm optimization (PSO) based linear quadratic regulator (LQR) controller; for position and sway control of a double pendulum crane system. The major problem of designing an LQR controller is that, the Q and R parameter are obtained by trial and error which is quiet laborious. This paper optimized Q and R parameters by using PSO algorithm to obtain the best or optimal results. Simulation studies were carried out on the double pendulum nonlinear crane model which is created in MATLAB/Simulink environment. The simulation results show the effectiveness of the proposed controller in terms of reducing the trolley position percentage overshoot, hook oscillation and payload oscillation by 65%, 89.8% and 88.6% respectively.**

Keywords—Double pendulum crane, position tracking, swing suppression, LQR control, PSO

I. INTRODUCTION

Over the years, the use of crane as a viable equipment for the transportation of heavy loads and hazardous material in industries, factories, construction sites and harbor among others is well elucidated and appreciated. The choice of any particular type of crane such as; overhead crane, gantry crane, boom crane and tower cranes depend solely on its application. Nevertheless, it has a common problem[1]. One of its major challenges in industrial applications is the payload oscillation during point-to-point movement of such crane[2].

The general control objective desirable of an overhead crane system is the fastness and accuracy of trolley's positioning, together with effective payload swing suppression[3].

Study shows that most control approach treats the payload oscillation as a single pendulum without consideration of the hook mass and additional cable as worthy variables[4]. However, in practice, double pendulum is used. Even at that, efficient and safe operation of the crane system is paramount, thus a system with fast and accurate positioning with minimum hook and payload sway is most desirable for increased industrial productivity, as well as enhanced safety of operation. Dynamic of a double pendulum crane system is so complicated that its control design is very tasking and interesting. The double pendulum crane (DPC) system is an under actuated non-linear system, this implies that a single control input (trolley force) is required to achieve three control variables task of; (trolley position, hook and payload oscillation angles)[5].

Research into DPC system is dated back to 1998[4]. Thereafter, it has increasingly become an area of interest for researchers and numerous control techniques for DPC system are explored. Interestingly, these efforts are becoming an attractive benchmark and they involve delayed feedback control[4], decoupling control[6], wave-based control[7], passivity–based control[8], sliding mode control[9] and vibration control[10]. In [2] acceleration profile was generated for DPC system maneuvers involving hosting using an iterative learning control. However, the measurement of the payload oscillation was not discussed. Other work includes energy based control[11], super twisting based control[12], online motion planning[13]. In [14] and [15] time-optimal trajectory planning strategy with swing suppression was proposed. In[16], LMI based control technique was used on the DPC system which did not only solve the problems of having to use more than one controller but have a better performance response. However, it involves the use of complex mathematics. More so, in [17] adaptive control method was suggested for the DPC system which successfully suppressed the maximum hook and payload oscillation with less maximum control force and lower transportation time. Nevertheless, all these control methods involve rigorous mathematical analysis. GA tuned PID for DPC systems is used in[18]. Reference

[5] and [19] proposed a PSO tuned PID control method for the DPC system, this method successfully tracked the position but took longer time for the sway angles to settle and also many controllers are used. Reference [20] used new time domain performance criterion PSO tuned PID-PD for the DPC system, the controllers are effective in moving the overhead crane as fast as possible without any steady state error with low load oscillation however, many controllers are used.

In this paper, a PSO based parameter optimization (Q and R) of an LQR controller was designed and implemented via simulation for position and swing suppression of the DPC system. To be precise, the parameters of the controller; Q and R are tuned using PSO.

The rest of the paper is organized such that: section 2 describes the dynamic model of the DPC system, while section 3 describes the PSO algorithms for tuning the LQR parameters, and section 4 details the LQR control scheme. Furthermore, Section 5 presents the simulation results while the conclusion is enumerated at section 6.

II. Double-Pendulum Crane Dynamic Model

DPC consists of a three independent generalized coordinates: the trolleys position, the hook angle and the payload angle as shown in Fig. 1:

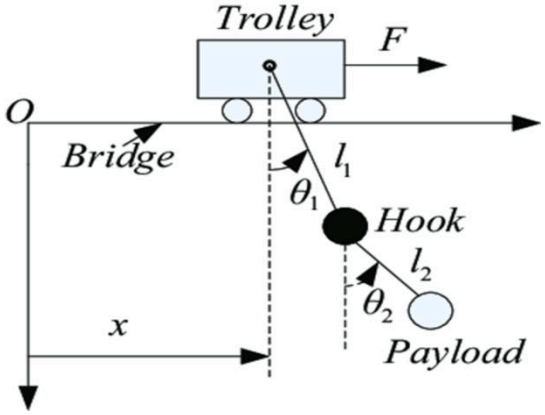


Fig. 1: Double Pendulum Crane System[7]

The parameters and variables of the DPC are shown in table 1[7].

Table 1: Double Pendulum Crane parameters

Parameter	Value	Unit
Trolley mass (m)	6.5	kg
Mass of hook (m_1)	2.0	kg
Mass of payload (m_2)	0.6	kg
Cable length from trolley to hook (l_1)	0.53	m
Cable length from hook to payload (l_2)	0.4	m
Gravitational constant (g)	9.8	m/s^2
Trolley position (x)		m
Hook angle (θ_1)		rad
Payload angle(θ_2)		rad

The dynamics of the DPC is represented using (1) – (3) below [6]:

$$(m + m_1 + m_2)\ddot{x} + (m_1 + m_2)l_1\ddot{\theta}_1 \cos\theta_1 + m_2 l_2 \ddot{\theta}_2 \cos\theta_2 - (m_1 + m_2)l_1\dot{\theta}_1^2 \sin\theta_1 - m_2 l_2 \dot{\theta}_2^2 \sin\theta_2 = F \quad (1)$$

$$(m_1 + m_2)l_1\ddot{x}\cos\theta_1 + (m_1 + m_2)l_1^2\ddot{\theta}_1 + m_2 l_1 l_2 \ddot{\theta}_2 \cos(\theta_1 - \theta_2) + m_2 l_1 l_2 \dot{\theta}_2^2 \sin(\theta_1 - \theta_2) + (m_1 + m_2)\, g l_1 \sin\theta_1 = 0 \quad (2)$$

$$m_2 l_2 \ddot{x}\cos\theta_2 + m_2 l_1 l_2 \ddot{\theta}_1 \cos(\theta_1 - \theta_2) + m_2 l_2^2 \ddot{\theta}_2 - m_2 l_1 l_2 \sin(\theta_1 - \theta_2) + m_2 g l_2 \sin\theta_2 \quad (3)$$

The equations of motion are linearized around the equilibrium point by the use of Taylor series expansion, and similarly the following assumptions are worthy of note: $\theta_1 = \theta_2 = 0$. Hence $\sin\theta_1 = \theta_1$ and $\sin\theta_2 = \theta_2$ and $\cos\theta_1 = \cos\theta_2 = 1$ and also $\cos(\theta_1 - \theta_2) = 1$ then;

$$\ddot{x} = -\frac{l_1(m_1+m_2)}{(m+m_1+m_2)}\ddot{\theta}_1 - \frac{m_2 l_2}{(m+m_1+m_2)}\ddot{\theta}_2 + \frac{F}{(m+m_1+m_2)} \quad (4)$$

$$\ddot{\theta}_1 = -\frac{1}{l_1}\ddot{x} - \frac{m_2 l_2}{l_1(m_1+m_2)}\ddot{\theta}_2 - \frac{g}{l_1}\theta_1 \quad (5)$$

$$\ddot{\theta}_2 = -\frac{1}{l_2}\ddot{x} - \frac{l_1}{l_2}\ddot{\theta}_1 - \frac{g}{l_2}\theta_2 \quad (6)$$

Applying the parameters in table 1, the state space description of the system can be obtained as

$$\dot{x} = Ax + BU \quad (7)$$

$$y = Cx \quad (8)$$

Where

$$x = [x\ \dot{x}\ \theta_1\ \dot{\theta}_1\ \theta_2\ \dot{\theta}_2]$$

$$U = F$$

$$y = [x\quad \theta_1 \theta_2]^T$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.9178 & 0 & 0.0013 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -31.4308 & 0 & 5.5461 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 31.8513 & 0 & -31.8531 & 0 \end{bmatrix}$$

$$, \quad B = \begin{bmatrix} 0 \\ 0.1539 \\ 0 \\ -0.2904 \\ 0 \\ 0 \end{bmatrix},$$

$$C = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0].$$

III. Particle Swarms Optimization (PSO)

The idea of PSO was first introduced by Resell Eberhart and James Kennedy in 1995 [21], and it was infused by social behavior of birds flocking and fish schooling[5]. Each individual within the swarm is delineated by a particle in search space. The particle has one assigned vector, which determines the next movement of the particle called the velocity vector[22]. The PSO algorithm is initiated with position being randomly initialized, such that individual particle moves through a multidimensional search space[23]. Each of the particles placed in the problem space has a fitness value evaluated by an objective function to be optimized at its current location. For every iteration, the particles in a local neighborhood share memories of their best visited position[24]. Then it uses these memories to adjust their own velocities and position according to (9) and (10).

$$V_i(t+1) = \omega V_i(t) + C_1\varphi_1\left(P_{best} - x_i(t)\right) + C_2\varphi_2\left(g_{best} - x_i(t)\right)$$ □□□□ □□□

$$x_i(t+1) = x_i(t) + V_i(t+1)$$ □□□□ □□□□

For each particle, its previous best position is used to obtain the best fitness position. The position is called personal best position; “Pbest” and the fitness value for this position is then stored[25]. The PSO algorithm also tracks the best position achieved by the individual particle in the swarm. This value is called global best position; “gbest”.

In this paper, the PSO technique is used to determine the most optimum weighting values for Q and R of the controller to minimize the sway angle of the hook and payload as well as to track the position of the cart. Q and R are the entries that form the particles in the PSO algorithm and are chosen to be diagonal matrices with positive real elements. The fitness value used in this paper is calculated based on the integral of square error (ISE) between the desired and actual value of the state to be controlled as presented in (11).

$$ISE = \sum(y_{desired} - y_{actual})^2 \quad (11)$$

The particle with the minimum ISE value of all the particles is then chosen as the global best “gbest”.

IV. Controller Design

This section describes in details the design of the control strategies for the state feedback LQR controller for the DPC system

A. State feedback LQR controller

This type of controller uses linear model of a system in state space form to estimate the controller gain.

The purpose of the controller is to minimize the cost function:

$$J = \int_0^{Tf}(x^TQx + u^TRu)dt \quad (12)$$

Where: Q and R are positive definite or positive semi definite, real and symmetric matrices[25].

The gain $K = R^{-1}B^TP$ of the control law $u = -kx(t)$ is obtained by solving the algebraic Riccati equation shown in (13):

$$A^TP + PA + PBR^{-1}B^TP + Q = 0 \quad (13)$$

The selection of the optimal values of Q and R was done by PSO and were found to be

$Q = diag(62.6696\ 3.2370\ 18.5614\ 36.795\ 76.3446\ 35.6301)$, $R = 0.03$ and the controller gain $K = [k_0 - k_i]$ is computed

using LQR MATLAB command $K = lqr(A, B, Q, R)$ and the gain were:

$$k_o = [52.4306 \quad 57.1565 \quad -318.6918 \quad -29.9244 \quad 119.5803 \quad -21.3701]$$

$$K_i = 20.2012$$

B. Proposed LQR control scheme

To reduce steady stated error integral action is added to the system as shown in Fig. 2. The constant gain Ki was used as pre-compensator.

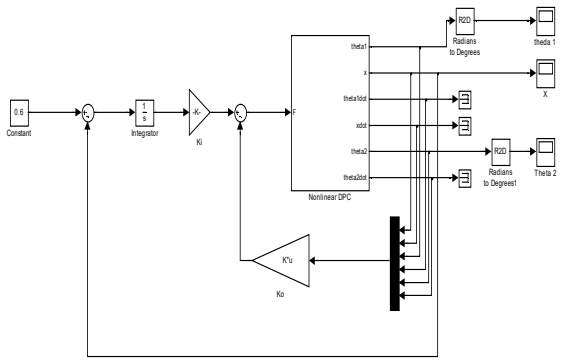


Fig. 2: LQR control scheme for overhead crane system.

V. Simulation Results and discussion

The performance of the proposed controller was investigated via simulations in Simulink. The simulation results were compared with those from the work in [5] which uses a combination of three Proportional Integral Derivative (PID) controllers. The parameters of the PID controllers are given in Table 2. The robustness of the proposed controller was investigated with respect to reference input tracking and payload variation.

Table 2: PID ControllerParameters[5]

PID controllers	Parameters	Values
PID1 (for Trolley position)	*Kp*	*19.7443*
	Ki	*0.0046*
	Kd	*15.9720*
PID2 (for hook oscillation)	*Kp*	*0.9709*
	Ki	*29.5439*
	Kd	*7.2471*
PID2 (for payload oscillation)	*Kp*	*0.6627*
	Ki	*1.5400*
	Kd	*0.1484*

A. Tracking performance

The simulation was carried out with the desired trolley position of 0.6m and 1.0m and a swing angle of $0^0$for both the hook and payload.

Case I: For reference step input of 0.6m, the position, hook, payload response and control signal are shown in Fig. 3 through Fig. 6.

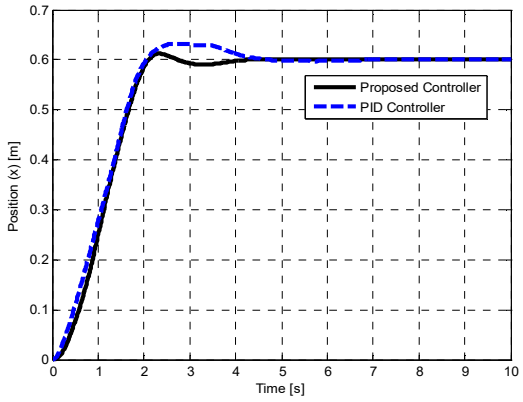


Fig. 3: Trolley position

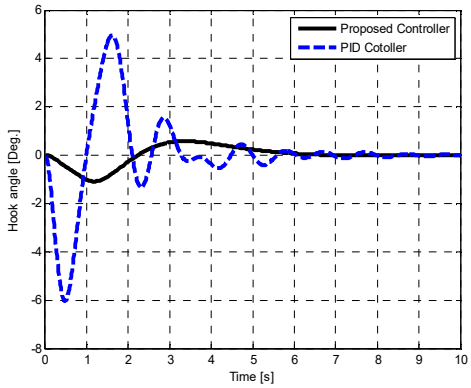


Fig. 4: hook swing angle

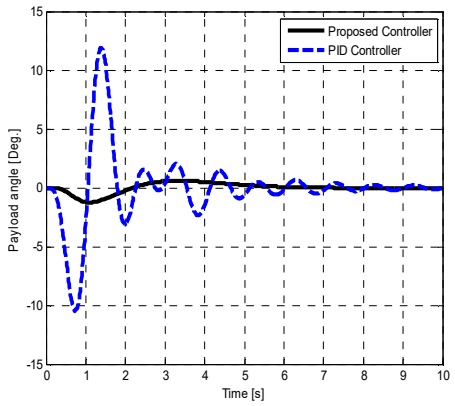


Fig. 5: payload swing angle

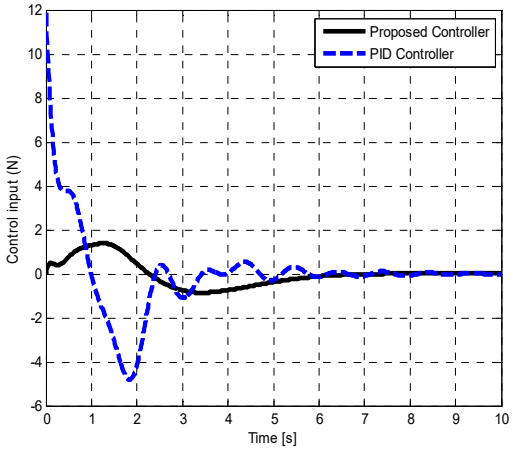


Fig. 6: Control signal

Both controllers tracked the trolley position as well as the swing angle as shown in the figures above. From Fig. 3 it can be seen that; the proposed controller settled the trolley position in 2.3816 seconds with an overshoot of 2.0465% whereas the PID controller settled the trolley position in 4.0032 seconds with an overshoot of 5.3417%.The proposed controller damped out the hook oscillation in 5.6849 seconds with a peak swing angle of 1.0853^0 whereas the PID controller damped out the hook oscillation in 7.8155 seconds with a peak swing angle of 6.0319^0 as shown in Fig. 4. The proposed controller damped out the payload oscillation in 5.5812 seconds with a maximum swing angle of 1.2470^0 whereas the PID controller damped out the payload oscillation in 8.3495 seconds with a maximum swing angle of 11.9157^0 as shown in Fig. 5. From Fig. 6, it can be seen that the initial control input required by the PID control scheme is higher when compared with that required by the proposed controller. Table 3 summarizes the simulation results.

Table 3: Results comparison for 0.6m step reference input

Controllers		PID	Proposed
Trolley position	*Overshoot (%)*	*5.3417*	*2.0465*
	Settling time(s)	*4.0032*	*2.3816*
Hook oscillation	$\theta_{1max}(deg)$	*6.0319*	*1.0853*
	Settling time(s)	*7.8155*	*5.6849*
Payload oscillation	$\theta_{2max}(deg)$	*11.9157*	*1.2470*
	Settling time(s)	*8.3495*	*5.5812*

Case II: For reference step input of 1.0m, the responses are shown in the Fig. 7 through Fig 10.

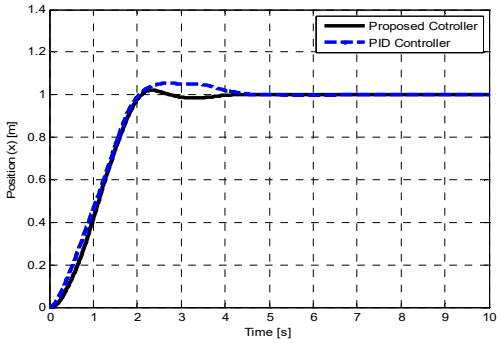


Fig. 7: Trolley position

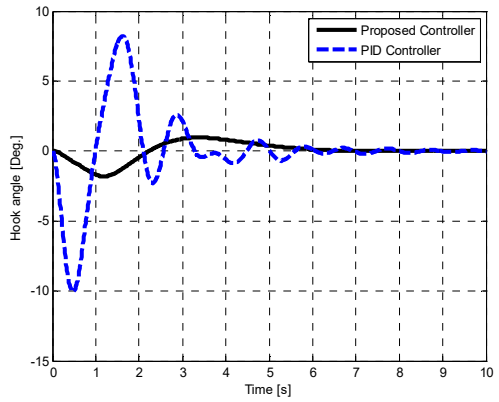


Fig. 8: hook swing angle

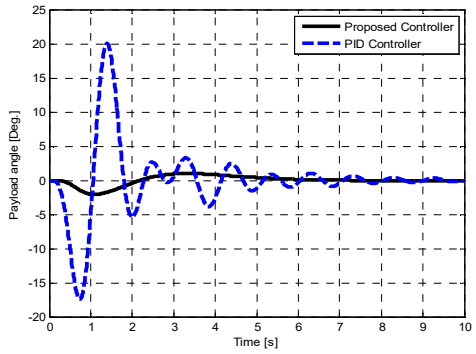


Fig. 9: payload swing angle

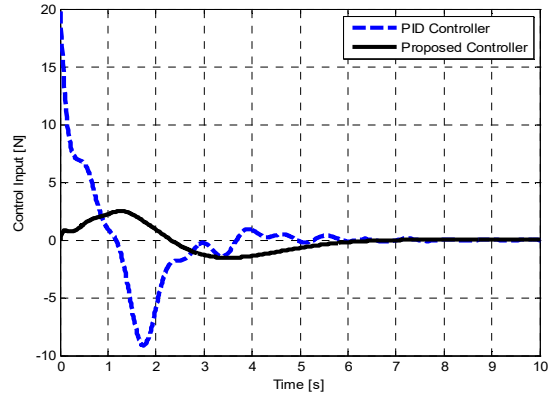


Fig. 10: Control signal

Increasing the step reference input from 0.6m to 1.0m, with the proposed controller; the settling time for the trolley's position increases from 2.3816 to 2.3875 seconds, the overshoot also increases from 2.0465% to 2.0608%, while the PID controller's settling time increases from 4.0032 seconds to 4.0038 seconds, and the overshoot decreases from 5.3417% to 5.3319% as shown in Fig. 7. The proposed controller's hook oscillation damping time increases from 5.6847 seconds to 5.6849 seconds with a peak swing angle of 1.0853^0 whereas the PID controller's damping time of the hook oscillation increases from 7.8155 seconds to 7.8187 seconds with a peak swing angle of hook increases from 6.0319^0 to 10.0515^0 as shown in Fig. 8.The proposed controller damped out the payload oscillation in 5.5815 seconds with a maximum swing angle of 2.0782^0 whereas using PID, the oscillation damping time of the payload increases from 8.3495 seconds to 8.3715 seconds with the peak angle increasing from 11.9157^0 to 20.0148^0 as shown in Fig. 9. From Fig. 10, it can be seen that the initial control input required by the PID control scheme is higher than that required by the proposed controller. Table 4 summarizes the simulation results.

Table 4: Results comparison for 1m step reference input

Controllers		PID	Proposed
Trolley position	*Overshoot (%)*	*5.3319*	*2.0608*
	Settling time(s)	*4.0038*	*2.3875*
Hook oscillation	$\theta_{1max}(deg)$	*10.0515*	*1.0853*
	Settling time(s)	*7.8187*	*5.6849*
Payload oscillation	$\theta_{2max}(deg)$	*20.0148*	*2.0782*
	Settling time(s)	*8.3715*	*5.5815*

B. Payload variation effect

The robustness of the proposed controller is validated with payload variation and the result is

compared with that obtained from the PID control scheme. The simulation was carried out with the payload mass doubled. Fig. 11 through 14 show the simulation results.

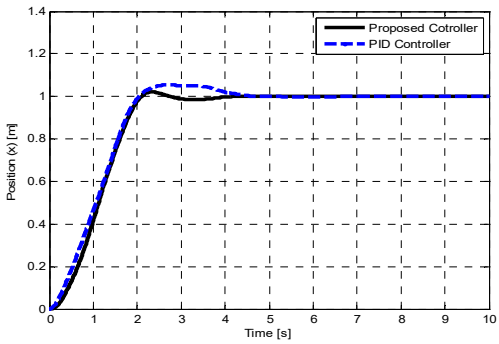


Fig. 11: Trolley position

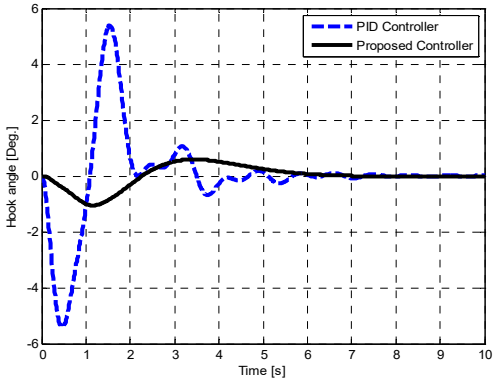


Fig. 12: hook swing angle

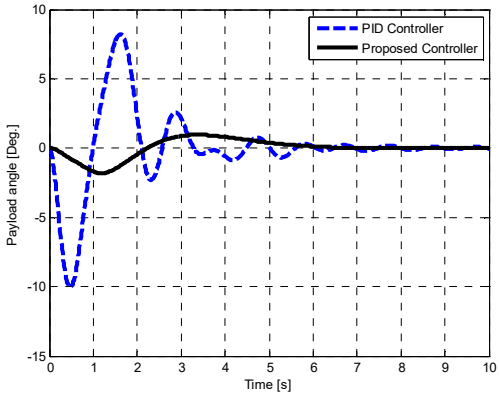


Fig. 13: payload swing angle

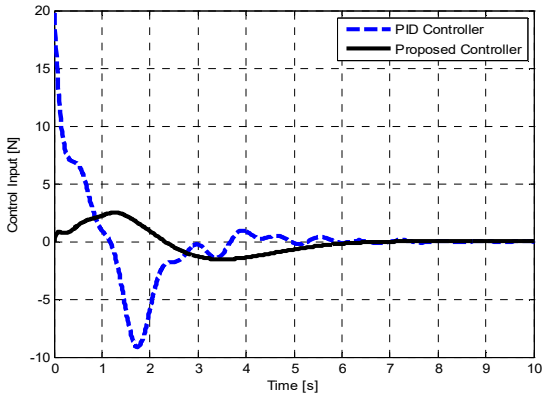


Fig. 14: Control signal

An attempt to double the payload and setting a reference step input of 0.6m. The proposed controller shows that; the settling time for trolley's position increases from 2.3816 seconds to 3.6438 seconds and the overshoot increases from 2.0465% to 2.9169% whereas, using the PID controller; the settling time remains as 4.0032 seconds, but the overshoot increases from 5.3417% to 5.3672% as shown in Fig. 11. With the proposed controller, the damping time of the hook's oscillation reduces from 5.6849 seconds to 5.5240 seconds and maximum swing angle reduces from 1.0853^0 to 1.0650^0. Whereas with the PID controller oscillation damping time reduces from 5.3417 seconds to 5.3315 seconds and the peak swing angle reduces from 6.0319^0 to 5.3852^0 as shown in Fig. 12. With the proposed controller, the damping time for the payload's oscillation increases from 5.5812 seconds to 5.9353 seconds and the maximum swing angle reduces from 1.2470^0 to 1.1492^0 whereas with the PID controller, the payload's oscillation reduces from 8.3495 seconds to 6.4139 seconds and the maximum swing angle reduces from 11.9157^0 to 9.5508^0 too as shown in Fig. 13. Fig. 14 shows the control input under the proposed controller condition; this shows that the initial control input required by the PID control scheme is higher when compared to that required by the proposed controller. Table 5 summarizes the simulation results under this condition.

Table 5: Results comparison for 0.6m step reference input with payload double

Controllers		PID	Proposed
Trolley position	*Overshoot (%)*	*5.3672*	*2.9169*
	Settling time(s)	*4.0032*	*3.6438*
Hook oscillation	$\theta_{1max}(deg)$	*5.3852*	*1.0650*
	Settling time(s)	*5.3315*	*5.5240*
Payload oscillation	$\theta_{2max}(deg)$	*9.5508*	*1.1492*
	Settling time(s)	*6.4139*	*5.9353*

VI. Conclusion

PSO is a robust, simple and very efficient optimization algorithm. In this paper, the optimization of the weighted matrix of the LQR controller was done with PSO, which avoided the choice of the matrices Q and R by trial error. When the optimal LQR controller of the double pendulum crane system was devised, researches and results show that optimization of the LQR controller with PSO has an efficient and satisfactory control effects that are feasible, universal and practical. The proposed method overcomes the shortcomings of the weighted matrices selection difficulty. Also, the simulation results show that meta-heuristic technique illustrates best performance in terms of the effectiveness of the trolley's movement to achieve the desired position with low hook and payload oscillation as compared to a conventional PID controller. Consequently, the proposed controller will be beneficial to many industrial applications when adopted and implemented. Hence it will enhance several production factors such as; economic priorities, safety, reliability as well as speed.

ACKNOWLEDGMENT

I wish to express my sincere gratitude to the Almighty Allah who is above everything (Alhamdulillah) for giving me the health, strength and ability to accomplish this research work. I am especially grateful to Associate Professor Mustapha Muhammad of Mechatronics Engineering Department, Bayero University Kano for his supervision, continued support, invaluable guidance, co-operation, and for providing the necessary facilities and sources during the entire period of this Project. Big thanks to Engr. Muhmmad Sani Dahiru who funded by scholarship and gave me chance to be here. Also the support I received from a friend and a brother in person of Umar Farouk Dahiru is thankfully acknowledged. A warm thanks to my beloved parents, by beloved siblings Haj. Halima Abdulhamid Bako and Yaya Sani Abdulhamid Bako for their endless support and patience.

REFERENCES

[1] A. A. Al-mousa, A. H. Nayfeh, and P. Kachroo, "Control of rotary cranes using fuzzy logic," *Shock Vib.*, vol. 10, pp. 81–95, 2003.

[2] K. A. Alhazza, A. M. Hasan, K. A. Alghanim, and Z. N. Masoud, "An iterative learning control technique for point-to-point maneuvers applied on an overhead crane," *Shock Vib.*, vol. 2014, pp. 1–11, 2014.

[3] W. E. Singhose and S. T. Towel, "Double-pendulum gantry crane dynamics and control," in *Proceedings of the 1998 IEEE International Conference on Control Applications*, 1998, no. September, pp. 1205–1209.

[4] Z. N. Masoud and A. L. I. H. Nayfeh, "Sway Reduction on Container Cranes Using Delayed Feedback Controller," *Nonlinear Dyn.*, vol. 34, pp. 347–358, 2004.

[5] H. I. Jaafar and Z. Mohammed, "PSO-Tuned PID Controller for a Nonlinear Double-Pendulum Crane System," in *Modeling, Design and Simulation of Systems: 17th Asia Simulation Conference, AsiaSim*, 2017, vol. 752, pp. 203–215.

[6] S. Lahres, H. Aschemann, O. Sawodny, and E. P. Hofer, "Crane automation by decoupling control of a double pendulum using two translational actuators," in *Proceedings of American Control Conference*, 2000, pp. 1052–1056.

[7] W. O. Connor and H. Habibi, "Gantry crane control of a double-pendulum , distributed-mass load , using mechanical wave concepts," *Mech. Sci.*, vol. 4, pp. 251–261, 2013.

[8] G. U. O. Weiping, L. I. U. Diantong, Y. I. Jianqiang, and Z. Dongbin, "Passitivity-Based-Control For Double-Pendulum type Overhead Cranes," in *IEEE Region 10 Conference TENCON*, 2004, pp. 546–549.

[9] L. A. Tuan and S. Lee, "Sliding Mode Controls of Double-Pendulum Crane Systems," *J. Mech. Sci. Technol.*, vol. 27, no. 6, pp. 1863–1873, 2013.

[10] H. I. Jaafar, Z. Mohamed, M. A. Shamsudin, N. A. Mohd Subha, L. Ramli, and A. M. Abdullahi, "Model reference command shaping for vibration control of multimode flexible systems with

application to a double-pendulum overhead crane," *Mech. Syst. Signal Process.*, vol. 115, pp. 677–695, 2019.

[11] N. Sun, Y. Fang, S. Member, H. Chen, and B. Lu, "Energy-Based Control of Double Pendulum Cranes," in *The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, 2015, pp. 258–263.

[12] N. Sun, Y. Fang, S. Member, H. Chen, and Y. Fu, "Super-Twisting-Based Antiswing Control for Underactuated Double Pendulum Cranes," in *IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, 2015, pp. 749–754.

[13] M. Zhang, X. Ma, H. Chai, X. Rong, X. Tian, and Y. Li, "A novel online motion planning method for double-pendulum overhead cranes," *Nonlinear Dyn.*, vol. 85, no. 2, pp. 1079–1090, 2016.

[14] H. Chen, Y. Fang, and N. Sun, "A swing constrained time-optimal trajectory planning strategy for double pendulum crane systems," *Nonlinear Dyn.*, vol. 89, no. 2, pp. 1513–1524, 2017.

[15] C. He, F. Yongchun, S. U. N. Ning, and Q. Yuzhe, "Pseudospectral Method Based Time Optimal Trajectory Planning for Double Pendulum Cranes," in *Proceedings of the 34th Chinese Control Conference*, 2015, pp. 4302–4307.

[16] M. Muhammad, A. M. Abdullahi, A. A. Bature, S. Buyamin, and M. M. Bello, "LMI-Based Control of a Double Pendulum Crane," *Appl. Model. Simul.*, vol. 2, no. 2, pp. 41–50, 2018.

[17] M. Zhang, X. Ma, X. Rong, X. Tian, and Y. Li, "Adaptive tracking control for double-pendulum overhead cranes subject to tracking error limitation, parametric uncertainties and external disturbances," *Mech. Syst. Signal Process.*, vol. 76–77, pp. 15–32, 2016.

[18] M. H. Abdel-razak, A. A. Ata, K. T. Mohamed, and E. H. Haraz, "Proportional–integral-derivative controller with inlet derivative filter fine-tuning of a double-pendulum gantry crane system by a multi-objective genetic algorithm," *Eng. Optim.*, vol. 0273, no. May, 2019.

[19] H. I. Jaafar *et al.*, "Efficient control of a nonlinear double-pendulum overhead crane with sensorless payload motion using an improved PSO-tuned PID controller," *JVC/Journal Vib. Control*, vol. 25, no. 4, pp. 907–921, 2019.

[20] N. I. M. Azmi, N. M. Yahya, H. J. Fu, and W. A. W. Yusoff, "Optimization of the PID-PD parameters of the overhead crane control system by using PSO algorithm," *MATEC Web Conf.*, vol. 255, p. 04001, 2019.

[21] M. Sweta and C. Yamini, "A Review on PSO and Association Rule Mining for Item Set Generation," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 1, pp. 480–484, 2017.

[22] A. S. M. Nor, H. Selamat, and A. J. Alimin, "Optimal controller design for a railway vehicle suspension system using particle swarm optimization," *J. Teknol.*, vol. 54, pp. 71–84, 2011.

[23] A. Al-mahturi and H. Wahid, "Optimal Tuning of Linear Quadratic Regulator Controller Using a Particle Swarm Optimization for Two-Rotor Aerodynamical System," *Int. J. Electr. Comput. Energ. Electron. Commun. Eng.*, vol. 11, no. 2, pp. 184–190, 2017.

[24] A. E. Muñoz Zavala, "A comparison study of PSO neighborhoods," *Adv. Intell. Syst. Comput.*, vol. 175 ADVANC, pp. 251–265, 2013.

[25] S. Howimanporn, S. Thanok, S. Chookaew, and W. Sootkaneung, "Design and implementation of PSO based LQR control for inverted pendulum through PLC," *SII 2016 - 2016 IEEE/SICE Int. Symp. Syst. Integr.*, pp. 664–669, 2017.

Problems and Prospects of Internet of Things to the Automobile Industry in Nigeria

[1]Aliyu Mustapha, [1Isah] Abubakar Ndakara,[1]Abubakar Mohammed Idris, [2]Abdullahi Raji Egigogo, [3]Abdullahi Abubakar Kutiriko, [1]Mohammed Abdulkadir, [4] Haruna Dokoro Ahmed

[1]Department of Industrial and Technology Education, Federal University of Technology, Minna, Nigeria

[2]Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria
[3]Department of Informatics, Kings College London
[4]Department of Computer Science, Gombe State Polytechnic, Bajoga, Nigeria

***Abstract*—This study was carried out to identify the problems and prospects of the Internet of Things to the automobile industry in Nigeria. Two research questions were answered. The descriptive survey design was employed and the target population was made up of experts in Automobile Technology and Information and Communication Technology (ICT). The instrument used for data collection was a structured questionnaire. The questionnaire was subjected to face and content validation by five experts; 2 in Automobile Technology and 3 in Computer/ Cyber Security. Cronbach Alpha was used to determine the reliability coefficient of the questionnaire and it was found to be 0.87. The data collected from the respondents were analyzed using mean. The findings on the problems of IoT to the automobile industry in Nigeria include among others security, availability and stability of internet network service. The Findings related to the prospects of IoT in the automobile industry in Nigeria include among others, IoT connects and adds security to the vehicles, it also turns a vehicle into a hub of infotainment. Based on the findings, it was recommended among others that the government should provide financial support to strengthen data confidentiality, availability and stability of Internet network service.**

Keywords— ***Automobile Industry, Internet of Things, Innovation and Technology, Problems and Prospects***

I. INTRODUCTION

The automobile industry is a corporation that involves in the manufacture and overhaul of motorized vehicles together with nearly all components in the motor vehicle, such as the engines and bodies but debar of among others, batteries and fuel. The industry's prime wares are lightweight trucks, pickups, traveller vehicles, vans as well as Sports Utility Vehicles (SUV). Since the days of cart and horse, people have happily possessed and passionately cared for their automobiles. Despite that, technologies such as smartphone, Internet, wireless and cellular communication is softening the impact of the bonds as it widens the frontiers of chances to an unending supply of conveniences and services. Individuals and enterprises alike are adopting the digital innovation, deploying mobile interactive devices to liaise, make choices and smooth the way for purchases [1].

The automotive industry has been around for quite some time and it has developed ever since, but the major shift that is on nowadays from vehicles driven by humans (dependent) to autonomous vehicles (driven by themselves) will keep on taking part in a significant function as transportation changes on a global scale. Today's automobiles have exploded thanks to IoT in two ways, these include, embedded and tethered. The former utilizes a built-in antenna and chipset while the latter use hardware to enable drivers to interface with their automobiles using smart/ cell phones [2]. Moreover, application integration is becoming commonplace in today's automobiles, Google Maps and other route devices have started to supplement works in the Global Position System (GPS) frameworks. For example, GasBuddy demonstrates where the driver can locate fuel in their local area. The industry is on the threshold of change to change to the self-driving automobile industry and the impulsive force behind this is the Internet of Things (IoT). As technology is becoming apparent as a result of swift development in the contemporary wireless telecommunication, IoT has been given a lot of attention and is presumed to promote good to the automobile industry. Though the prevalent use of IoT in the automobile industry is so far in its early stage with the origination of 4G and 5G networks, the demographic, market and digital trends are coming together to transform the automotive industry to expedite swift transmission and processing of assorted data [3].

IoT is one of the leading trends that influence the evolution of Information and Communication Technologies (ICT) by connecting "everything" such as data, people, things and process to the Internet. [4] stated IoT is a turbulent technology where the cyber world run into the physical world. Therefore, IoT is an independent communication between non-living objects to profit human beings. As an emerging paradigm, IoT comes together with three principal innovations namely, things (the objects to trace, things-oriented vision), networking (the linking of the objects to the Internet and connection between the objects, Internet-oriented vision) and representation (the portrayal of the objects on the Internet,

semantic-oriented vision). IoT also links automobile parts and services, machines, people and vehicles to smooth-running the flow of data, permit real-time judgment and ameliorate automotive experiences [3].

In the automobile industry, the "things" in IoT may well be any objects that require to be tracked in practice, for example, products in inventory and cars on road. The "things" also focus on the three basic attributes to be recognizable, able to interact and smart [3]. To encapsulate these attributes, several technologies such as Radio Frequency Identification (RFID) tags and two-dimensional (2D) bar code has been put to use to distinctively identify the statuses such as movement, temperature and location of an object [3]. Sensors fused with communication technologies such as Wireless Sensor Networks (WSN) and RFID Sensor Networks (RSN) can also be employed to trace the state of an object and transfer the data to the Internet.

Manufacturers in the automobile industry such as Honda pioneer the RFID to control the demand chain. CarMax, the monumental automotive retail merchant within the U.S. attaches RFID tags to every vehicle therefore on track the life cycle of a vehicle from buy-in to overhaul and finally to the retail/auction [5]. Similarly, BMW has embraced the intelligent-Drive (iDrive) as an intelligent informatics system that uses assorted tags and sensors. The iDrive supports the environment information and driving data to aid drivers to make an instantaneous judgement while driving on the road. With an integrated GPS, the iDrive tracks the road condition and vehicle location to give driving routes. For instance, If there is a theft of a BMW car, the owner could locate through the BMW tracking system the stolen car. Furthermore, other leading automobile manufacturers such as General Motors, Ford and Toyota also make use of the GM OnStar, Ford SYNC and Toyota G-Book respectively in their recently released models in the market.

The services in the automobile industry ordinarily include sales, marketing, overhauling and recycling of used automobile parts. Through the IoT, the merchandiser-ship is susceptible to track car usage and examine end-user choice in addition to the life cycle to heighten the efficiency and enhance end-user satisfaction [6]. The preceding modality breaks the "normal" pattern to change the automobile services to:

1. Product innovation: this involves the improvement or development of an existing or new product
2. Process innovation: this involves the enhancement of practices used in creating the product.
3. Position innovation: this involves a product that is re-positioned in a transformed condition.
4. Paradigm innovation: this involves the substantial revamping the operation that demands an adjustment in company worth and dominance constitution.

For a respective class above, novelty can stretch out to duo distinct degree, these include the 'Do better' and 'Do different'; the former deals with continuous ground-breaking activities on the same path while that latter deals with the change that transform the actual function activities. In the automobile services, the transformation includes the following types of innovation: product/service, process and paradigm innovation. The innovation may stretch out to independent levels at a separate time frame. IoT, which are demanding new technology, is weird to the African continent and Nigeria in particular. This is borne out of the fact that Nigeria is deficient of the basic Information Technology (IT) infrastructure demands such as stable electrical energy and meagre internet connectivity service for the successful taking up of the technology, therefore, this study intended to address the following:

1. What are the problems of IoT to the Automobile industry in Nigeria?
2. What are the prospects of IoT to the Automobile industry in Nigeria?

II. METHODOLOGY

A descriptive survey was selected to identify the problems and prospects IoT to the automobile industry in the Federal Capital Territory (FCT) Abuja, Nigeria. FCT Abuja was purposely chosen due to the presence of high traffic flux and its strong internet network. The aimed population was 218 respondents, consisting of 145 experts in Automobile Technology in Apo and 73 experts in ICT in Zone 3 Plaza, Neighbourhood centre respectively. Since the population is of manageable size, the entire population was studied; hence no sampling technique was utilized for the study. The instrument used for the data collection was a structured questionnaire which comprises twenty-one (21) items. These include fourteen (14) items dealing with the problems of IoT to the Automobile industry in Nigeria and seven (7) items dealing with the prospects of IoT to the Automobile industry in Nigeria. The study adopted a four-point rating scale using real limit of numbers. These include Strongly Agree (SA)= 3.50-4.00, Agree (A)= 2.50-3.49, Disagree (D)= 1.50-2.49, Strongly Disagree (SD)= 1.00-1.49. To ensure the validity of the instrument five validates validated the instrument; these comprise of three from the field of ICT and two from the field of automobile. This is to make certain that the instrument was able to bring forth the fundamental information needed for the study. To determine the internal consistency of the instrument, Cronbach Alpha (α) was used to establish its reliability. The reliability coefficient was found to be 0.86 and the data collected from the respondents were analyzed using mean.

III. RESULTS

A. Research Question 1

What are the problems of IoT to the Automobile industry in Nigeria?

Table 1: Mean response of the respondents on the problems of IoT to the Automobile industry in Nigeria

N_1=145, N_2=73

S/N	Item	X_1	X_2	X_T	Remark
1	Erratic power supply	3.42	3.34	3.38	A
2	Lack of political will to the genuine growth of ICT	3.37	3.19	3.28	A
3	Corruption	3.73	3.67	3.70	SA
4	Availability of internet service	3.90	3.97	3.94	SA
5	Stability of internet service	3.37	3.40	3.39	A
6	Data privacy	2.90	3.01	2.95	A
7	Security	3.86	3.92	3.89	SA
8	Information secrecy	2.47	2.67	2.57	A
9	Scalability of the storage	2.53	2.52	2.53	A
10	The reputation of service providers	2.49	2.56	2.52	A
11	Liability issue	3.67	3.52	3.60	SA
12	Lack of good software	2.73	2.81	2.77	A
13	Lack of good sensors	3.85	3.75	3.80	SA
14	Lack of good maps	9.69	2.74	2.72	A
	X_g			**3.22**	**A**

Keys: X_1 = Mean of experts in the automobile; X_2 = Mean of experts in ICT; X_T = Average of Mean response of the respondents; X_g = Grand Average of Mean response of the respondents; N_1 = Number of experts in the automobile; N_2 = Number of experts in ICT.

The result in Table 1 revealed that the respondents strongly agreed on item number 3, 4, 7, 11 and 13 and agreed on the item number 1, 2, 5, 6, 8, 9, 10 and 12. Having a grand average of mean of 3.22, This gives the impetus to conclude that the respondents agreed on the problems of IoT to the automobile industry in Nigeria.

B. Research Question 2

What are the prospects of IoT to the Automobile industry in Nigeria?

Table 2: Mean response of the respondents on the prospects of IoT to the Automobile industry in Nigeria

S/N	Item	X_1	X_2	X_T	Remark
1	Pay-as-you-go services	3.97	3.84	3.91	SA
2	Intelligent parking cloud services	3.63	3.59	3.61	SA
3	Revolutionary changes in financing	2.83	3.92	3.38	A
4	Revolutionary changes in insurance	3.59	3.71	3.65	SA
5	Real-time traffic alert	3.83	3.79	3.81	SA
6	Evolution of driverless cars	3.84	3.77	3.81	SA
7	Evolution of connected cars	3.60	3.55	3.58	SA
	X_g			**3.68**	**SA**

The result in Table 2 revealed that the respondents strongly agreed on item number 1, 2, 4, 5, 6 and 7 and agreed on item number 3. Having a grand average of mean of 3.68, This gives the impetus to conclude that the respondents strongly agreed on the prospects of IoT to the Automobile industry in Nigeria

IV. DISCUSSION OF FINDINGS

The inference on the problems of IoT revealed that security, availability and stability of internet network service, are the major problems of IoT to the Automobile industry in Nigeria. This result concurs with [7] that the top critical problem to adopt IoT widespread is security. [8] also corroborated that the main concerns of IoT include among others, accessibility and steadiness of the internet network and invulnerability. [6] also supported the claim that another prominent problem is the global benchmark in protection, secrecy, structure and subject area to give a wide berth to antagonism between and mystification of locally established benchmarks in industry or an establishment.

The findings on the prospects of IoT in the Nigerian Automobile industry revealed that the automobile industry can take the benefit of novel alternatives for competitive demarcation in mobility services that is the Pay-as-you-go services and connected vehicles technology. [9] noted that the IoT can be linked-to reinforce security to the vehicles. The findings also revealed that connected vehicles will be transformed into a centre of infotainment. This is in line with [2] that in time to come, vehicles will be available embedded with telematics that will bring about pleasure in driving. This means that the driver will have applications on the dashboard that would convey real-time trip and traffic reports to the driver, convert speech-to-text to thwart the woes of typing while driving and hand gesture sensors to help avert road traffic crashes.

V. CONCLUSION

Internet of Things is undoubtedly a technology that has come to stay and play a critical part in the automobile industries from vehicles driven by humans to vehicles driven by themselves and also gets connected to smartphones, register real-time traffic alerts and offer emergency roadside assistance. Thus, identifying the problems and prospects of IoT to the automobile industry in Nigeria, IoT will become a player in the automobile industry because of its ability to benefit customers, manufacturers, whole economic system and automobile dealers.

VI. RECOMMENDATIONS

Based on the findings of this study, the following recommendations were made:

1. The government should provide financial support to strengthen data confidentiality, availability and stability of Internet network service.
2. Efforts should be made by introducing the energy-efficient and intelligent vehicle to tackle the bone of contention of safety, environment and energy.

REFERENCES

[1] G. L. Ackerman. Efficacious Technology Management: A Guide for School Leaders.Retrieved 20th June 2019 http://hackscience.net/etm/Efficacious%20Technology%20Management%20ver.%201.2.pdf. 2018.

[2] B. S. Davika and A. Kumari. Review paper on IoT based technology in automobiles. International Research Journal of Engineering and Technology (IRJET*)* vol. 5 No. 3 pp. 3140- 3142. 2018.

[3] K. B. Ruhi., R. Madhuri and A. Himanshu. Internet of Things (IoT) In The Smart Automotive Sector: A Review. IOSR Journal of Computer Engineering (IOSR-JCE) vol. 1 pp. 36-44. 2018.

[4] I. Nick. The impact of the Internet of Things. Retrieved 20th June 2019 from https://www.information-age.com/impact-internet-things-iot-123467503/. 2017

[5] E. Turban., C. Pollard and G. Wood. Information Technology for Management: On-Demand Strategies for Performance, Growth and Sustainability. London: Wiley. 2018.

[6] Q. Erwa., L. Yoanna., Z. Chenghong and H. Lihua. Cloud Computing and the Internet of Things: Technology Innovation in Automobile Service. Berlin Heidelberg: Springer-Verlag Berlin. 2013.

[7] G. Alasdair. IoT Security Issues. Germany: CPI book GmbH, Leck 2017.

[8] F-L. Paula., M. F-C. Tiago., S-A. Manuel., C. Luis. and G-L. Miguel. A Review on the Internet of Things for Defense and Public Safety. Sensors (Basel) *vol.* 16 No. 10, pp. 1644. 2016

[9] A. Nadeem., A. Muhammad., T. Noshina., B. Thar and A. Sohail. A Mechanism for Securing IoT-enabled Applications at the Fog Layer. Journal of Sensor and

Intrusion Detection Based on Support Vector Machine Optimized with Cat Swarm Algorithm

Suleiman Idris
Department of Computer Science,
School of Information & Comm. Tech.,
Federal University of Technology,
Minna, Nigeria.
sidris27@gmail.com

Oyefolahan Ishaq O.
Department of Infor. & Comm. Tech.,
School of Information & Comm. Tech.,
Federal University of Technology,
Minna, Nigeria.
o.ishaq@futminna.edu.ng

Ndunagu Juliana N.
Department of Computer Science,
Faculty of Sciences,
National Open University of Nigeria
Abuja, Nigeria.
jndunagu@noun.edu.ng

***Abstract*—intrusion detection system (IDS) like firewall, access control and encryption mechanisms no longer provide the much-needed security for systems and computer networks. Current IDS are developed on anomaly detection which helps to detect known and unknown attacks. Though, these anomaly-based IDS feature a high false rate. To reduce this false alarm rate, in this paper, we proposed an intrusion detection model based on support vector machine (SVM) optimized with Cat swarm optimization (CSO) algorithm. We use the information gain (IG) for attribute reduction and perform classification using the optimized Support vector. The result obtained shows that our model performs well with the least false alarm rate and good accuracy value compare with other classification algorithms evaluated using the same datasets.**

Chapter 1 *Index Terms*—Intrusion Detection, Support vector machine, Cat Swarm Optimization, Information Gain, NSL-KDD

I. INTRODUCTION

One of the major technological achievement in recent time is the possibility of connecting computer systems for the purpose of sharing resources. Furthermore, the advent of the internet has made it possible for people to communicate from different part of the globe through connected computer networks. However, these interconnection of computer devices came with its own cons. One of the major issues with this technology is in the area of security. computer networks and the internet at large are faced with many security attacks. These attacks aim to compromise the three security goals confidentiality, integrity and availability of any system, network and their resources.

Many protection techniques have been employed to manage the security risks involved with computers and networks. Techniques like encrypting confidential data, access control and software and hardware firewall policies.

However, these techniques are not enough as each one of the techniques possess significant limitations. Therefore, it becomes important to use other additional defense mechanism like intrusion detection system (IDS) [1]. IDS is a software application or a hardware device that is configured to monitors computer system or network for abnormal activities and report or prompt for appropriate action [2]. Many researches have been carried out by researchers to determine an intrusion detection technology with good detection in regards to the accuracy value and minimum training time. Although, many issues still exist with IDS, issues like poor capability for detection, high false positive rate [3].

Many methods have been introduced to improve the performance of IDS in recent times. One of the popular research methods in IDS is support vector machine (SVM). SVM is one of the novels machine learning method that has become a well-known research method in the area of intrusion detection. This is because its generalization performance is good, unavailability of local minimal and it uses minimum time for execution [1]. Although, the performance of Support Vector Machine still depend on how well its parameters are appropriately selected. [1]. If the selection of its parameters are not done appropriately, it will perform poorly. In this study, an IDS that is based on SVM with its parameters optimized using Cat Swarm Optimization algorithm has been proposed.

II. INTRUSION DETECTION SYSTEM

An IDS could be a hardware device or an application software configured to monitor traffics that moves in and out of a computer system or network for activities that are classified as malicious or breach of policy and produces report to a management station. Some of these systems some time may try to completely stop an attempt to get unauthorized access however, it is not compulsory component of a monitoring system [2]. Because of the increasing number of connectivity between computers, intrusion detection has become important in the area of network security [4]. Techniques available for intrusion prevention for instance access control, encryption and firewalls have not provided the security level required to protect systems and networks from increasing security attacks [5]. Therefore, it becomes crucial to deploy an IDS as an additional security measure to detect these security attacks before they course havoc in the system [5]. IDS is developed primarily to detect different kinds of traffics that are malicious and abnormal computer usage that a typical firewall will not be able to detect. The concept of machine learning has been used to develop many IDS. Specifically, integrating two or more learning techniques

have yielded better detection performance compare to a single detection technique [6].

Categorizing IDS can be achieved in different ways. The most common categories or types are misuse based IDS and anomaly-based IDS [7][1].

A. Misuse-based Intrusion Detection System

This is also referred to as signature-based intrusion detection system. This type of detection techniques scanned packets or audit logs and compared with commands or events that are previously known to be a sign of an attack [6]. This type of IDS performs very well in detecting attacks that are previously known. It has a low false alarm rate. However, it performs poor when it comes to detecting new attacks that are not previously known or contained in the database [4][7].

B. Anomaly-Based Intrusion Detection System

This category of IDS is developed based on normal behavior features. It uses these identified features with normal traffics to pinpoint any action that significantly deviates from the normal features. It uses data taken from normal usage to identify patterns [7]. Anomaly IDS make use of patterns associated with behavior that could mean unacceptable activities and analyse previous activities to know whether the observed behavior are normal. [6].

III. Categories or Classes of Intrusion Attacks

Intrusion attacks can be classified into: Remote to local (R2L) attack, User to Root (U2R) attack, Probing attack and Dos or DDoS attacks [8].

A. Remote to Local Attacks (R2L)

In an R2L attack, the adversary aim is to acquire a local right to a machine. To achieve this, the attacker send packets that are capable of compromising the target system over the network, the machine loopholes or vulnerability are then exploited to gain unauthorized access. Attackers with ability to communicate with their target device but have no account on that device uses this kind of attack to exploit weaknesses that exist on the target system to acquire local access on the target device [9]. Attacks like this can be carried out by making use of ports that are open on the target system, using the system loopholes, password guessing [7].

B. User to Root Attack (U2R)

In this type of attack, a normal user tries to escalate his/her privileges by taking advantages of weaknesses found in a system to gain administrative access or root access. This attack is like R2L attacks. The difference is that the attacker here is already a normal user and he/she wants to escalate his/her privilege. [7]. User to root (U2R) attack simply refer to a situation where a legit or normal user wants to gain higher privilege in other to carryout illegal or unauthorized activities. [9].

C. Probing Attack

This class of attack has to do with reconnaissance, gathering information by scanning systems and networks to find weaknesses that exist with them. The found loopholes are used to exploit the systems and networks [9].

D. Dos/DDoS Attack

Denial of Service (Dos) attack often involves attacker sending traffics that are more than what the victim system can handle making such system deny legitimate users' access to services [10]. DoS attack usually originated from a single source. A DoS attack becomes a DDoS attack if the traffics originated from sources more than one [11]. DDoS attacks are usually carried out by deploying many compromised systems (usually called botnet or zombies) to overwhelm their victim [12]. Dos and DDoS attack are attacks targeted at compromising the availability of computer system, router, network and their resources [13]. These attacks are carried out by sending illegitimate traffics capable of draining the system memory or network bandwidth [14][15]. These attacks can be carried out at different layer of the open system interconnection model like the physical layer (the first layer), network layer(the third layer), transport layer (the fourth layer) and application layer (the seventh layer)[12][15] At the physical layer, the attack can simply be to remove a power or network cable connecting a server to the network. Attacks at the network layer are achieved using network layer protocol example of protocol that can be used to achieve this attack at the network layer is the Internet Control Message Protocol (ICMP) [14]. At the transport layer, the attacks can be achieved using layer four protocol like the user datagram protocol which is a connectionless protocol, another protocol that can be used at layer four is the transmission control protocol (TCP) [14]. Hypertext transmission protocol (http) is one major protocol used to carry out denial of service attack this protocol is used at the application layer level. Other protocols used at the application layer to carry out DoS attack are Simple Mail Transmission Protocol, Domain Name System, Voice over internet protocol (VoIP).

IV. Support Vector Machine

Support vector machine is a machine learning algorithm that has gained importance in the area of pattern classification. SVM primarily aim at finding the best hyperplane to divide two classes in a dataset. Several machine learning algorithms exist for dataset classification, SVM standout of these algorithms because of its outstanding generalization capability and its good record for achieving high accuracy level in the training datasets [1].

Classification problem has several major challenges, one of them is the separation of data tending differently, making it difficult for linear separation. [16]. Usually, the dataset is not separable linearly. To overcome this issue of

linearly inseparable datasets, the dataset can be mapped into dimension feature space that is higher and then the hyperplane that separate linearly vectors mapped. That is to say x_i will be substituted with where K gives the mapping with the higher dimension (K is also refer to as the kernel function). Commonly, kernel functions are of three main kinds: polynomial, sigmoid and radial-basis kernel function (RBF) [1].

A. Polynomial kernel function

This kernel can be used to solve problems were the samples for the training datasets are normalized. It is non-stationary. Using this kernel, some parameters have to be settled. The parameters are the gamma slope, r being the constant term and d being the polynomial degree (hence d=3, r=0) [17]. The polynomial function is represented as follows.

$$K(x_i, x_j) = (\sigma x_i^T x_j + r)^d, \sigma > 0 \qquad (1)$$

B. Radial-basis Kernel Function (RBF)

This family of kernel functions have a distance measure smoothed by an exponential function. It maps samples nonlinearly into space dimension that is higher. It is good with instances where attributes and class label do not have linear relations. In addition, the linear kernel can be described as a subset of RBF because, a linear kernel having the penalty parameter C perform similar way with RBF kernel with some parameters (C, Gamma) [17]. The RBF kernel is represented as follows

$$K(x_i, x_j) = \exp(-\sigma \left\| x_i - x_j \right\|^2), \sigma > 0 \qquad (2)$$

One of the parameters that plays a major role is the adjustable parameter represented as σ this parameter should be turned carefully. If it is overestimated, it will cause the exponential to behave like a linear function and the nonlinear power of the higher dimensional projection will begin to. If the adjustable parameter is underestimated, the regulation power of the function will be loss and the boundary for decision will become highly sensitive to noise in training data. Therefore, Support vector machine behavior basically depends on how well the choice of the width parameter σ is made [17].

C. Sigmoid kernel

One requirement of this kernel is that it must satisfies Mercer's theorem, for this to happen, the kernel has to be positive definite. Although, this kernel despite its popular acceptance and usage, it is still not positive semi-definite for some of its parameter's values. Therefore, a carefully chosen parameter for σ, r is very important. If these parameters are not well chosen, it will lead to a very wrong result [17].

$$K(x_i, x_j) = \tanh(\sigma x_i^T x_j + r) \qquad (3)$$

σ can be seen as a parameter that could be measured using scale of the input samples, and r as a shifting parameter, the shifting parameter that controls the threshold of mapping (hence r =0). Generally speaking, RBF and linear kernels are better than the sigmoid function [23].

V. Cat Swarm Optimization

One of the types of optimization problem is feature selection. It is usually achieved by hybridizing a good an optimization algorithm with a classification algorithm. Two commonly used optimal algorithms are Particle Swarm Optimization (PSO) and genetic algorithm (GA). Recently, another optimization algorithm has been proposed Cat swarm optimization (CSO) and it has been proven to perform better compare to PSO [18][19]. CSO was built putting into considerations the behavior of cats, cats are known for hunting excellently and for also showing great level of alertness even at their resting positions. This behavior exhibited by cats can be described or explained by two modes. These modes are: Seeking and Tracing modes [18][20].

A. Seeking Mode

This mode describes the situation of the cat while resting. In this mode, the cat does more of thinking and takes decisions about where to move to next [19]. Four parameters are used to represent seeking mode in the CSO algorithm: one of the parameters is Seeking memory pool (SMP), the second parameter is the Seeking Range of the selected dimension (SRD), third parameter is the count of dimension to change (CDC) and the fourth parameter is Self-Position Consideration (SPC) [21]. Procedure of seeking mode is described below

Step1: produce j replica of the current state of cat_k, where j = SMP. Check SPC if it is true, j = (SMP-1), then accept current status to be one among the candidates.

Step2: For each replica, following the CDC, in no order add or subtract SRD percent of the current values and change existing ones.

Step3: determine the values of the fitness (FS) for all candidates points

Step4: in the case where all FS values are not the same, determine the selecting likelihood of every candidate point by (4), else make all selecting likelihood of every candidate point be 1.

Step5: in no order, choose the position to go to next from the candidate points, and change the position of cat_k."

$$Pi = \frac{|FSi - FSb|}{FSmax - FSmin} \quad where\ 0 < i < J \qquad (4)$$

In a situation where the fitness function aim is to look for the least solution then $FSb = FSmax$, else $FSb = FSmin$.

B. Tracing Mode

This mode describes the situation of the cat while chasing a target. A cat in a tracing mode changes position in accordance with its own velocity for each dimension [21]. The process of tracing mode is explained as follows "Step1: Each dimension (*vk,d*) velocities should be updated following (5).
Step2: velocities should be checked to ensure they are within maximum velocity range. In a situation where the range of the new velocity is over it should be set to be equal to the limit.

$$V_{k,d} = V_{k,d} + r_1 \times C_1 \left(X_{best\ d} - X_{k,d} \right) \quad (5)$$

Step3: the position of cat*k* chould be updated following (6).

$$X_{k\,d} = X_{k\,d} + V_{k\,d} \quad (6)$$

x_{best} represent the state of the cat with the most acceptable fitness value; $X_{k,d}$ is the state of $cat_{k.}$ c_1 represent constant and r_1 represent random value the random values are in the range [0,1]. "

VI. RELATED WORK

In the network intrusion detection algorithm developed by [22], two tree-based classifier models were combined. the random tree and Naïve Bayes tree classifiers. The paper aim is to have a hybrid classifier that can classify traffic entering a network into normal or attack with better accuracy compare to the individual classifiers. The study used the NSL-KDD dataset to assess how well their classifier perform. Detection accuracy of 89.24% was achieved. The future work proposed by the study is to test the effect of reducing the attributes on the training and testing datasets and the detection accuracy.

Also, [23], proposed a framework that detect and mitigate known and unknown distributed denial of service in real time environment using artificial neural network. The study used ANN to detect attack based on some features that separate DDoS attack from normal attack. The ANN was trained with data collected from a network setting that represented a mirror image of a real life network environment. In addition to the data collected from the mirror network, the study used old data to evaluate their work. A detection accuracy of 98% was recorded. The future work would be to train their approach using other dataset and compare the outcome with the outcome they got. Also, their work was not simulated in any network environment, one could simulate their approach to verify the detection accuracy of the work and the false alarm rate.

Bahrami, Bozorg-Haddad and Chu [21] proposed multilayer perception with genetic algorithm to detect DDoS attack at the seventh layer of the OSI model. Four features were considered from traffics entering that exhibit important alteration in their characteristics. The first parameter is the number of hypertext transfer protocol count. Features of hypertext transfer protocol like the GET, POST, OPTIONS, HEAD, DELETE, PUT, TRACE, and CONNECT were analyzed and normal features where recorded. The second parameter is the number of IP address that enters a network within a small-time window. The third parameter is the constant mapping function. The fourth parameter is the fixed frame length. When there is a change in these features, attack will be detected. Experiment result reveal that the technique gave 98.04% accuracy in detecting attacks at the seventh layer (application layer) of the OSI model with high false positive rate of 2.21%. The future work is to work on improving the detection accuracy and lowering the false positive rate.

Enache and Patriciu [1] proposed IDS using SVM combined with information gain. They used Information Gain to select the features of the dataset and the SVM was used for classification. The parameters for Support Vector Machine were selected using Particle Swarm Optimization which optimizes candidate solution through iteration and Artificial Bee Colony developed by observing honey bees behaviour. In other to evaluate the performance NSL-KDD dataset was deployed. The results gotten showed that optimized SVM with PSO or ABC performed better with the dataset compare to the normal SVM. The future work would be to apply other feature selection swarm intelleginces that could do better compare to the ones used here.

Rana et, al. [24] uses fuzziness based on semi supervised approach for intrusion detection system. To improve the classifier performance for the IDS, samples that are not labelled supported with supervised learning algorithm were used. Results gotten from experiment using this method reveal that samples that are not labelled belonging to the categories of low and high fuzziness groups provide the most input to increase the performance of the classifier compared to classifiers that are already existing examples random forest naive bayes, support vector machine. They got an accuracy of 84.12%. The future work proposed is to apply this method to increase the effectiveness of IDSs for detecting many types of attacks.

VII. PROPOSED MODEL

Fig. 1 depict our model. Acquiring the NSL-KDD dataset, Preprocessing and attribute selection are pre-requisite in any intrusion detection work. Our contributions begin with optimizing the parameters of the SVM using Cat Swarm Optimization Algorithm.

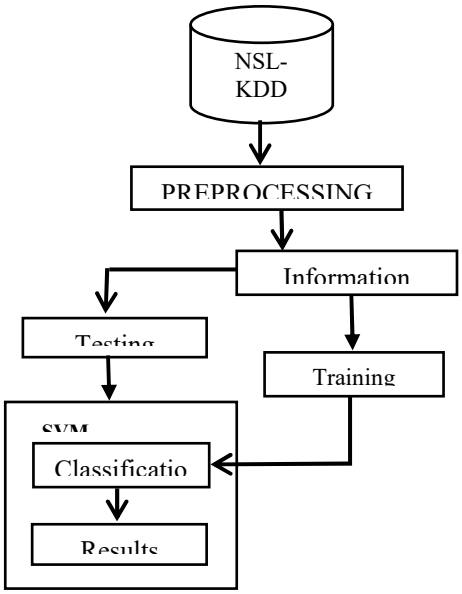


Figure 1: Proposed Model.

A. Data Processing/Feature Selection

In other to evaluate our system, NSL KDD dataset was used. The dataset is a KDD dataset that has been improved upon. It has numerous advantages over the KDD Cup 99. The advantages include: unavailability of redundant record in the train dataset, no duplicate records in the test dataset. The NSL KDD dataset consist of 41 features. However not all of the features that are relevant. Therefore, the need for feature selection. In other to carryout feature or attribute selection, Entropy (information gain) was used. Entropy is a common criterion used in machine learning to individually rank features or attributes with respect to class attributes. IG is calculated by a decrease in the uncertainty of knowing the class feature when the value of the feature is not known. Its idea is based on the principle of information theory usually deployed in ranking and chosing attributes with high value to reduce feature vector size and achieve improved classification with less complexity.

The entropy or information gain of a given feature A with relation to the class feature C, represented as I(C/A), is the decrease in uncertainty with the value of C knowing the value of A. Assuming C and A are whole numbers variables that draw from C = (c_1, …. c_k) and A = (a_1, …, a_n). H(C) is the IG, that calculate the uncertainty about the value of C. H(C/A) is the conditional entropy of C given A, that calculates the uncertainty about the value of C knowing the values of A. Therefore I(C/A) = H(C) - H(C/A) [1].

$$H(C) = -\sum_{I=1}^{K} P(Ct)log2(P(Cl)) \quad (7)$$

$$H(C/A) = -\sum_{j=1}^{n} P(aj) \sum_{i=1}^{k} P\left(\frac{Ci}{aj}\right) log2(P(Ci/aj))$$

where, P(ci/aj) is the posterior probabilities of C given the values of A. therefore I(C/A) is given as follows

$$I(C/A) = -\sum_{I=1}^{K} P(Ci)log2(P(Ci)) - \left(-\sum_{j=1}^{n} P(aj) \sum_{i=1}^{k} P\left(\frac{Ci}{aj}\right) log2(P(Ci/aj))\right)$$

B. Dataset classification

After feature selection, the next stage is dataset classification. We carry out the classification using the optimized SVM with CSO. First, we find the best parameters of the SVM using CSO then we use the optimal parameters to build the training sample as follows.

Step 1: Clearly state the parameters of the algorithm
Step 2: create first cats and velocity in no particular order
Step 3: spread the cats into the two modes tracing and seeking
Step 4: Check if cat is in seeking mode if yes start seeking mode otherwise start tracing mode
Step 5: recalculate fitness function and retain the cat with the best solution in the memory
Step 6: check to know if looping condition is satisfied. If it is,
Stop looping and give out the peak parameter (C and 6) else, return to step 2.
Step 7: use the peak parameter (C, 6) and training sample to build up SVM prediction model.

C. Building SVM Prediction Model with Optimal Parameter (C, 6)

In this work, we use the SVM constructed by Radial basis function (RBF). This family of kernel functions have a distance measure smoothed by an exponential function. It maps samples nonlinearly into space dimension that is higher. It is good with instances where attributes and class label do not have linear relations. In addition, the linear kernel is a subset of RBF because, a linear kernel with penalty parameter C perform the same way with RBF

kernel with some parameters (C, Gamma) [23]. The RBF kernel is represented in (2):

VIII. RESULT AND DISCUSSION

The experiment was carried on a java NetBeans platform with Weka.jar libraries to be able to access weka functionalities. First, we carry out feature selection on the 42 attributes in the NSL KDD dataset to know attributes that have high impacts and those without impact on our prediction.

After attribute selection using information gain (Entropy), some of the attributes have good entropy value while others have insignificant or zero (0) entropy value that is they have no impact on the prediction outcome. Attributes with insignificant entropy values were removed, table 1 shows attribute with good entropy value.

Table 1

Attributes selected after Information Gain

S/N	Attribute Name	S/N	Attribute Name
1	Arc_bytes	11	Count
2	Dst_bytes	12	Logged_in
3	Services	13	Same_srv_rate
4	Flag	14	Rerror_rate
5	Dst_host_srv_count	15	Srv_rerror_rate
6	Dst_host_same_srv_rate	16	Dst_host_srv_diff_host_rate
7	Dst_host_rerror_rate	17	Dst_host_same_src_port_rate
8	Dst_host_diff_srv_rate	18	Srv_fidd_host_rate
9	Dst_host_srv_rerror_rate	19	Dst_host_serror_rate
10	Diff_srv_rate	20	Dst_host_srv_serror_rate

A. Performance Evaluation

The performance of our model was evaluated based on the following metrics:

Accuracy: Proportion of total number of correct predictions

$$\frac{TP + TN}{P + N} \quad (8)$$

Precision: proportion of correct positive observation

$$\frac{TP}{TP + FP} \quad (9)$$

Recall: Proportion of positives correctly predicted as positive

$$\frac{TP}{P} \quad (10)$$

F-Measure: This is derived from precision and recall values. The F-Measure produces a high result when Precision and Recall are both balanced, thus this is very significant.

$$\frac{2 * Recall * Precision}{Recall + Precision} \quad (11)$$

FP Rate: with this model we can know if our model has many false alarms. It is calculated by taking the ratio of misclassified instances to normal instances.

The results obtained from applying our optimized support vector machine on the NSL KDD datasets is presented in table 2

Table 2
Results Obtained

Accuracy	Precision	Recall	F-Measures	FP Rate
96.3	95.4	97.9	96.7	0.02

B. Comparison of Detection Accuracy, Precision, Recall, F-Measure and False Positive Rate with Zero R and Other Classifiers

We compare the performance of our system with Zero R and some popular classification algorithms namely J48, NaiveBayes, RandomTree applied on the datasets. The performance of the algorithms are presented in fig. 2, 3, 4 and 5 table 3 summarizes the result obtained from each of the algorithms.

Table 3

Summary of Results with other Classification Algorithms

Classifier	Accuracy%	Precision %	Recall %	F-Measure%	FP Rate
J 48	95.7	96.1	94.4	95.2	0.32
RandomTree	95.1	95.7	93.3	94.5	0.035
NaiveBayes	84.3	76.3	94.7	84.5	0.24
Zero R	54.6	29.9	54.7	38.7	0.54
CSO-SVM	96.3	95.4	97.9	96.7	0.02

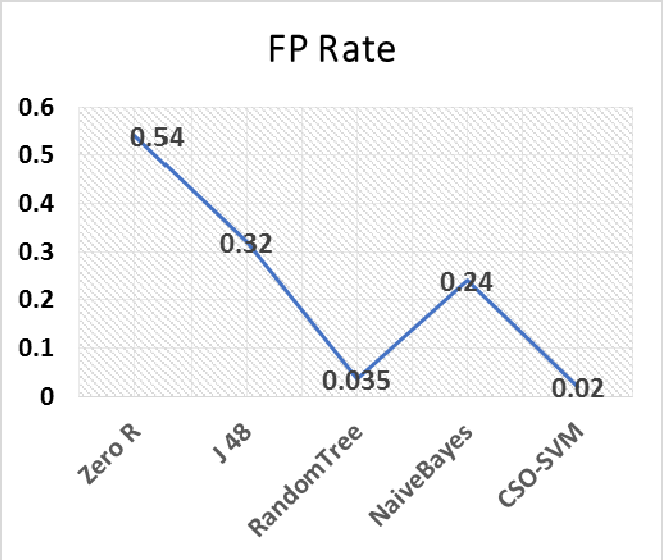


Figure 2: FP Rate

Fig. 2 shows the false positive rate for our classification algorithm and other classification algorithms. CSO-SVM has the lowest false positive rate of 0.02 compare with J48, RandomTree and NaïveBayes with false positive rate of 0.32, 0.035 and 0.024 respectively. While Zero R has the highest value of false positive rate of 0.54.

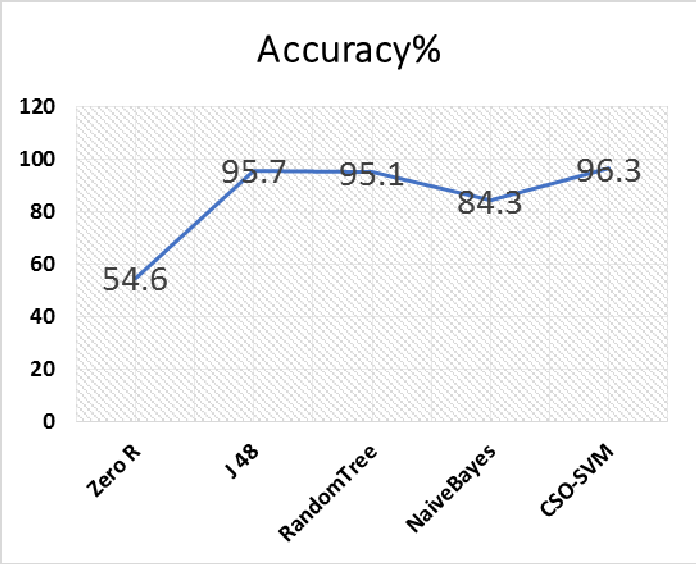


Figure 3: Accuracy

The performance of our algorithm in terms of accuracy in comparison with the Zero R, J48, RandomTree and NaïveBayes is presented in fig. 3, CSO-SVM performs better with accuracy of 96.3 percent compare with Zero R, J48, RandomTree and NaïveBayes with accuracy of 54.6, 95.7, 95.1 and 84.3 respectively.

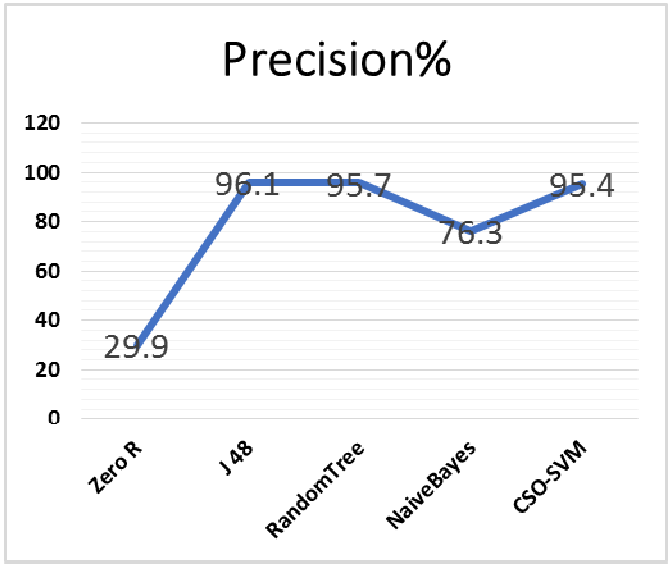


Figure 4: Precision

Interms of precision value, J48 has the highest precision value of 96.1 percent. CSO-SVM has precision value of 95.4. with RandomTree and NaiveBayes having precision values of 95.7 and 76.3 respectively. The base line classifier has precision value of 29.9.

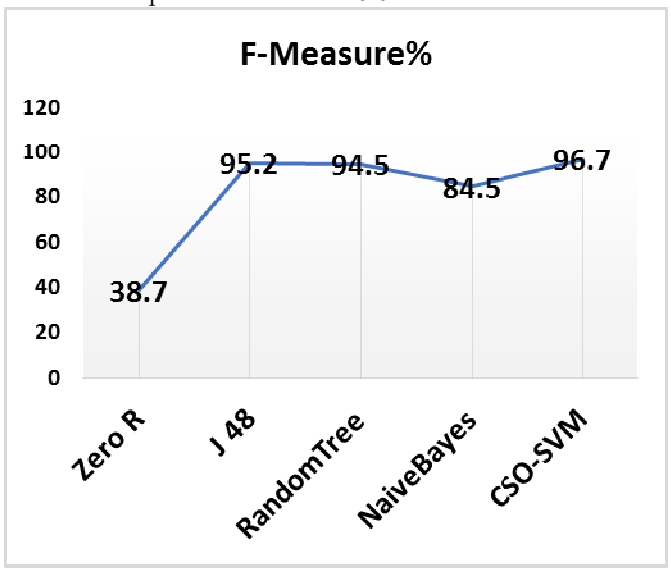


Figure 5: F-Measure

Fig. 5 shows the comparison of the CSO-SVM value compare with the other classification algorithms. F-Measure is high when you have a balanced Precision and Recall value. CSO-SVM has the highest F-Measure value of 96.7 percent followed by J48 95.2. RandomTree and NaiveBayes have F-Measure values of 94.5 and 84.5 respectively. The baseline classifier present an F-Measure value of 38.7.

IX. CONCLUSION

In this research work, we have been able to optimize the performance of support vector machine using Cat Swarm Optimization Algorithm. The NSL-KDD dataset

was used. the entropy value of each of the attributes was calculated with respect to the class value. Attribute with insignificant entropy value were removed during the preprocessing stage. The classification was done with the optimized SVM-CSO. The classification result shows that the CSO-SVM has better performance in all areas compare to the performance of the baseline classifier (Zero R). In terms of accuracy, and F-measure the CSO-SVM performs better compare to other clarification algorithms like the popular J48, Naïve Bayes and RandomTree. Most importantly, the CSO-SVM has low false positive rate of 0.02 compare to IG-PSO-SVM and IG-ABC-SVM with 0.04 and 0.03 respectively.

REFERENCES

[1] A. C. Enache, and V. V. Patriciu, "Intrusions Detection Based On Support Vector Machine Optimized with Swarm Intelligence," *9th IEEE International Symposium on Applied Computational Intelligence and Informatics,* pp. 153 – 158, 2014

[2] K. Amit, C. M. Harish, and M. Rahul, "A Research Paper on Hybrid Intrusion Detection System," *International Journal of Engineering and Advanced Technology,* vol. 2 no. 4, pp. 294 – 297, 2013.

[3] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Elsevier Journal of applied soft computing,* vol. 18, pp. 178 – 184, 2014.

[4] L. A. Wathiq, A. O. Zulaiha, Z. A. N Mohd, "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on ModiÞed K-means for Intrusion Detection System," *Expert Systems With Applications,* 2016.

[5] A. H. Asma, F. S. Alaa, and M. W. Talaat, "Intrusion Detection System Using Weka Data Mining Tool," *International Journal of Science and Research*. Vol. 6, no. 9, pp. 337 – 342, 2015.

[6] L. Wei-Chao, K. Shih-Wen, T. Chih-Fong, "An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge based systems,* 2015.

[7] A. Shadi, A. Monther, and B. Y. Muneer, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science,* 2017.

[8] L. Dhanabal, and S. P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 4, no. 6, pp. 446 – 452, 2015.

[9] R. Jamal, "A survey of Cyber Attack Detection Strategies," *International Journal of Security and Its Applications*, vol. 8, no. 1, pp. 247 – 256, 2014.

[10] P. D. Sheetal, R. H. Priti, A. D. Arundhati, "Denial of Service Attack Defense Techniques," *International Research Journal of Engineering and Technology*, vol. 4, no. 10, pp. 1532 – 1535, 2017.

[11] N. Vani, and P. K. Munivara, "Detection of Anomaly Based Application Layer DDoS Attacks Using Machine Learning Approaches," *i-manager's Journal on Computer Science*, vol. 4, no. 2, pp. 6, 2016.

[12] N. Hoque, H. Kashyap, and D. K. Bhattacharyya, "Real-time DDoS attack detection using FPGA," *Elsevier journal of Computer Communications,* vol. 110, pp. 48 – 58, 2017.

[13] V. K. Yadav, M. C. Trivedi and B. M. Mehtre, "An Approach to Handle DDoS (Ping Flood) Attack'" *Proceedings of International Conference on ICT for Sustainable Development, Advances in Intelligent Systems and Computing,* vol. 408, pp. 11 – 23, 2016.

[14] J. S. Khundrakpam, and D. Tanmay, "MLP-GA based algorithm to detect application layer DDoS attack'" *Journal of Information Security and Applications,* vol. 36, pp. 145 – 153, 2017.

[15] T. Zhiyuan, J Aruna, and H. Xiangjian, (2013). "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," *IEEE transactions on parallel and distributed systems,* 2013.

[16] A. N. Muhammad, B. S. Kudang N. Dodi and M. Akhiruddin, "A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection. *Open Access Journal of Information Science*, vol. 9, no. 5, pp. 418 – 424, 2018.

[17] A. Rimah, B. A. Dorra and E. Noureddine, "Practical Selection of SVM Supervised Parameters with Different Feature Representations for Vowel Recognition," *International Journal of Digital Content Technology and its Applications,* vol. 7, no. 9, 2013.

[18] L. Kuan-Cheng, Z. Kai-Yuan, H. Yi-Hung, C. H. Jason and Y. Neil, "Feature selection based on an improved cat swarm optimization algorithm for big data classification. *Springer Science+Business Media New York, 2016.*

[19] I. Israa and S. Mustafa, "Improved Cat Swarm Optimization for Efficient Motion Estimation," International Journal of Hybrid Information Technology. Vol. 8, no. 1, 279 – 294, 2015.

[20] M. Kumar, S. K. Mishra and S. S. Sahu, "Cat Swarm Optimization Based Functional Link Artificial Neural Network Filter for Gaussian Noise Removal from Computed Tomography Images," *Applied Computational Intelligence and Soft Computing,* pp. 1 – 5. 2016

[21] M. Bahrami, O. Bozorg-Haddad, and X. Chu, "Advanced Optimization by Nature-Inspired Algorithms," *Studies in Computational Intelligence*, pp. 9 – 18, 2018.

[22] K. Jasmin, J. Samed, and S. Abdulhamit, "An effective combining classifier approach using tree algorithms for network intrusion detection," *Neural Computing & Applications*, 2016.

[23] S. Alan, E. Richard and T. R. Overill, "Detection of known and unknown DDoS attacks using Artificial Neural Networks," 2015.

[24] A. R. A. Rana, W. Xi-Zhao, Z. H. Joshua, A. Haider and H. Yu-Lin, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484 – 4 97, 2017

Evaluating GrabCut Algorithm for Cervical Cancer Lesion Segmentation

Suleiman Mustafa
National Agency for Science and Engineering Infrastructure
NASENI
Abuja, Nigeria
mustafa.suleiman@naseni.org

Mohammed Dauda
National Agency for Science and Engineering Infrastructure
NASENI
Abuja, Nigeria
mdsmatt@naseni.org

***Abstract*—Automated segmentation of cervical lesions from plain medical photographs of affected cancerous regions has been evaluated. Cervical cancer is among the most common cancers affecting women in the world, particularly in developing countries due to the high costs of screening. For this reason, recently, a simple inexpensive test by visual inspection with acetic acid (VIA) is used where the cervix region was observed with the naked eye for change in color and texture appearance. This study consideres applying segmentation techniques to the captured images during VIA to be effective in assisting a computer-aided system for diagnosing the cervix region based on color and texture observations. The methodology involved estimating the average cervix positions and generating a segmentation mask to segment an input image into lesions of interest by GrabCut algorithm. For evaluation purposes, each cervix region was manually segmented one after another by hand. Then, those segmented regions were regarded as the correct foregrounds and were compared with pixels identified as the part of foregrounds in GrabCut automatic segmented images by counting overlapping pixels. As a result, GrabCut detected correct foreground and background pixels accurately up to about 87.1% and 88.9% respectively, which shows the potential segmentation ability of GrabCut.**

Keywords— GrabCut, Cervix Cancer, Segmentation evaluation, Mask generation

I. INTRODUCTION

Cervical cancer is the fourth most recurrent cancer in women with an expected 570,000 new cases from 2018 representing 6.6% of all female cancers. Around 90% of deaths from cervical cancer happened in low- and middle-income countries. The high death rate from cervical cancer worldwide could be reduced through a widespread approach that includes prevention, early diagnosis, effective screening and treatment programmes [1]. Its incidence in Sub-Saharan Africa is growing due to lack of or poor screening [2], while in contrast, its prevalence is declining rapidly in Western countries due the accessibility of resources [3]. Cervical cancer screening is commonly achieved by taking a pap smear test, where cells are scraped and collected from the cervix for microscopic examination. This is followed by investigation for irregularities with the aim of detecting potentially precancerous symptoms, i.e., cervical intraepithelial neoplasia (CIN) [4].

The high cost of the equipment used for pap smear screening and the complicated process of collection, sampling, preparation, staining, reading, reporting and the period of time it takes before providing test results, low-tech and inexpensive screening tools are being developed that could considerably reduce the burden of cervical cancer deaths [2]. Therefore, a low cost computer aided cervical cancer screening systems that can be used in the developing countries constitutes a viable alternative and thus has become an interesting research field recently.

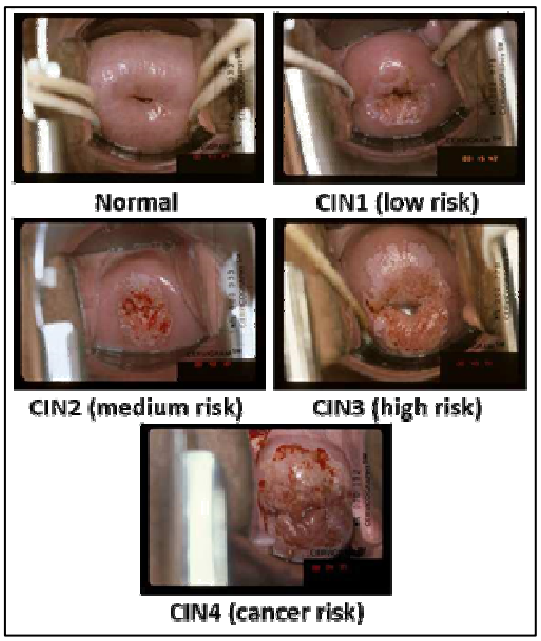
An easier and cost effective alternative screening method is based on the visual acetic acid (VIA) test [3] performed at tissue level rather than cellular level. It involves applying acetic acid to the region and observing change in colour using image processing techniques. The test works in the most remote of settings, bringing access to virtually all locations [4]. Although, VIA performed by well-trained medical personnel has shown the ability to correctly identify between 45% to 79% of cervix cancers [5], there is still room for improvement in the use of computer aided systems.

An essential requirement to the computer aided task of investigating of cervix cancer is segmentation. Previous studies [6]–[12] have highlighted that this step allows for better representation of the region under study and computational methods can provide suitable results for identifying cervix regions. It also allows for effective extraction of shape, colour and texture features used in the classification of cancerous and non-cancerous images. However, to determine the effectiveness of the segmentation method through objective method, is necessary to determine that sufficient regions of the affected part are covered for the diagnosing system to analyse.

II. METHODOLOGY

C. Image Acquisition

The data-set used consist of 1,000 images of which 500 images are of normal (no risk) and CIN1 (low risk) treated as



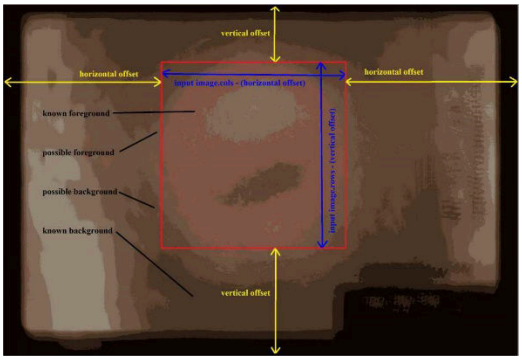
"Negatives", while the remaining 500 images are of CIN2, CIN3 and CIN4 considered as *"Positives"* (medium risk, high risk and cancer risk, respectively). The obtained images have the cervicography medical equipment from which they are captured from in the background of the images as shown in Fig.1. Therefore, effective segmentation is required to eliminate them and outer cervix lesions while preserving the affected part.

Fig. 1. Sample of Cervix images at different stages.

D. Mask Generation for GrabCut

Effective segmentation of the significant regions for differentiating cervix cancer from non-cervix cancer is very important but challenging because of similarities in structure and appearance [12] [13]. For the acquired images, segmentation is done to separate the cervix region (foreground) from the outer lesion (background). GrabCut segmentation [14] was used, for this purpose, to capture as much of the foreground region as possible. It is derivative from graph cuts [15] with the initially identified information about the foreground and the background represented by a rectangular selection around ROI.

Grab Cut method has demonstrated to be effective for segmentation of affected parts in medical images [16]. It usually entails user input by drawing rectangle or providing three variables i.e. foreground, background, and the unknown areas. The proposed method already estimates the mean location of the inputs and suggests possible rectangle regions to simplify the task of identifying the variables for a non-expert or eliminating the need to draw the rectangle. The affected part is located and covers most of the central region



in almost all the acquired images, but the region of the affected part varies for each image as shown in Fig.1. In previous studies involving cervix segmentation, semi-manual methods have been adopted [17], [18]. Therefore, to eliminate the need for user intervention to determine the desired rectangle region of interest (ROI) around the foreground object, a segmentation image mask was created by morphing randomly selected images from our database.

Morphing is achieved by taking two of the colour images f_u and f_v, creating an intermediate morphed image f_{mor} using the following equation:

Fig. 2. Coloured quantized image and our defined ROI.

$$f_{mor}(i,j) = (1-\alpha) f_u(i,j) + \alpha f_v(i,j) \quad (1)$$

where (i,j) represents pixel position and α is a parameter which controls the merger rate between f_u and f_v. In RGB colour space, the vector function F can be denoted $f(i,j) = (R(i,j), G(i,j), B(i,j))$, where *R, G* and *B* are the red, green and blue colour components, respectively.

$\alpha = 0.5$ was used to generate a balanced blend when applying the morphing process (1) iteratively to 100 of our randomly selected images f_u and f_v, then the average of 100 morphed images was taken. As a result of the averaged image being too noisy, additional treatment based on color quantization using *K*-means clustering was performed [19] in order to obtain the quantized image. The quantized image is shown in Fig.2. Based on this technique, four ROIs as shown in Fig.2 were manually determined; areas specified with high probability of known foreground (KF), known background (KB), possible foreground (PF) and possible background (PB). The motive for choosing these areas is that, the foreground area is inside of KF and the background is outside of KB in most of the image data. However, the foreground or background is a little larger (inside of PF or outside of PB) in some of the data. Therefore, an adjustment was made and uses any two of ROIs as the masks of the foreground and background in the subsequent GrabCut segmentation, as necessary.

To determine an optimal number of clusters for *K*-means, silhouette analysis was performed [20], on our 100

randomly selected images. The silhouette shows the measure of how close each point in one cluster is to points in the neighboring clusters and provides a method to decide the number of clusters visually. From the silhouette plot in Fig.3, it shows that using 4 or 16 clusters, which have the highest score among all cluster sizes tested, is suitable for *K*-means transformation of the morphed result to the quantized image.

E. Cervix Region Segmentation using GrabCut

As described in previous section, GrabCut [14] is an extension of graph cut algorithm used as colour image segmentation method. GrabCut was adopted to segment cervix region from every image data. It uses a full-covariance Gaussian mixture matrix (GMM) of *k* Gaussian components to model the foreground and the background, each pixel in the image corresponds to one GMM component

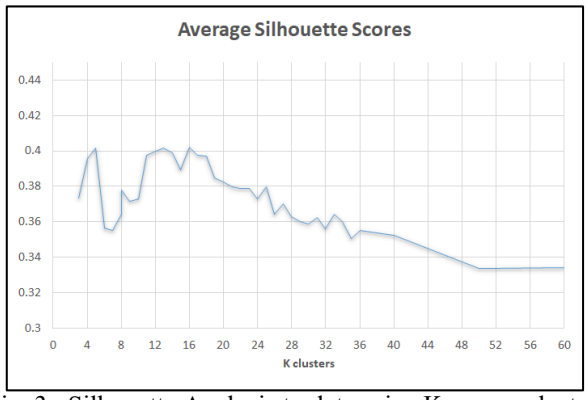


Fig. 3. Silhouette Analysis to determine K-means clusters.

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) \quad (2)$$

$$U(\alpha, k, \theta, z) = \sum_n D(\alpha_m, k_n, \theta, z_n) \quad (3)$$

$$V(\alpha, z) = \gamma \sum_{(m,n)} [\alpha_n \neq \alpha_m] \exp - \beta \|z_m - z_n\|^2 \quad (4)$$

$$\theta = \{\pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k), \alpha = 0, 1, k = 1 \ldots K\} \quad (5)$$

GrabCut redefines Gibbs energy *E* as (3), consisting of image data *Z*, a full covariance Gaussian mixture distribution of *k* components is defined for background pixels *(α = 0)*, data term *U* and smoothness term *V* , defined as equations (4) and (5) respectively. *D* the probability that a pixel belongs to a foreground or background, while θ defines as (5), it indicates all parameter that the algorithm will train. This reflects the discontinuous penalty β between neighborhood pixels *m* and *n* .

For every pixels of an input image either foreground (cervix) or background (non-cervix) is labelled in advance while referring to the masks KF and KB. Also KF is regarded as an initial specified area including cervix region. Then the labelled pixels and the initial region were inputted into GrabCut algorithm. The algorithm estimates an initial Gaussian mixture distribution (GMM) derived from colour information of the foreground and background pixels, and then assigns one of distributions with maximum likelihood among GMM individually to every foreground pixel. The parameters of GMM are updated after finishing all assignments, and background pixels are separated and extracted from the foreground by using minimum cut algorithm [13]. The above procedure is repeated several times, the GrabCut finally outputs well segmented foreground as show in Fig 4.

III. EXPERIMENTS

In this study, almost all tested images are well segmented by selecting ROIs suitably and applying the GrabCut algorithm as show in Fig. 4, compared with manual segmentation. However, in some few cases, as shown in Fig.5. the result becomes either under- or over-segmented due to noise in the image from tissue, characters, marks or light reflection.

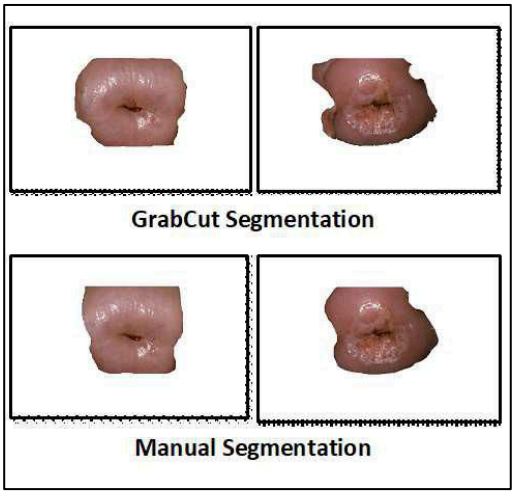


Fig. 4. Example of segmentation results

In Fig.5 the automatic segmented region of the proposed model is indicated by the white boundary. For evaluation in next section, the segmented boundary was manually modified to the appropriate region indicated by the yellow line as shown in Fig.5.

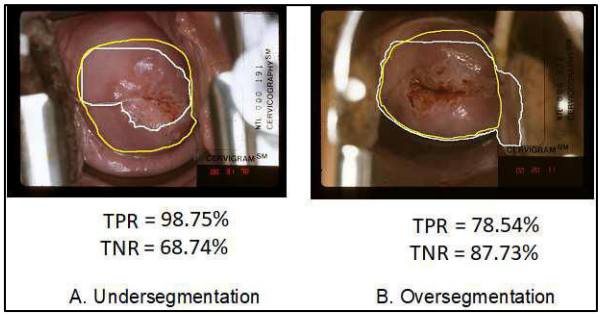


Fig. 5. Examples of poor segmentation results.

To validate the effectiveness of our segmentation we compare the pixel positions of GrabCut (white boundary) with that of the manual segmentation (yellow boundary) as show in Fig.5. For this purpose, pixels correctly recognized as part of the background are true positive (TP), and those correctly recognized as part of the lesion are true negative (TN). However,
lesion pixels in GrabCut images incorrectly identified as part of the background in manual segmented images are (FN), while the background pixels incorrectly identified as part of the lesion is (FP). Furthermore, the True Positive Rate (TPR) and True Negative Rate (TNR) which are generally used for evaluation measures were calculated. In the experiments, TPR is the percentage of accurately detected lesion pixels, while TNR is the percentage of accurately detected background pixels, and their relationship is represented as:

$$TPR = \frac{TP}{TP+FN} * 100\%,\ TNR = \frac{TN}{FP+TN} * 100\% \quad (6)$$

TABLE I DEFINITIONS OF TRUE/FALSE POSITIVE/NEGATIVE.

		Detected pixel	
		Lesion	**Background**
Actual pixel	**Lesion**	True positive (TP)	False negative (FN)
	Background	False positive (FP)	True negative (TN)

The objective measure is established on the notions of true/false positive/negative defined in Table I. For instance, if a lesion pixel is identified as part of the background skin, this pixel is measured to be a false negative (FN). On the other hand, if a background pixel is identified as part of the lesion, it is measured as a false positive (FP).

IV. RESULTS

The comparison of the proposed method with other algorithms evaluated in terms of sensitivity and specificity shown in Table.II, suggests that the adopted methodology captures sufficient lesions for colour and texture feature extraction. However, there have been some either higher TPR or TNR results achieved by other methods, it is necessary to consider both measures collectively. A good measure should consider also a ratio of the detection area compared to the ground truth area, because it may have a detection that covers the whole image, and have a perfect true positive score, but include regions that are part of the background. Furthermore, results show in silhouette analysis, 4 or 16 clusters are suitable choices for the given image data to generate an effective segmentation mask. This is due to the presence of clusters with below average silhouette scores compared to other cluster sizes.

TABLE II. COMPARISON OF SEGMENTATION ALGORITHMS.

Method	**TPR** (Sensitivity)	**TNR** (Specificity)
Liu et al. [7]	89.13%	89.31%
Li et al. [8]	60.77%	93.41%
Zhang et al.[9]	71.15%	81.67%
Gordon et al.[10]	98%	75%
Bai et al. [11]	96.70%	81.99%
Proposed	**87.10%**	**88.11%**

V. CONCLUSION

The use of GrabCut algorithm for automatic segmentation of cervical cancer from cervigraph images of affected cervix regions was successfully evaluated in this paper. Our method consists of GrabCut mask generation from morphing images to derive the average position and segmentation image processing technique. As a result of applying the generated mask to the algorithm, the adopted method accurately detected lesion pixels about 87.1% and non-lesion 88.11% per image, which suggests that the segmentation captures enough lesions for shape, colour and texture feature extraction and can potentially speed up and improve affected lesion classification in computer aided diagnosis as an inexpensive method.

REFERENCES

[1] World Health Organization, Cervical cancer, http://www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/.

[2] J. Jeronimo, P.E. Castle, R. Herrero, R.D. Burk and M. Schiffman, "HPV testing and visual inspection for cervical cancer screening in resource-poor regions," International Journal of Gynecology and Obstetrics, vol. 83, 2003, pp. 311-313.

[3] H. S. Saleh, "Can visual inspection with acetic acid be used as an alternative to Pap smear in screening cervical cancer?," Middle East Fertility Society Journal, vol. 19, no. 3, 2014, pp. 187-191.

[4] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D Forman and F. Bray, “Cancer Incidence and Mortality Worldwide: GLOBOCAN 2012”, International Journal of Cancer, vol. 136, 2015, pp. E359-E386.

[5] C. Rother, V. Kolmogorov, A. Blake, “GrabCut: Interactive foreground extraction using iterated graph cuts,” ACM Transactions on Graphics (SIGGRAPH), vol. 23, 2004, pp. 309-314.

[6] R. Sankaranarayanan, L. Gaffikin, M. Jacob, J. Sellors, S. Robles, “A critical assessment of screening methods for cervical neoplasia,” International Journal of Gynecology and Obstetrics 89: S4-S12, 2005. doi:10.1016/j.ijgo.2005.01.009.

[7] J. Lui, L. Li, L. Wang, “Acetowhite region segmentation in uterine cervix images using registered ratio image,” Computers in Biology and Medicine, vol. 93, 2018, pp. 47-55.

[8] W. Li, J. Gru, D. Ferris, A. Poirson, “Automated Image Analysis of Uterine Cervical Images,” Proc. SPIE 6514, Medical Imaging 2007: Computer-Aided Diagnosis, 651401, 2007; doi: 10.1117/12.736969

[9] S. Zhang, J. Huang, D. Metaxas, W. Wang, X. Haung, “Discriminative sparse representations for cervigram image segmentation,” Proc. IEEE international conference on Biomedical imaging, 2010, pp. 133-136.

[10] S. Gordon, S. Lotenberg, R. Long, S. Antani, J. Jeronimo, H. Greenspan, “Evaluation of uterix segmentation using ground truth from multiple experts,” Australasian Physical and Engineering Sciences in Medicine, vol. 33, 2009, pp. 205-216.

[11] B. Bai, P. Liu, Y. Du, Y. Luo “Automatic segmentation of cervical region in colposcopic images using K-means,” vol. 41, 2018, pp. 1077–1085.

[12] D. Song, E. Kim, X. Huang, J. Patruno, H. Munoz-Avila, J. Heflin, L.R. Long and S. K. Antani.“Multi-modal Entity Coreference for Cervical Dysplasia Diagnosis, ”IEEE Transactions in Medical Imaging, vol. 34, no.1, 2015, pp. 229-245.

[13] Z. Xue, L. R. Long, S. Antani, J. Jeronimo, G. R. Thoma, “ A Web-accessible content-based cervicographic image retrieval system,” Proc. of SPIE vol. 6919, 691907-1, 2008.

[14] C. Rother , V. Kolmogorov, A. Blake, “Grabcut: interactive foreground extraction using iterated graph cuts”, ACM Trans. Graph., vol. 23, no. 3, 2004, pp. 309-314.

[15] Y. Boykov, G. Funka-Lea, “Graph cuts and efficient N-D image seg-mentation,” International Journal on Computer Vision, vol. 70, 2006, pp. 109-131.

[16] S. Mustafa, A. B. Dauda, M. Dauda, "Image processing and SVM classification for melanoma detection", IEEE Computing Network and Informatics (ICCNI), October, 2017.

[17] A. Alush, H. Greenspan and J. Goldberger, “Automated and Interac-tive Lesion Detection and Segmentation in Uterine Cervix Images,” IEEE Transactions on Medical Imaging, vol. 29, no. 2, 2010, pp. 488-501. doi: 10.1109/TMI.2009.2037201.

[18] Q. Ji, J. Engel, E. Craine, “Texture analysis for classification of cervix lesions,” IEEE Transaction on Medical Imaging, vol. 19, no. 11, 2009, pp. 1144-1149.

[19] O. Verevka, J. Buchanan,, “Local K-Means Algorithm for Colour Image Quantization,” Proceedings of the Graphics/Vision Interface Conference, 1995, pp. 128-13

[20] Peter J.Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” Journal of Computational and Applied Mathematics, vol. 20, 1987, pp. 53-65.

Fetal Heart Rate Estimation: Adaptive Filtering Approach vs Time-Frequency Analysis

Ashraf A. Ahmad*, Dominic S. Nyitamen, Sagir Lawan, Chiroma L. Wamdeo
Dept of Electrical/Electronic Engineering,
Faculty of Engineering and Technology,
Nigerian Defence Academy (NDA),
Kaduna, Nigeria.
*aaashraf@nda.edu.ng

Abstract—**Fetal heart rate (FHR) estimation is an essential aspect of fetal heart rate observation utilized to determine fetal health condition during the antenatal period. This paper proposes two methods for FHR estimation from an abdominal electrocardiogram (ECG) signal based on adaptive filtering and time-frequency techniques. The fetal ECG signal is the main aspect of the abdominal ECG marred by other interference/ noise signals. These signals include propagated maternal ECG signal, the power-line noise and the white noise. The adaptive filtering approach is based on the least mean squares (LMS) and requires the maternal ECG signal as reference signal for the filtering processing for FHR estimation. The time-frequency analysis is based on the spectrogram time-frequency distribution (TFD), where instantaneous power (IP) and peak analyses are subsequently carried out from this TFD for FHR estimation. The adaptive filtering is made of a single step but requires an external reference signal while the time-frequency analysis involves more steps but does not require a reference signal. Results obtained showed that both methods are able to estimate the FHR of 160 beats per minute (bpm) accurately at signal-to-noise ratio (SNR) of slightly higher noise at -1dB. Results obtained showed that the proposed techniques are efficient for fetal heart monitoring and choice of technique will depend on available resources.**

Index Terms—**Electrocardiogram (ECG), Fetal Heart Rate (FHR), Instantaneous Power (IP), Least Mean Squares (LMS), Time-Frequency Distribution (TFD), Signal-to-Noise Ratio (SNR).**

I. INTRODUCTION

Estimating the heart rate of a baby (foetus) from a pregnant mother's abdominal ElectroCardioGram (ECG) signal measured at the womb is an important part of Fetal Heart Rate (FHR) monitoring. This estimation helps to determine the health condition of the baby. The condition of the baby may either be normal within average heart rate or in distress. If it is in distress then heart rate will either be lower than the normal average or excessively high and may be associated with pathologic condition such as asphyxia [1]. Thereafter further investigations may be carried out on the baby in distress in order to determine other causes of this heart rate and how to mitigate it. Furthermore, FHR is sometimes estimated from the fetal ECG signal extracted from the abdominal ECG which is usually impeded with various sources of noise and interferences. These sources can be the heart beats of pregnant mother (propagating maternal ECG signal) or the ones from the ECG device (power-line and white noise) or other sources [2, 3]. It is therefore imperative to use appropriate signal processing tools for ECG signal analysis suitable for FHR monitoring.

These tools with unique performance, limitations and objectives include support vector machine [4], spatial filtering [5], extended Kalman filtering [6], Fourier transform [7], wavelet transform [8, 9] and principal component analysis [10] among others. Recently, feature extraction involving segmentation and morphology were used to determine the peaks in an ECG image [11]. Results obtained were validated by cardiologists. Also fetal ECG extraction and analysis was carried out using adaptive noise cancellation and wavelet transformation technique [12]. More recently, a robust method of Empirical Mode Decomposition (EMD) and Multiple Signal Classification (MUSIC) were utilized for FHR estimation carried out in the frequency domain [13]. In the simulated experiment, a low estimation error of 2 beats per minute (bpm) is obtained at Signal-to-Noise Ratio (SNR) of -30dB. Furthermore, a hierarchical probabilistic framework was used for fetal R-peak detection using ECG waveform and heart rate information [14]. Results obtained showed an overall detection accuracy of 99.6% at low SNR for the recordings considered in the paper. A recent comprehensive review and appraisal for various methods involving fetal ECG processing can be accessed in [15].

This paper identifies the two main signal processing tools for heart rate estimation at a good SNR; the adaptive filtering and time-frequency analysis. The filtering technique uses an adaptive algorithm to match the output to a reference signal after iterative process leaving an unmatched part of the output as an error signal. In this paper, the Least Mean Squares (LMS) method is used as the adaptive algorithm and the maternal ECG signal (conventionally obtained from the electrode connected to the mother's chest cavity) is used as the reference signal. Recently, the derivative of the normalized LMS was used for estimation of noise in non-stationary signals [16]. The time-frequency analysis uses a mathematical expression (known as Time-Frequency Distribution (TFD)) to represent in signal of interest in time and frequency domains simultaneously to extract desired information. In this paper, square magnitude of the Short-Time Fourier Transform (STFT), the spectrogram is used to process the abdominal ECG is order to show the various components of the abdominal ECG. Recently, the spectrogram was used to classify the different forms of radar signals based on time-frequency agility [17].

The rest of the paper is organized as follows: section II presented the simulation set-up for FHR estimation from the abdominal ECG signal, while section III presented the steps undertaken based on two techniques for the FHR estimation. Finally results obtained and explanations associated with them are presented in section IV followed by the section on conclusion.

II. SIMULATION SET-UP

The fetal ECG ($F(t)$), maternal ECG signal ($M(t)$) and noise signals forms the abdominal ECG signal ($A(t)$) at the mother's womb. The fetal ECG signal ($F(t)$) is typically of a low signal amplitude of 0.25mV and of normal heart rate of between 110bpm and 160bpm [3]. The maternal ECG signal at the mother's chest cavity serves two purposes in this paper; as a reference signal for the adaptive filtering technique and as a non-linear propagated interference signal to the fetal ECG signal. It has signal amplitude of 3.5mV and heart rate of 89bpm [18]. The sampling frequency in this work is 4KHz based on standard desktop ECG devices [19], which facilitates the conversion of the ECG signal into samples for digital

processing. This heat rate values indicates that more heartbeats are captured in the fetal signal than that of the maternal signal. The noise model for the fetal ECG signal consider for this paper is composed of three sections; white noise ($W(t)$), power-line noise ($P(t)$), and the propagated maternal ECG ($\ddot{M}(t)$). The simulation set-up of the abdominal ECG signal used in this paper is shown in Fig. 1.

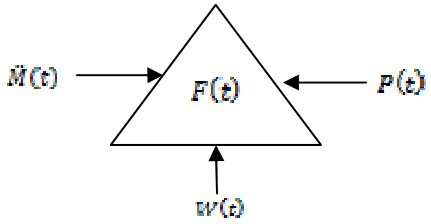


Fig. 1. Pictorial depiction of the abdominal ECG signal model

The two dots above the maternal ECG signal shown in Fig.1 indicate the non-linear propagation of this signal in the form of the interference to the fetal ECG signal at the abdominal level. The power-line noise originates from the ECG measuring device limitations and it is modeled by a sinusoidal equation shown. A signal amplitude of 0.001mV and frequency of 50Hz is chosen in-line with the authors' national power supply standard. Duration of this signal is matched to the duration of the abdominal ECG signal.

The white noise is simply an Additive White Gaussian Noise (AWGN) with three unique characteristics; constant frequency power spectrum, Gaussian probability density function and impulse response autocorrelation [20]. The white noise is obtained by generation of normally distributed pseudorandom numbers of signal length similar to the abdominal ECG signal. The maternal ECG signal of the abdominal ECG signal is modeled around a non-linear propagation from the chest cavity to the abdomen. This non-linear propagation is modeled by a linear Finite Impulse Response (FIR) filter of ten randomized coefficients based on the hamming window [21]. The total noise can therefore be given in (1) and abdominal signal in (2) based on the model discussion in this section.

$$N(t) = P(t) + W(t) + \ddot{M}(t) \quad (1)$$

$$A(t) = F(t) + N(t) \quad (2)$$

Equation (2) shows that the abdominal ECG signal can be simply presented as the sum of the fetal signal and total noise. It is also important to point out that the noise model may be bigger than this model but the paper restricts itself to the most prominent ones.

III. methodology

Two main methods in the form of adaptive filtering and time-frequency analyses are considered in this paper as earlier mentioned as a way of assessing this method's strengths and weakness. Moreover the use of these methods also provides choice for practical implementation based on available resources.

Adaptive filtering varies multiplier coefficients to achieve a targeted optimum performance through the use of an incorporated adaptive algorithm. Varying these coefficients ensures that an error function such as the one in (3) is minimized [22].

$$e(t) = x(t) - y(t) \quad (3)$$

Where $e(t)$ is the error signal, $x(t)$ is a desired reference signal and $y(t)$ is the filter output. The error signal, $e(t)$ is fed into to the adaptive filter through the adaptation algorithm at various increments of time, t. Over this time, output of the adaptive filter, $y(t)$ becomes a better and better match to the desired reference signal, $x(t)$ through this process [23]. In this paper, the basic form of the least mean squares (LMS) adaptation algorithm is used for the fetal heart rate estimation based on the steepest-descent algorithm origin (4).

$$a(t+1) = a(t) + 2\alpha e(t)x(t) \quad (4)$$

Where a represents the multiplier filter coefficients and α represents the convergence factor of limit range (5) in order to ensure the convergence of the algorithm.

$$0 < \alpha < \frac{1}{NE[x^2(t)]} \quad (5)$$

N is the order of the filter, E is the expectation operator and $E[x^2(t)]$ represents the average power of the input signal. Based on the simulation model of this paper, the desired signal is a combination of maternal signal and small amount of white noise to model measurement probe noise. At the end of the adaptive process, the fetal ECG signal is returned as the error signal as the abdominal ECG signal tries to the match the desired/reference signal, i.e. the maternal ECG signal. The LMS algorithm used in this paper is of default 10 multiple coefficients, 0.0001 step size and 1.0 leakage factor. From the obtained fetal ECG signal, the FHR is estimated at the interval between the first and second peak (known as the R-R interval) of the obtained signal using a 75th percentile percent threshold. The threshold is selected in this way so as to cater for the noise associated with the extracted fetal ECG signal. Thereafter, the obtained heart rate in samples is converted and rounded to the nearest figure in bpm using the sampling frequency.

Time-frequency analysis is the usage of mathematical expressions known as time-frequency distribution (TFD) for analysis of time-varying spectra of signals in order to solve arising problems in different fields [24]. The STFT is the oldest

linear TFD that uses sliding window to concentrate signal contents to the time domain [25]. It is mathematically given in (6).

$$F_x^w(t,f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t)\, e^{-j2\pi f\tau} d\tau \qquad (6)$$

Where $x(\tau)$ is the signal and $w(\tau)$ is the window function used for the signal contents concentration. The spectrogram TFD used in this paper is the square magnitude of the STFT and allows for better suppression of noise when compared to the normal STFT. It is mathematically given in (7).

$$S_x^w(t,f) = \left| \int_{-\infty}^{\infty} x(\tau) w(\tau - t)\, e^{-j2\pi f\tau} d\tau \right|^2 \qquad (7)$$

The spectrogram output is a function power, time and frequency and as such further analysis is required before FHR estimation. This paper identified the use of instantaneous power (IP) for this further analysis whose output is a function of power and time only. This IP is obtained from the integral of the TFD (in this case, the spectrogram) with respect to frequency, i.e. time marginal [24]. It is mathematically given in (8).

$$P_i(t) = \int_{-\infty}^{\infty} S_x^w(t,f) df \qquad (8)$$

However, the obtained IP representation of the abdominal ECG contains both the maternal and fetal ECG; as such peak analysis is carried out to extract the fetal R-R interval peaks. The peak analysis is done using the inbuilt MATLAB 'findpeaks' command through specification of minimum peak height of 0.015W, minimum peak distance of 0.1s (400 samples) and suppression of peak height greater than 0.026W. These selected values are based on the observation of IP graph with the objectives of emphasizing the fetal ECG signal peaks. FHR is simply estimated from the first and second peaks by measuring their time separation in samples and converted to the nearest round figure in bpm using the sampling frequency.

IV. RESULTS AND DISCUSSION

To test the performance analysis of the two proposed methods, a standard FHR of 160 bpm was considered at higher noise level SNR of -1dB. The SNR in decibels is obtained mathematically by the expression in (9).

$$SNR(db) = 10log_{10}\frac{P_s}{P_N} \qquad (9)$$

Where P_s is the power of the signal of interest, fetal ECG, and P_N is the power of the total noise/interferences as given in (1). A graphical representation of the abdominal signal containing the fetal ECG and the rest of model noise signals at chosen test SNR is given in Fig. 2.

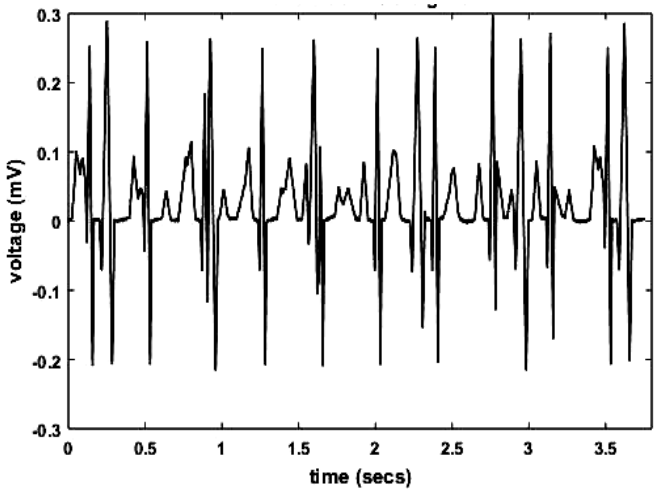


Fig. 2. Graphical depiction of the abdominal ECG at SNR of -1dB

It can be seen from Fig. 2 that there are various peaks and sub peaks in the plot indicating the mixture of different ECG signals and noises. The adaptive filtering technique described in section was then applied to the abdominal ECG signal and the fetal ECG signal is extracted before FHR estimation from this signal. The extracted fetal ECG is shown in Fig. 3.

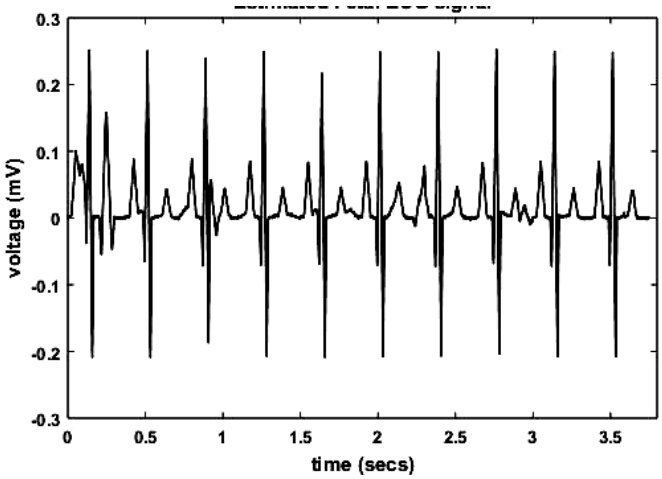


Fig. 3. The extracted fetal ECG signal using adaptive filtering

It is seen from Fig 3 that the fetal peaks are clear and as such FHR is estimated easily by counting the number of samples within the specified threshold of 75 percentile of the maximum peak. The threshold caters for the noise within the fetal ECG signal that the adaptive filtering couldn't cater for. It is also important to point out that fetal ECG of Fig.3 shows different aspect of an ECG signal such as the S and T points of an ECG signal, but however irrelevant to this research as the scope of the paper is restricted to FHR estimation. After successful FHR estimation using the adaptive filtering, the

time-frequency analysis is applied to the abdominal ECG of Fig. 2 using the aforementioned spectrogram and IP is obtained from its output using (8). The graphical representation of obtained IP is shown in Fig. 4.

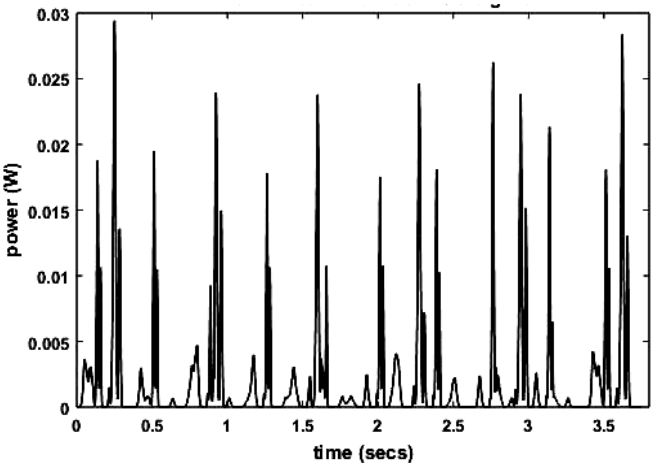


Fig. 4. The IP representation of abdominal ECG signal

It is observed from Fig. 4 that the time-frequency analysis emphasized the peak contents of the abdominal ECG signal containing the fetal and propagated maternal ECG signal. Therefore this paper undertook further analysis involving peak analysis as described in the methodology section carefully guided by the extracted fetal ECG signal of Fig. 3. The stem plot of the refined peaks of the obtained IP is given in Fig. 5 for graphical appreciation.

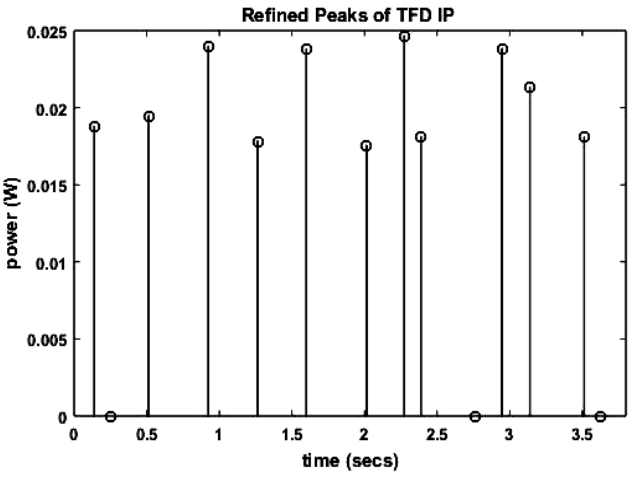


Fig. 5. Refined peaks of the abdonimal ECG IP

It will be observed that peaks of the stem plot of Fig. 5 correspond to those of extracted fetal ECG signal of Fig. 3. This therefore indicated that FHR can easily be obtained by measuring the intervals between the peaks. However, a key difference in FHR estimation from the time-frequency analysis is the absence of threshold requirement as the refinement process has already emphasized the desired R-R peak for the estimation. Furthermore, it can also be observed form Fig. 5. that peak refinement process is not completely accurate due to the failure to suppress some unwanted peaks around between 2nd and 3rd seconds of observation. Further research will involve investigating this limitation and proposing solutions.

V. CONCLUSION

The paper presented two key methods for FHR estimation of a simulated abdominal ECG signal at standard FHR of 160bpm and SNR of -1dB. The adaptive filtering techniques based on LMS algorithm was one of these methods while the other was time-frequency based on spectrogram TFD. FHR is obtained from both methods by measuring 1st and 2nd peak interval (R-R interval) in samples and conversion to bpm with the aid of sampling frequency. Choice of method for accurate FHR estimation will depend on available resources. A reference signal is required for adaptive filtering while more subsequent analyses are required for time-frequency analysis. Further research will explore the estimation accuracies of different FHR values including non-standard ones at different SNRs to determine other strengths and weaknesses of proposed methods. Furthermore, computational complexity analysis as performance indication can also be carried out in order to determine practical implementation complexities.

REFRENCES

[1] P. Rajesh, K. Umamaheswari, and V. N. Kumar, "A Novel Approach of Fetal ECG Extraction Using Adaptive Filtering," *International Journal of Information Science and Intelligent System,* vol. 3, pp. 55-70, 2014.

[2] M. Nasiri, "Fetal electrocardiogram signal extraction by ANFIS trained with PSO method," *International Journal of Electrical and Computer Engineering,* vol. 2, p. 247, 2012.

[3] M. Ghodsi, H. Hassani, and S. Sanei, "Extracting fetal heart signal from noisy maternal ECG by singular spectrum analysis," *Journal of Statistics and its Interface, Special Issue on the Application of SSA,* vol. 3, pp. 399-411, 2010.

[4] S. S. Mehta and N. S. Lingayat, "Support Vector Machine for Cardiac Beat Detection in Single Lead Electrocardiogram," in *IMECS*, 2007, pp. 1630-1635.

[5] M. Xu-Wilson, E. Carlson, L. Cheng, and S. Vairavan, "Spatial filtering and adaptive rule based fetal heart rate extraction from abdominal fetal ECG recordings," in *Computing in Cardiology 2013*, 2013, pp. 197-200.

[6] P. Hamelmann, R. Vullings, M. Mischi, A. F. Kolen, L. Schmitt, and J. W. Bergmans, "An Extended Kalman Filter for Fetal Heart Location Estimation During Doppler-Based Heart Rate Monitoring," *IEEE Transactions on Instrumentation and Measurement (Early Access),* pp. 1-11, 2018.

[7] K. Lweesy, L. Fraiwan, C. Maier, and H. Dickhaus, "Extraction of fetal heart rate and fetal heart rate variability from mother's ECG signal," *World Academy of Science Engineering and Technology,* vol. 3, pp. 599-604, 2009.

[8] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE transactions on Biomedical Engineering,* vol. 53, pp. 2507-2515, 2006.

[9] R. Almeida, H. n. Gonsalves, A. P. Rocha, and J. Bernardes, "A wavelet-based method for assessing fetal cardiac rhythms from abdominal ECGs," in *Computing in Cardiology 2013*, 2013, pp. 289-292.

[10] R. Petrolis, V. Gintautas, and A. Krisciukaitis, "Multistage principal component analysis based method for abdominal ECG decomposition," *Physiological measurement,* vol. 36, p. 329, 2015.
[11] S. Hartati, R. Wardoyo, and B. Y. Setianto, "The Feature Extraction to Determine the Wave's Peaks in the Electrocardiogram Graphic Image," *International Journal of Image, Graphics & Signal Processing,* vol. 9, pp. 1-13, 2017.
[12] P. Sutha and V. Jayanthi, "Fetal electrocardiogram extraction and analysis using adaptive noise cancellation and wavelet transformation techniques," *Journal of medical systems,* vol. 42, p. 21, 2018.
[13] Z. Wei, L. Xiaolong, Z. Jin, W. Xueyun, and L. Hongxing, "Foetal heart rate estimation by empirical mode decomposition and MUSIC spectrum," *Biomedical Signal Processing and Control,* vol. 42, pp. 287-296, 2018.
[14] G. J. Warmerdam, R. Vullings, L. Schmitt, J. O. Van Laar, and J. W. Bergmans, "Hierarchical probabilistic framework for fetal R-peak detection, using ECG waveform and heart rate information," *IEEE Transactions on Signal Processing,* vol. 66, pp. 4388-4397, 2018.
[15] R. Jaros, R. Martinek, and R. Kahankova, "Non-adaptive methods for fetal ECG signal processing: A review and appraisal," *Sensors,* vol. 18, p. 3648, 2018.
[16] S. Rathnakara and V. Udayashankara, "Estimation of Noise in Nonstationary Signals Using Derivative of NLMS Algorithm," *International Journal of Image, Graphics and Signal Processing,* vol. 9, p. 9, 2017.
[17] A. A. Ahmad, A. Saliu, A. E. Airoboman, U. Mahmud, and S. Abdullahi, "Identification of Radar Signals Based on Time-Frequency Agility using Short-Time Fourier Transform," *Journal of Advances in Science and Engineering,* vol. 1, pp. 1-8, 2018.
[18] J. Raj and A. K. Sahu, "Fetal electrocardiogram extraction and analysis." vol. B. Tech Rourkela, India.: National Institute of Technology, 2014.
[19] K. Narayana and A. B. Rao, "Noise removal using adaptive noise canceling, analysis of ECG using Matlab," *International Journal of Engineering Science and Technology,* vol. 3, pp. 2839-2844, 2011.
[20] A. Andreas, "Digital signal processing: Signals, systems, and filters," McGraw-Hill, New York, 2006.
[21] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, J. E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE,* vol. 63, pp. 1692-1716, 1975.
[22] S. S. Haykin, *Adaptive filter theory*: Pearson Education India, 2005.
[23] B. Widrow and E. Walach, *Adaptive inverse control, reissue edition: a signal processing approach*: John Wiley & Sons, 2008.
[24] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*, 2nd ed.: Academic Press, 2015.
[25] A. A. Ahmad, S. A. Ahmad, and A. L. Muhammad, "Recent developments in the use of time-frequency analysis for radar-based applications," in *AFRICON 2015*, 2015, pp. 1-5.

A Resonant Fault Current Limiting Prediction Technique based on Auditory Machine Intelligence

Biobele A. Wokoma
Dept. of Electrical Engineering
Rivers State University
Port-Harcourt, Nigeria
biobele.wokoma@ust.edu.ng

Emmanuel N. Osegi
Dept. of Information Technology
NOUN
Nigeria
emmaosegi@gmail.com

***Abstract*—Faults are a major problem encountered by power system operators particularly single-line-to-ground faults. To mitigate such faults and assure enhanced services to consumers, power system operators need to deploy appropriate hard and soft-computing solutions. In this paper, we present a novel approach to fault mitigation based on a new type of artificial intelligence technique dedicated to time series prediction called Auditory Machine Intelligence (AMI). The actual fault mitigation approach uses a Resonant Fault Current Limiter (RFCL) to fine-tune inductances in circuit in order to estimate the clearance times for a fault. The fault mitigation approach is cast as a time series problem where the resonant inductances (L) and associated clearance times (t_c) are re-sequenced in a temporal aggregated fashion; this approach is then applied to a double-circuit transmission line (Alaoji-Afam sub-transmission) of the Nigerian power network. The results using the proposed technique on a generated L-t_c sequence are compared with that of the Group Method of Data Handling for time series (GMDH $_{\text{time-series}}$) which is a state-of-the-art neural network; the results indicate that the both techniques are competitive but the AMI technique will outperform the GMDH $_{\text{time-series}}$ to the tune of 0.57% for a number of GMDH $_{\text{time-series}}$ and AMI equal simulation trials.**

***Index Terms*— artificial intelligence, fault mitigation, prediction, power systems.**

I. INTRODUCTION

One of the important challenges of electrical power systems is the provision of a reliable fault-free power network. However, due to the complexity of existing power networks and incessant dynamic changes in today's power system network, it is difficult if not entirely impossible to attain a fault-free power system network state. Thus, depending on the fault prevention requirements or particular mitigation solution needed, different solutions exist. For instance, fault mitigation strategies have been proposed based on the idea of optimal power flow and the knocking down of power transmission lines [1] or in the use of High Order Statistics (HOS) and Artificial Neural Networks (ANNs) for protecting transmission lines [2]. These active research directions are typically geared towards devising means or techniques for assuring the security and reliability of the power system network in the context of fault prevention. As the demand for increased reliability and security of power system network increase, the research for alternate fault mitigation approaches is also on the increase; several works abound with different solution context or objectives, as can be found in [3-10]. However, we can still see that the primary objective of these solution methods is to mitigate faults that occur on the transmission lines. In this paper, we investigate the potential of predictive fault mitigation in the context of Resonant Fault Current Limiting (RFCL) protection of power transmission lines during three-phase faults on the Alaoji-Afam sub-transmission.

Our primary objective in this research paper will be to predict the clearance time of a faulted circuit given the amount of resonant inductance inserted into a Resonant Fault Current Limiting (RFCL) circuit.

II. METHODOLOGY

A. Resonant Fault Current Limiting

Resonant Fault Current Limiting (RFCL) is a fault mitigation measure that uses the concept of in-or-out-of-circuit inductance to minimize the short-circuit current that flows in the power system circuit during a three-phase fault. In principle, an RFCL will offer zero impedance in the fault mitigation circuit under normal situations and induce an inductance in the circuit under fault situations.

In an RFCL, the current flow in the FCL part of the fault mitigation circuitry and the voltage drop across the RFCL can be computed using (1) and (2) as:

$$i_k^p = C_k^p \frac{\partial v_{C,k}^p}{dt} + \frac{v_{C,k}^p}{R_k^p} \quad (1)$$

$$v_{C,k}^p = v_k^p - L_k^p \frac{\partial i_k^p}{dt} \quad (2)$$

where,

R_k^p = resistance offered by the Metal-oxide varistor (MOV) or surge arrester

C_k^p = capacitance of resonant circuit

L_k^p = inductance of resonant circuit

v_k^p = voltage across the RFCL circuit

i_k^p = current flowing through the RFCL circuit

$v_{C,k}^p$ = voltage across the capacitor of the resonant circuit

$\frac{\partial i_k^p}{dt}$ = rate of change of current flow in the inductor of the resonant circuit

$\frac{\partial v_{C,k}^p}{dt}$ = rate of change of voltage across the capacitor of the resonant circuit

p = a superscript denoting the phase quantities

k = a subscript denoting the time quantities

A comprehensive detail of the RFCL model applied in this work can be found in [11].

For the purpose of fault simulations, we have adopted the flow diagram of an RFCL simulation model (see Fig.1) which is described as follows:

i. Define the Alaoji-Afam sub-transmission line parameters.

ii. Initiate/or de-activate a fault on say Phase-A in circuit 1.

iii. If fault is initiated in Step-2 above, adjust inductance, L accordingly using a pre-defined value.

iv. Run the simulation model

v. Read and record the clearance time (t_c) for the considered L value.

vi. Repeat aforementioned steps until the desired number of L values and corresponding clearance times (t_c) have been estimated.

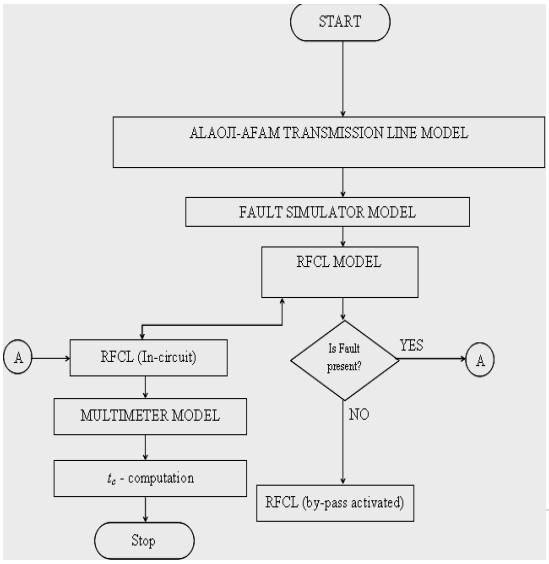


Fig.1: Flow Diagram for Fault Mitigation Simulation

Note that the fault is initiated in fault simulator model and the fault limit in the RFCL model; estimates for clearance times for each L considered are obtained from the multi-meter model after a simulation run.

B. Auditory Machine Intelligence Technique

The Auditory Machine Intelligence (AMI) introduced earlier in [12] under the name "Deviant Learning Algorithm" is a novel type of deterministic artificial intelligence technique for time series prediction. AMI was developed in [13] to exploit the idea and findings about the mismatch negativity effect (MMN) to build an algorithm that can give more precise predictions in a timely manner. This algorithm basically occurs in two-phases: Phase-1 or low-level prediction that make a prediction in the current time step based on a history of data points in the previous time step and a Phase-2 or high-level prediction that performs look-ahead predictions several time steps ahead. The AMI is currently on an experimental phase but the algorithm is sound enough to be applied across different domains or problem types requiring a predictive solution. For full details about this approach, the readers are referred to [12-14].

In the current version, the AMI uses only a Phase-1 prediction. The interesting features of the AMI is that it specifically requires only the computation of a single formula and does not require a stochastic fine-tuning or perturbation to learn on the data; in fact, in the AMI, no parametrization is needed implying that we do not need to fine tune any parameter to learn on the data.

In the AMI, sequences of data points (values) are learnt sequentially or temporally in an adaptive manner and a mean deviant point estimated using (3):

$$S_{dev(mean)} = \frac{\left(\left(\frac{\sum[S_{dev}]}{(n-1)}\right) + S_{deviant}\right) - 2}{n+1} \quad (3)$$

where,

n = number of temporally aggregated sequences

$S_{deviant}$ = the *(n-1)th* value of a temporal aggregated sequence

S_{dev} = the difference between $S_{deviant}$ and S_{stars}

where,

S_{stars} = the *(n-2)th* values of a temporal aggregated sequence

To make a prediction, the formula in equation (4) is used as:

$$S_{pred} = S_{deviant} + S_{dev(mean)} \quad (4)$$

The mathematical treatment of Phase-1 in the AMI is provided in [13].

III. EXPERIMENTAL DETAILS AND RESULTS

A. Experimental details

Experiments have been performed using the MATLAB-SIMULINK software tool. The experiments study the prediction performance of the proposed Auditory Machine Intelligence (AMI) technique in relation to the Group Method of Data Handling (GMDH) technique. These predictions are interpreted in terms of the Mean Absolute Percentage Error (MAPE), a popular metric for assessing the quality of time series data prediction.

The experiments use a SIMULINK model of the Alaoji-Afam sub-transmission (see Fig. 2 in Appendix) for generating L-t_c sequence data used in the predictions; this data can be re-sequenced using the temporal-aggregation procedure proposed in [14]. The re-sequenced data is provided in Table I. The AMI code has been developed in the MATLAB language and is available in the MATLAB central repository.

L (H)	*t (ms)*
0.0020	21.70
0.0100	20.40
0.1000	16.70
0.2000	16.70
0.3000	16.70
0.4000	16.70
0.5000	16.70

TABLE I. L-t_c Sequence Data

1) *The Group Method of Data Handling*

The Group Method of Data Handling (GMDH) is a polynomial root-finding inductive and self-organizing deep learning neural network that was originally invented in [15] to tackle the limitations and inherent challenges faced in the reproduction of reliable multi-layered feed-forward neural networks; this type of neural network is characterized by possessing very deep layers [16]. In this study, we use the GMDH for time series (GMDH $_{\text{time-series}}$) which have been developed in the MATLAB language by the Yarpiz team (www.yarpiz.com).

B. Results

Simulation results considering base parameters of GMDH were performed for 10 different simulation trial runs; the AMI program was also run for the same number of trial runs. The estimated Mean Absolute Percentage Errors (MAPE) for each of the different simulation runs are presented in Table II.

Simulation Step Count	**GMDH$_{MAPE}$**	**AMI$_{MAPE}$**
1	0.1560	0.2501
2	0.2513	0.2501
3	0.1560	0.2501
4	0.1705	0.2501
5	0.2244	0.2501
6	0.4227	0.2501
7	0.4328	0.2501
8	0.2238	0.2501
9	0.4198	0.2501
10	0.1781	0.2501
Mean:	0.2635	0.2501

From the results in Table II, it is clear that the GMDH exhibit stochastic behavior as the MAPE predictions are variable across most of the simulation trials while on the other hand, the AMI exhibit deterministic behavior with a constant MAPE prediction. Also, GMDH gives least and maximum MAPE value on certain trial runs (see MAPE values **0.1560** under GMDH column in Table II). In general, the AMI technique will give a better prediction on the average when compared to the GMDH.

IV. CONCLUSIONS AND FUTURE WORK

A Resonant Fault Current Limiting (RFCL) prediction technique for mitigating faults on a sub-transmission network is presented. The prediction technique is based on an Artificial Intelligence (AI) technique called the Auditory Machine Intelligence (AMI) which is well suited for time series data prediction. The AMI technique was compared to another very popular and well-studied prediction technique called the Group Method of Data Handling (GMDH) for the task of predicting the clearance time given the amount of resonant inductance inserted into the Fault Current Limit (FCL) part of an RFCL circuit. The results of simulations are promising for the new (AMI) technique indicating that deterministic AI techniques with a very simple algorithm can still outperform stochastic alternatives with very complex structures. Therefore, the paper's contribution to knowledge is in the application of this AI technique to fault mitigation on power system sub-transmission networks.

Future research directions include the prediction of the short-circuit level and clearance times under noisy and calm conditions and its application in real-time embedded automatic power systems fault mitigation and control circuitry.

Acknowledgment

The authors gratefully acknowledge the contributions of the Yarpiz Team for making their GMDH source codes freely available.

References

[1]. S. Poudel, Z. Ni, and N. Malla, "Real-time cyber physical system testbed for power system security and control," International Journal of Electrical Power & Energy Systems, vol. 90, pp. 124–133, Sep. 2017.

[2]. D.V. Coury, and J.R.A. Carvalho (2017), "Novel HOS/ANN Approach for Transmission Line Protection Using Current Signals Only", *The 12th Latin-American Congress on Electricity Generation and Transmission - Clagtee 2017*

[3]. N. Mamdouh, R. A. Swief, and M. A. L. Badr, "Application of DSTATCOM coupled with FESS for Power Quality Enhancement and Fault Mitigation," Renewable Energy and Sustainable Development, vol. 3, no. 1, pp. 55–63, Mar. 2017.

[4]. M. Panteli, C. Pickering, S. Wilkinson, R. Dawson, and P. Mancarella, "Power System Resilience to Extreme Weather: Fragility Modeling, Probabilistic Impact Assessment, and Adaptation Measures," IEEE Transactions on Power Systems, vol. 32, no. 5, pp. 3747–3757, Sep. 2017.

[5]. Ł. Staszewski and W. Rebizant, "DLR-supported overcurrent line protection for blackout prevention," Electric Power Systems Research, vol. 155, pp. 104–110, Feb. 2018.

[6]. H. Makwana and B. Bhalja, "Distance Relaying Algorithm for a Line-to-ground Fault on Single Infeed Lines," Electric Power Components and Systems, vol. 42, no. 12, pp. 1227–1238, Jul. 2014.

[7]. M. M. Tripathy, and H. O. Gupta, "Investigation of an advanced compensated Mho relay on double circuit series compensated transmission line," TENCON 2015 - 2015 IEEE Region 10 Conference, Nov. 2015.

[8]. M. R. Aghamohammadi, S. Hashemi, and A. Hasanzadeh, "A new approach for mitigating blackout risk by blocking minimum critical distance relays," International Journal of Electrical Power & Energy Systems, vol. 75, pp. 162–172, Feb. 2016.

[9]. F. Ferdous and Ruma, "Zone Protection System of Transmission Line by Distance Relay using Matlab/Simulink," 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Nov. 2018.

[10]. F. Cadini, G. L. Agliardi, and E. Zio, "A modeling and simulation framework for the reliability/availability assessment of a power transmission grid subject to cascading failures under extreme weather conditions," Applied Energy, vol. 185, pp. 267–279, Jan. 2017.

[11]. A.N. R. L. Sirisha and A. K. Pradhan, "Parameter Estimation of Resonant Fault Current Limiter for Protection and Stability Analysis," IEEE Transactions on Power Systems, vol. 32, no. 3, pp. 2288–2295, May 2017.

[12]. E.N. Osegi, V.I.E. Anireh, "Deviant Learning Algorithm: Learning Sparse Mismatch Representations through Time and Space", hal-01359723v3, 2016.

[13]. E.N. Osegi, V.I.E. Anireh. AMI: An Auditory Machine Intelligence Algorithm for Predicting Sensory-Like Data (2019, in-press).

[14]. E. N. Osegi, V. I. Anireh, and C. G. Onukwugha, "pCWoT-MOBILE: a collaborative web based platform for real time control in the smart space", iSTEAMS SMART-MIINDs Conference, pp. 237-250, June, 2018

[15]. A.G. Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics," Automatica, vol. 6, no. 2, pp. 207–219, Mar. 1970.

[16]. J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, Jan. 2015

APPENDIX

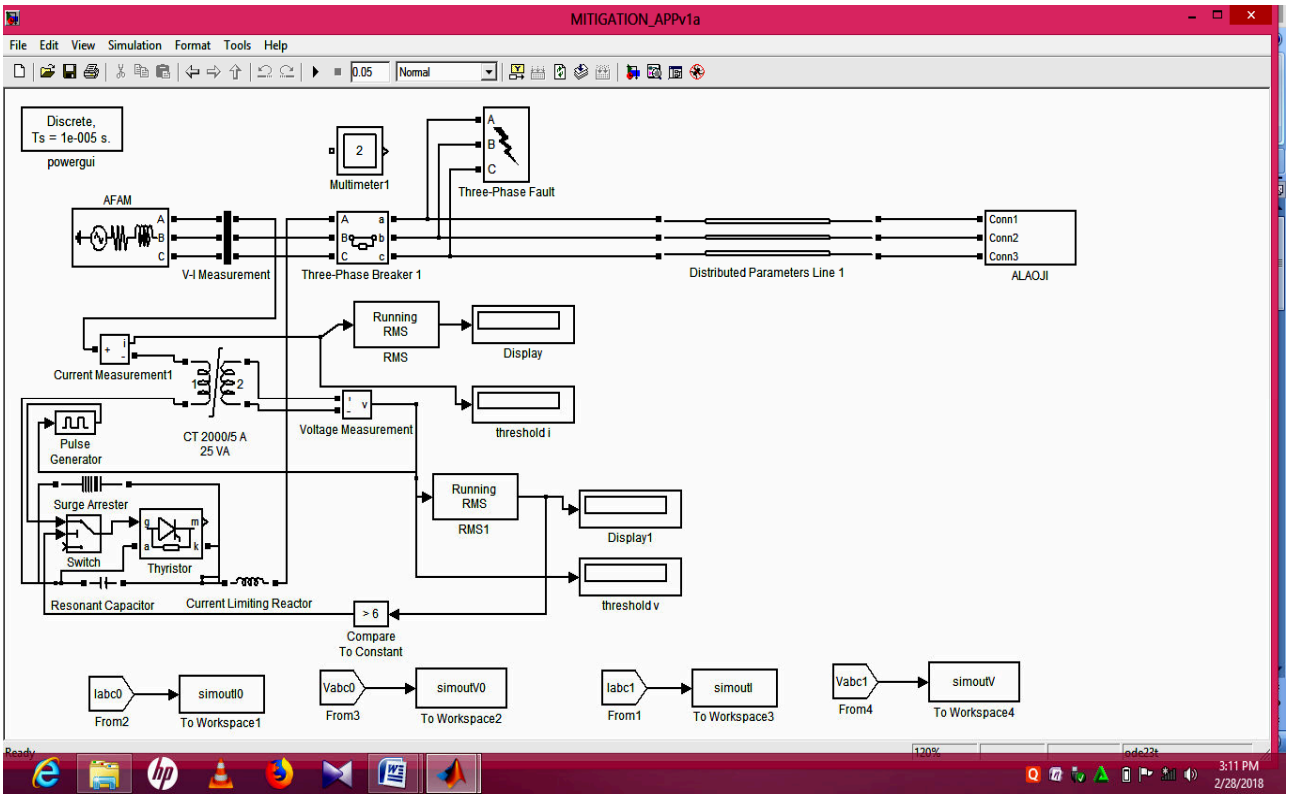


Fig.2: SIMULINK Model of Alaoji-Afam Sub-Transmission

Development of an Improved Intelligent Hybrid Expert System for Diagnosis of Lassa fever

Okoh, N. A. O.
Department of Computer Science
Ambrose Alli University
Ekpoma, Nigeria
augustaokoh@gmail.com

John-Otumu, A. M.
Training, Research & Innovations Unit, Directorate of ICT
Ambrose Alli University
Ekpoma, Nigeria
macgregor.otumu@gmail.com

Nnebe, S. E.
Department of Computer Science
Ambrose Alli University
Ekpoma, Nigeria
nnebese@aauekpoma.edu.ng

Oshoiribhor, E. O.
Department of Computer Science
Ambrose Alli University
Ekpoma, Nigeria
emmaoshor2001@gmail.com

***Abstract*— Lassa fever also known as Lassa Hemorrhagic Fever (LHF) is a zoonotic disease, meaning that humans become infected from contact with infected animals. About 80% of the people who become infected with Lassa virus have no specific symptoms. One in every five infections is severe, in which the virus affects several organs such as liver, spleen and kidney, which has become very worrisome. Research has shown that diagnosing Lassa fever is clinically difficult. In this study, we identified from literature that rule-based, fuzzy logic and Neuro-fuzzy based techniques are the only approach that has been used to develop an expert system for diagnosing the dreaded Lassa fever to aid clinical decisions. However, it is noted that rule-based expert systems cannot be very efficient in the design and development of expert systems based on its shortcomings such as opaque relations between rules, ineffective search strategy, and its inability to learn. While using fuzzy based techniques only for designing expert system is also less efficient when compared to the hybrid approach in terms of learning and other measures. We developed a web-based intelligent hybrid expert system driven by Neural Network (NN), Fuzzy logic (FL) and Case Based Reasoning (CBR) using JavaScript, HTML, CSS, MySQL based on wide coverage area usage. The system was tested using about 17 laboratory confirmed positive cases dataset. The test results revealed that the performance of our model developed is efficient, fast and also consistent with the specialist opinion with about 94.12% classification accuracy rate, and about 5.88% classification error rate.**

Keywords*— *Intelligent, Hybrid Model, Neuro-fuzzy CBR, Expert System, Lassa fever

I. INTRODUCTION

Lassa fever is an acute viral haemorrhagic fever that is awfully infectious [1]. Lassa fever occurs very frequently in different parts of Nigeria and affects approximately 100,000-500,000 persons per year in West Africa [2]. The fever was first exposed and reported in a town called Lassa in Borno State, Nigeria.

According to Eze *et al* [3] the Lassa fever virus is a single stranded RNA virus belonging to the arenaviridae family.

The incubation period for Lassa fever varies between 6 – 21 days. It is indicative and typically characterized by involuntary impulse to vomit, abnormal and frequent stooling, abdominal pain, sore throat, liver enlargement, inflammation of the membranous sac enclosing the heart, low blood pressure, fast heart rate, cough, chest pain, shortest of breath, severe headache, hearing loss, and body temperature above 38°C [4] – [5].

Lassa virus is zoonotic [42] – [43] and infected rodents in the mastomysnatalensis species complex are reservoirs capable of excreting the virus through urine, saliva, excreta and other body fluids to man [3]. Lassa fever presents no definite signs; clinical analysis is often problematic especially at the early onset of the disease. Accurate diagnosis therefore can be assisted with differential laboratory testing, clinical manifestations, epidemiological findings since definitive diagnosis requires investigations available only in highly specialized laboratories [6]. Early diagnosis and treatment of Lassa fever is very vital for survival. It requires specific treatment using the guanosine analogue ribavirin and special caution must be taken to circumvent spread of the disease [7].

Expert systems use human knowledge to solve problems which could normally require human intelligence. These Expert Systems represent the expertise knowledge as data or rules within the computer. These rules and data can be called upon when needed to solve problems. There are so much knowledge embedded in books and manuals, this knowledge cannot be useful if they are not read and interpreted by humans. Majority of Expert Systems are developed through specialized software called shells.

An Expert System shell is a software development environment containing the basic components for building expert system. It does not contain knowledge experts in a particular area. Thus, the shell-based approaches for building a system focus mainly on the system components but little on the user interface, making shell-based systems very suitable for users with programming skills [8].

Artificial Neural Network (ANN) as a tool for medical diagnosis has become the most popular in the last few decades due to its feasibility and accuracy. ANN was developed after getting inspiration from biological neurons [9]. Neural networks employs learning paradigm that includes supervised, unsupervised and reinforcement learning [10]. It has been applied in stock market, prediction, credit assignment, monitoring the condition of machinery and medical diagnosis [11].

Fuzzy Expert System is an Expert System that uses fuzzy logic instead of Boolean logic. Fuzzy Expert System is a collection of fuzzy rules and membership functions that are used to reason about data [12]. Fuzzy logic was introduced in the 1930 by Jan Lukasiewicz, a Polish Philosopher (extended the truth tables between 0 and 1). In 1937, Max Black defined first sample fuzzy set in 1965, Lotfi Zadeh rediscovered fuzziness, identified and explored it [13].

Neuro fuzzy systems are a kind of enhancement over simple fuzzy expert systems. The term neuro-fuzzy derives from the terms ANN and fuzzy logic [14]. The term neuro-fuzzy was first proposed by J.S.R. Jang. A Neuro Fuzzy system is actually based on a fuzzy system with a trained learning algorithm that has been designed from neural network. The learning procedure which is heuristic in nature mainly applies on the host data, and doing modification for making developed training sets with fuzzy system. Neuro-fuzzy is one of the most advanced techniques that are mainly concatenation of two model-neural networks and the fuzzy logic for solving complex problems of this nature [15].

1.1 Statement of the problem

Traditional rule-based expert system have been used over the time past to design expert systems for diagnosing medical conditions such as malaria fever, typhoid fever, leukemia, cerebral palsy, viral infection, influenza [16] - [21].

Literature also shows that rule-based and fuzzy logic are the only approach that has been used to design an expert system for diagnosing the dreaded Lassa fever [22] – [23].

However, due to the shortcomings of rule-based expert systems in the areas of opaque relations between rules, ineffective search strategy, imprecision tolerance, adaptability, knowledge discovery and inability to learn. These shortcomings have made expert systems based on rule base very rigid and less efficient. While using fuzzy based techniques only for designing expert system is also less efficient in its inability to learn but also good in areas like knowledge representation, uncertainty tolerance, imprecision tolerance, and explanation ability. Also, the adaptive neuro-fuzzy expert system developed to predict a suspected case of Lassa fever [49] cannot be fully evaluated based on human expert opinion that is based of laboratory results. Moreover, the system accepts only few input parameters.

As a follow up from above, combining the strength of Fuzzy Logic (FL), Case Based Reasoning (CBR), alongside Artificial Neural Network (ANN) in the areas of learning ability, maintainability, knowledge recovery, adaptability, uncertainty tolerance, and imprecision tolerance is a better choice of forming an intelligent hybrid model for diagnosing Lassa fever that will yield a more effective and efficient result for better decision making by medical practitioners.

1.1 Aim and Objectives of the Study

The overall aim of this study is to develop an intelligent hybrid expert system for diagnosis of Lassa fever using Neuro-fuzzy Case Based Reasoning approach.

The specific objectives of the study are:

i. To capture user's responses and accurately analyze and diagnose the signs/symptoms in order to detect suspected case of Lassa fever using the neural network and case based reasoning techniques.
ii. To further classify the diagnosed suspected Lassa fever case into degree of severity "Mild", "Moderate" and "Severe" using the fuzzy logic model.
iii. To generate result for patient that can aid quick decision making by medical practitioners.

II. RELATED WORKS

In this section, we reviewed about twelve (12) related research work done on expert systems for medical diagnosis using hybridized approach.

Imianvan and Obi [24] designed a cognitive neuro-fuzzy expert system for hypotension control. The system was designed using a set of eight (8) clinical symptoms as input and the output was clustered into Low, Moderate and High. The neuro-fuzzy model was developed using Microsoft Access and Microsoft Visual Basic. Neuro-solution and Crystal report were used in the neural network analysis and graphical representations, while Unified Modeling Language (UML) was used to depict the system in different views. The proposed system is seen to be self-learning, adaptive and is able to handle the uncertainties often seen with the diagnosis of hypotension.

Similarly, Obi and Imianvan [17] designed an interactive neuro-fuzzy expert system for diagnosis of Leukemia. The system was designed using a typical dataset containing fourteen (14) clinical symptoms and the output was structured into three clusters (With Leukemia, Might be Leukemia, Not Leukemia). The neuro-fuzzy model was developed using Microsoft Excel, Microsoft Access and Microsoft Visual Basic. The Neuro-solution and Crystal report were used in the neural network analysis and graphical representations.

Ephzibah and Sundarapandian [25] developed Neuro-fuzzy expert system for heart disease diagnosis. The neuro-fuzzy system incorporated the human-like reasoning style of fuzzy systems through the use of fuzzy sets and a linguistic model consisting of IF-THEN fuzzy rules. The model classification system is seen to be effective in the heart disease diagnosis.

The research work of Tom and Anebo [26] proposed a Neuro-fuzzy model for diagnosis of Monkeypox diseases. A dataset of 18 symptoms were captured to form the rule-base, and the output was designed to be classified into mild, moderate, severe and very severe. A simulation of the proposed model was done using Matrix Laboratory (MATLAB). The system as recommended is a decision support system for young medical practitioners.

In a similar study done by Shaabani, Banirostam and Hedayati [27], a neuro-fuzzy system was implemented to diagnose multiple Sclerosis. Back Propagation and Least Square Error (LSE) were the training techniques used and the K-fold cross validation was used to optimize the input/output. The system was implemented using MATLAB. The simulated results showed that the proposed system has about 96% accuracy.

A Neuro-fuzzy system for the classification of cells as cancerous or non-cancerous was developed by [28]. Images of cells were pre-processed using median filter algorithm, while segmentation was done using marker-controlled watershed algorithm and finally, gray-level co-occurrence was used in the extraction process. The computed topography scan image dataset of the lung cell was downloaded from the cancer imaging archive dataset. ANN Back propagation algorithm was used in training the system, while the cancerous cells were passed into the fuzzy inference engine for classification. Results showed about 70% accuracy and 89% precision.

Egwali and Obi [29] also proposed an adaptive neuro-fuzzy model for diagnosing Ebola Hemorrhagic Fever (EHF). The proposed system used nine (9) clinical symptoms of the EHF and 29 clinical signs and symptoms further classified into 5 Tiers. Back propagation learning algorithm was used in the training. MATLAB and Fuzzy Logic tool box were used to simulate the entire process.

A hybrid neuro-fuzzy expert system for diagnosing thyroid diseases was proposed by [30]. The dataset used in the system was obtained from the UCI machine learning repository, and the dataset were preprocessed using Microsoft Excel to organize and classify the dataset values before being used in the expert system. Back propagation and fuzzy logic were used in the training. Results showed that the system could classify into three categories; which are normal, hyper and hypo.

Goni *et al* [31] developed an intelligent system for diagnosing Tuberculosis using adaptive neuro-fuzzy methodology. About eleven symptoms of tuberculosis which are persistent were used and the system was able to categorize the severity level as mild, moderate, severe and very severe. A yes or no served as input to the adaptive neuro-fuzzy inference system. MATLAB 7.0 was used to simulate the entire process.

In similar manner, a neuro-fuzzy approach for diagnosing and controlling tuberculosis was also proposed by [32]. The system consists of 11 input variables as symptoms and a rule base that consist of about 120 rules to determine the output parameter. The center of gravity method for defuzzification was also used in the research work. Simulation of the system was also done using MATLAB 7.0

Also, Maskara, Kushwaha and Bhardwaj [33] emphasized on the power of neuro-fuzzy system in developing an expert system for diagnosing diseases to the India people who cannot afford to consult the expert doctors owing to high consultation fees and unavailability of expert in rural areas. Results revealed that the expert system is fast and perfect for disease diagnosis and recommended for any patient or doctors.

An ARM Cortex-M3 Based Interactive Neuro Fuzzy Expert System was proposed by [34] for diagnosing Breast Cancer in order to assist physicians, radiologist and others in clinical diagnosis. They designed the breast cancer detection rules using digital Mammographic dataset. The proposed system used a wide range of Breast Imaging Reporting and Data System (BIRAD) classification schemes in order to enhance the level diagnosis accuracy.

A case of lung disease prediction based on the observed symptom value using hybrid neuro-fuzzy system was also proposed by [35]. The significant symptoms were identified based on Pearson's correlations performed on all the observed data. The proposed system achieved a sensitivity, specificity and accuracy of 100%, 75% and 95% respectively. The experimental results revealed that the performance of the proposed system was comparable with other standard measures.

III. METHODOLOGY

This section discusses the area of study of the research work, the data gathering methods used and the research design methods.

3.1 Area of Study

The research work was conducted in Irrua Specialist Teaching Hospital (ISTH) formerly Otibhor Okhae Teaching Hospital (OOTH). ISTH was established by Decree 92 of 1993 as the 14th Teaching Hospital in Nigeria to provide tertiary health care delivery services to the people of Edo State and beyond.

The hospital is located in Irrua town, Edo Central Senatorial District, along the Benin-Abuja highway at about 87 kilometers north of Benin City, the Edo State Capital. The Location of the Hospital on the Benin-Abuja highway has positioned it to become a notable Centre for the Treatment of

Accident Victims and the dreaded Lassa fever at the teaching hospital Institute of Lassa Fever Research and Control (ILFRC). The hospital was commissioned on the 21st of November 1991 by the former Vice President of the Federal Republic of Nigeria, Admiral Augustus Aikhomu. Clinical activities did not however, commence until May, 1993.

3.2 Sources of Data

The data for this study was gathered using both primary and secondary sources.

The secondary data came from Journals, conference proceedings and documentary which were used as the basis for forming most of the twenty-nine signs and clinical symptoms established (See Table 1)

TABLE 1: Signs and clinical symptoms of Lassa fever

Symptoms	Code
Involuntary impulse to vomit [40 – 41], [46 - 48]	P01
Abnormal and frequent stooling [40 - 41, 48]	P02
Abdominal pain [40 – 41], [46 - 48]	P03
Infrequent and difficulty in stooling	P04
Difficulty swallowing	P05
Liver enlargement [4] – [5]	P06
Yellow discoloration of the skin, whites of the eyes, etc	P07
Inflammation of the membranous sac enclosing the heart	P08
High blood pressure	P09
Low blood pressure	P10
Fast heart rate [4] – [5]	P11
Cough [43]	P12
Chest pain [43]	P13
Shortness of breath	P14
Sore throat [40] – [42]	P15
Abnormal sleepiness or sleeping sickness	P16
Severe headache, and stiffness of the neck or back muscles [4] – [5]	P17
Hearing loss involving one side only or double sides [4-5, 41]	P18
Loss of attention or sleepiness or an outburst of great uncontrollable laughter with loss of consciousness	P19
Body temperature > 37 °C [41]	P20
Dehydration	P21
General body weakness [41]	P22
Abnormal large amount of protein in the urine [41, 48]	P23
Swollen face [40] – [41]	P24
Mucosal bleeding (Mouth, Nose, and Eyes) [44-45]	P25
Abnormal decrease in the number of blood platelets	P26
Internal bleeding [40]	P27
Frequent backache [4-5, 41]	P28
Confusion or disorientation	P29

The primary source of data came from both observation and key informant interview.

We observed the Polymerase Chain Reaction (PCR) procedures among other methods used in the laboratory for diagnosing Lassa fever. We also had detailed interview sessions with both medical doctors and medical laboratory scientist in the key domain area of our research. The information gathered assisted to a large extent in forming the basis of establishing the weight values assigned to the different symptoms used for computation by the neural network (See Tables 2, 3 and 4) and the clustering into different degrees of severity using the fuzzy rules (See Figure 2).

TABLE 2: Weight value for the first stage diagnosis

Code	Fixed input value	Responses		
		Yes	No	Not Sure
P01	1	2	0.5	1
P02	1	2	0.5	1
P03	2	2	0.5	1
P04	1	2	0.5	1
P05	1	2	0.5	1
P06	2	2	0.5	1
P07	2	2	0.5	1
P08	2	2	0.5	1
P09	1	2	0.5	1
P10	2	2	0.5	1
P11	1	2	0.5	1
P12	1	2	0.5	1
P13	1	2	0.5	1
P14	1	2	0.5	1
P15	1	2	0.5	1
P16	1	2	0.5	1
P17	1	2	0.5	1
P18	2	2	0.5	1
P19	1	2	0.5	1
P20	2	2	0.5	1
P21	1	2	0.5	1
P22	2	2	0.5	1
P23	2	2	0.5	1
P24	2	2	0.5	1
P25	2	2	0.5	1
P26	2	2	0.5	1
P27	2	2	0.5	1
P28	1	2	0.5	1
P29	2	2	0.5	1

TABLE 3: Weight value for the second stage diagnosis (BP and Temp)

Code	Input value	Value < 100	Value > 100
P30	2	2	1
Code	Input value	Value > 39	Value < 39
P31	2	2	1

TABLE 4: Weight value for the third stage diagnosis

Code	Various Input Weights			
	Fixed input value	Yes	No	Not Sure
P32	2	2	0.5	1
P33	2	2	0.5	1
P34	2	2	0.5	1

3.3 The Proposed Neuro-fuzzy CBR Architecture

System architecture can be referred to as the intangible model that defines the arrangement, comportment, and more views of a system [36]. It is also known as a formal description and representation of a system which is organized in a way that supports reasoning about the structures and behaviors of the system. A system architecture consist of system modules and the sub-systems developed, that will work together to implement the overall system [37] – [38].

Figure 1 depicts the proposed Neuro-fuzzy CBR architectural blueprint for detecting suspected cases of Lassa fever and classification of the suspected cases detected into various degrees of severity.

The Neuro-fuzzy CBR Algorithm developed by [39] was adopted for explaining the system architectural blueprint presented in Figure 1, because it provides a self-learning and adaptive system that is capable of handling uncertainty and imprecise data as if it was a human expert.

Step 1: Input observed personal symptoms as a new case

Step 2: Submit the input symptoms for diagnoses

Step 3: Activate the Euclidean distance as a Nearest Neighbor Algorithm

Step 4: Compare the new case with all the sets of past confirmed positive and negative Cases stored in the Case Based Library for a measure of similarity match using the Euclidean distance model

$$\sqrt{[(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2 + \cdots + (z_1 - z_2)^2]}$$

Step 5: Which past case does the new case resemble the most?

Step 6: IF a measure of similarity match is found Continue processing Else Goto **Step 9**

Step 7: Re-use that solution of the most similar past Case

Step 8: Adapt the past Case solution to get a solution for the new Case using the Neural Network Model

Step 9: Activate the Neural Network (NN) Model

Step 10: Compute for $X = \sum_{i=1}^{n} x_i * w_j$

Step 11: Compute for

$$Y = \begin{cases} Positive\ case\ of\ Lassa\ fever\ detected, & if\ x > \theta \\ Negative\ case\ of\ Lassa\ fever\ detected, & if\ x < \theta \end{cases}$$

Step 12: Activate the Fuzzy Logic (FL) Model to classify the final output from the neural network into its degree of severity based on the initial twenty-nine clinical symptoms observed.

Step 13: Compute for degree of severity

$$\mathbf{Y} = \sum_{i=1}^{n} L_3(xi)$$

Step 14: generated classified output:

$$Output = \begin{cases} Mild & 0.6 \leq Y < 0.7 \\ Moderate & 0.7 \leq Y < 0.8 \\ Severe & 0.8 \leq Y \leq 1.0 \end{cases}$$

Step 15: Retain the new Case and its solution into the Case Based Library for reference purpose in solving future cases.

Step 16: Update the Case Based Library.

The thirty-four (34) input parameters fed into the system in two diagnostic stages by a given user (patient) in Figure 1 are obtained from users experiences in form of signs and also their laboratory results if known and available for usage.

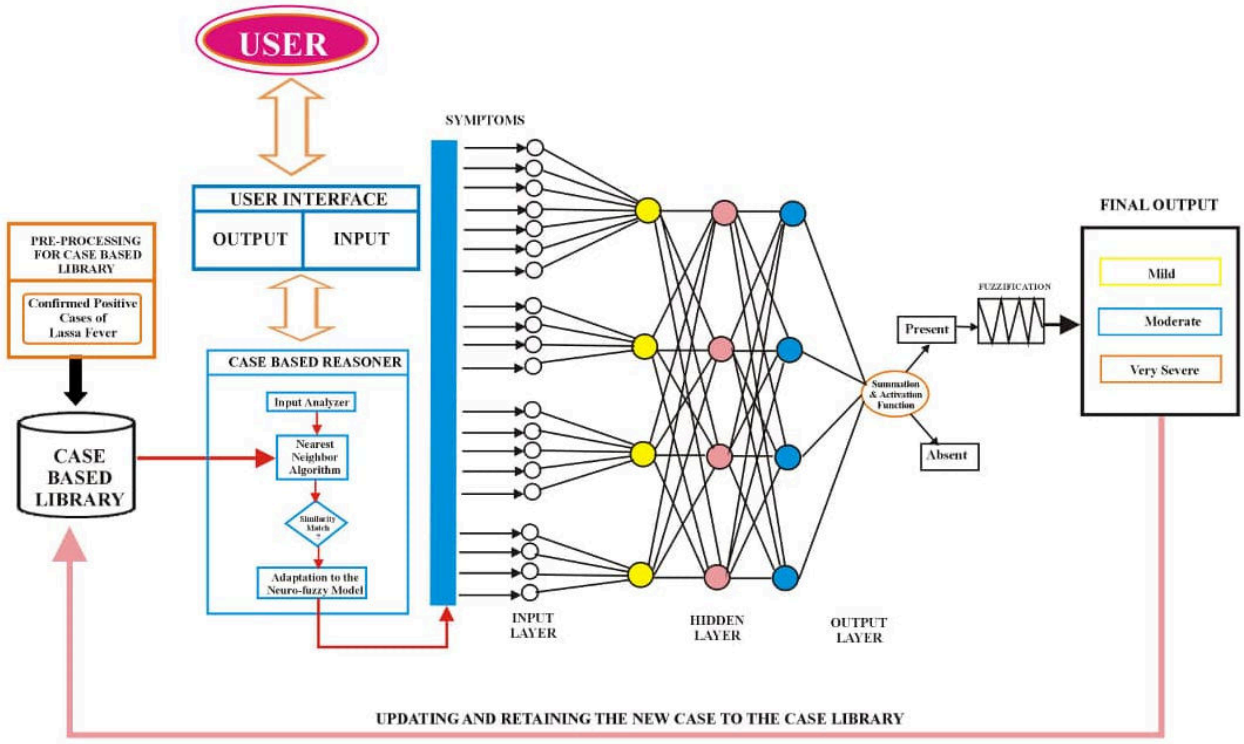


Figure 1: Neuro-Fuzzy CBR Architecture **Adopted from [39]**

Rule 1: IF ($60 \leq Y < 70$)
THEN SUSPECTED Lassa fever CASE is MILD
Rule 2: IF ($70 \leq Y < 80$)
THEN SUSPECTED Lassa fever CASE is MODERATE
Rule 3: IF ($80 \leq Y \leq 100$)
THEN SUSPECTED Lassa fever CASE is SEVERE

Figure 2: Derived Fuzzy rule for classification [39]

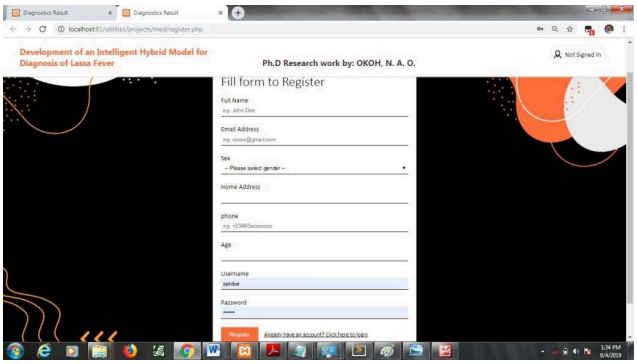


Figure 3: Signup interface for registering new user

IV. RESULTS AND DISCUSSION

This section discusses the results obtained from the research work carried out. Figures 3, 4, 5, 6, 7, 8, 9, 10 and 11 depict screenshots of the different interface that can be used to perform certain actions as a result of the developed intelligent hybrid expert system for diagnosis of Lassa fever.

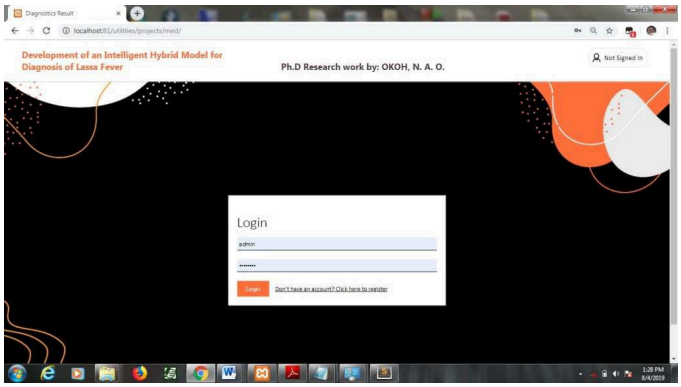


Figure 4: Login interface

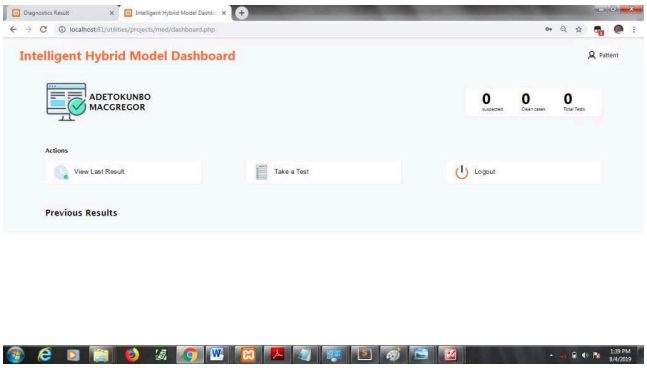


Figure 5: User dashboard

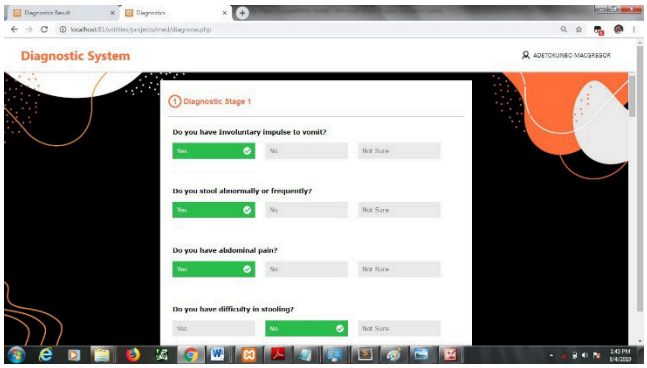


Figure 6a: Interface for capturing patient's sign/symptoms

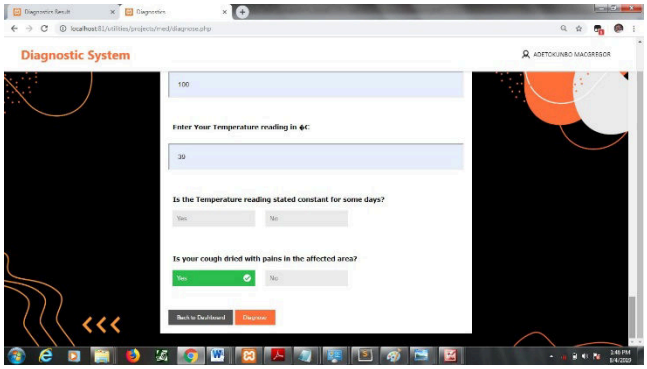


Figure 6b: Interface for capturing patient's sign/symptoms

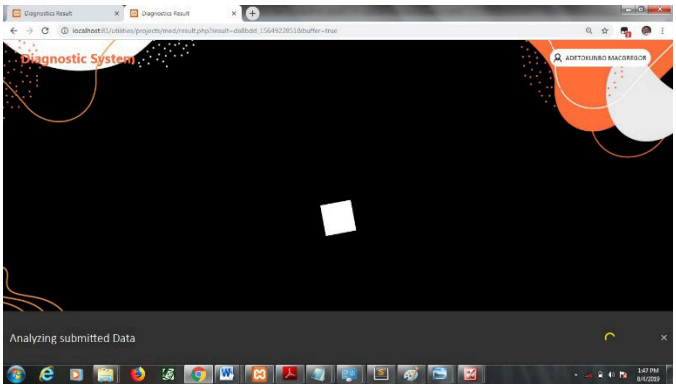


Figure 7: Interface for processing diagnosis

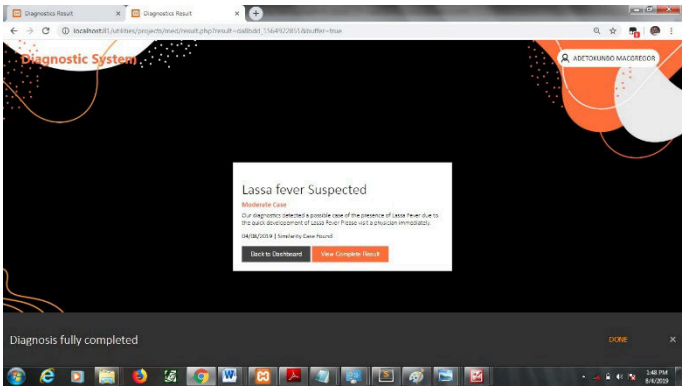


Figure 8: Interface for displaying diagnostic report

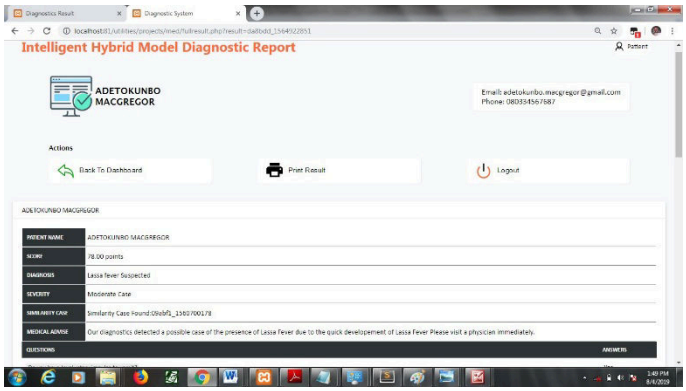


Figure 9: Interface for viewing detailed diagnostic report

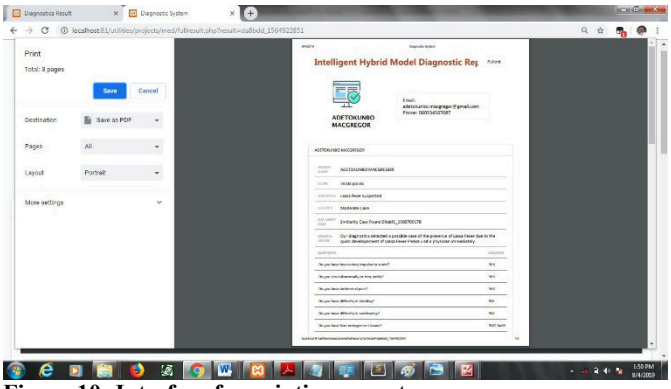


Figure 10: Interface for printing report

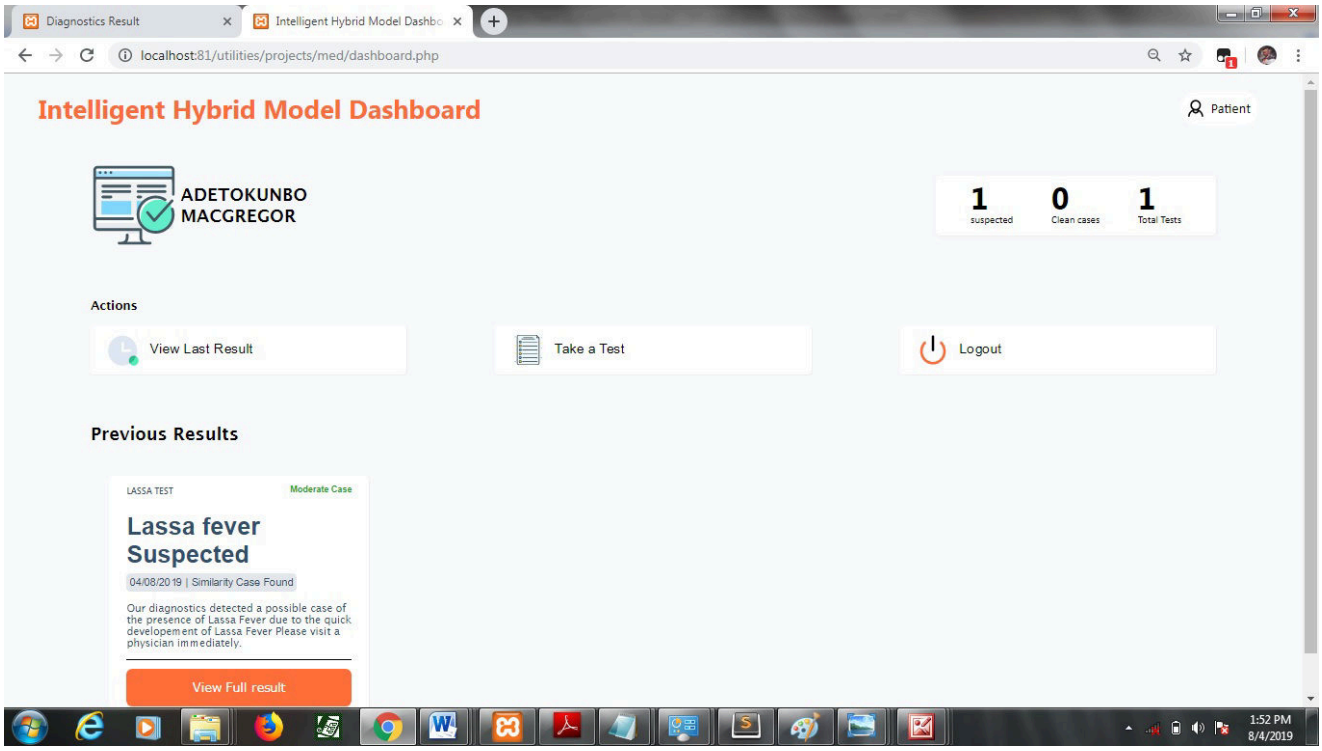


Figure 11: Action dashboard

4.1 Evaluation of expert system techniques

This section shows the evaluation of the number of techniques used in implementing our expert system model as compared to four different research scholars techniques used for analyzing and diagnosing Lassa fever.

Table 5 reveals the different types of techniques used in developing the expert system.

TABLE 5: Types of techniques used

S/N	Techniques	No
1.	Rule-based [22]	1
2.	Rule-based [18]	1
3.	Fuzzy-based [23]	1
4.	Neuro-fuzzy [49]	2
5.	Adopted & Implemented Neuro-fuzzy CBR Model [39]	3

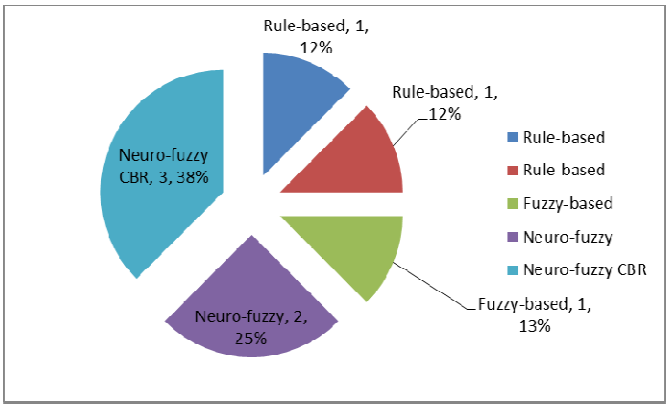


Figure 12: Graphical representation of Table 5

Figure 12 depicts the graphical representation of the techniques used in Table 5 in terms of percentage rating. Tables 6 shows a sample dataset of patients and laboratory results confirming them as positive cases against our implemented intelligent hybrid expert system results for diagnosing Lassa fever. Table 7 shows a comparative summary of the sample dataset shown in Table 6.

TABLE 6: Sample dataset and diagnostic results

	SIGNS AND CLINICAL SYMPTOMS																																	RESULTS	
CASE FILE NO.	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	INTELLIGENT HYBRID EXPERT SYSTEM RESULT	EXPERT OPINION BASED ON LAB RESULTS
CF001	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	99	32	Yes	No	LF SUSPECTED	SUSPECTED
CF002	Yes	Yes	Yes	No	No	Not Sure	No	Not Sure	No	Not Sure	Not Sure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not Sure	Not Sure	Yes	Yes	Not Sure	No	Yes	Not Sure	Not Sure	Yes	Yes	Yes	100	38	Yes	Yes	LF SUSPECTED	SUSPECTED
CF003	Yes	No	Yes	Yes	No	Yes	No	No	No	Not Sure	Not Sure	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Not Sure	Yes	Yes	No	No	Yes	Not Sure	Not Sure	Yes	Yes	Yes	99	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF004	Yes	No	Yes	Not Sure	No	Yes	No	Yes	Not Sure	No	Yes	No	Not Sure	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Not Sure	No	Not Sure	Yes	No	Yes	No	120	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF005	Yes	No	Yes	No	Yes	No	No	No	No	Not Sure	Not Sure	Not Sure	Not Sure	Not Sure	No	No	No	No	No	Not Sure	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	120	40	No	Yes	LF ABSSENT	SUSPECTED
CF006	Yes	No	Yes	No	No	Not Sure	No	Not Sure	No	Not Sure	No	Yes	Yes	Not Sure	Yes	Yes	Yes	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	Yes	Yes	Yes	100	39	Yes	Yes	LF SUSPECTED	SUSPECTED
CF007	Yes	No	Yes	Not Sure	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not Sure	Not Sure	Yes	Yes	Yes	Yes	Yes	Yes	100	39	Yes	Yes	LF SUSPECTED	SUSPECTED
CF008	No	Yes	Yes	No	No	Not Sure	Not Sure	Not Sure	No	Not Sure	Not Sure	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Not Sure	No	Yes	Not Sure	Yes	Yes	Yes	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF009	No	No	Yes	No	Yes	No	No	Not Sure	Not Sure	Not Sure	Yes	Yes	Yes	No	Yes	No	No	Not Sure	Not Sure	Yes	Yes	Yes	Not Sure	Not Sure	No	Not Sure	Yes	Yes	No	Yes	100	39	Yes	Yes	LF SUSPECTED	SUSPECTED
CF010	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not Sure	Not Sure	Not Sure	Yes	Not Sure	Yes	Yes	Not Sure	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF011	No	No	Yes	Yes	Not Sure	Not Sure	No	Not Sure	No	Not Sure	Not Sure	Yes	Yes	Yes	Yes	Not Sure	No	Yes	Not Sure	Not Sure	Yes	Yes	Not Sure	No	Yes	Not Sure	Yes	Yes	Yes	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF012	No	No	Yes	No	Yes	Not Sure	No	Not Sure	No	Not Sure	Yes	Yes	Yes	Yes	Yes	Not Sure	Yes	Not Sure	No	Yes	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF013	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Not Sure	No	Yes	Yes	Yes	Yes	Yes	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF014	Yes	No	Yes	Yes	No	No	Not Sure	Not Sure	Not Sure	Not Sure	Yes	No	Yes	No	Yes	Yes	Yes	No	No	Yes	No	Yes	Not Sure	No	No	Not Sure	Not Sure	Yes	No	Yes	100	39	Yes	Yes	LF SUSPECTED	SUSPECTED
CF015	Yes	Yes	Yes	No	No	Not Sure	Not Sure	Not Sure	No	Yes	Yes	Yes	Yes	Not Sure	No	Not Sure	No	Yes	No	Yes	Yes	Yes	Not Sure	No	Yes	Not Sure	Yes	Yes	Not Sure	Yes	100	39	Yes	Yes	LF SUSPECTED	SUSPECTED
CF016	No	No	Yes	No	Yes	Not Sure	No	Not Sure	No	Not Sure	Yes	Yes	Yes	Yes	Yes	Not Sure	Yes	Not Sure	No	Yes	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	Yes	Not Sure	Yes	100	40	Yes	Yes	LF SUSPECTED	SUSPECTED
CF017	Yes	No	Yes	No	No	Not Sure	Not Sure	Not Sure	No	Not Sure	Yes	No	No	No	Yes	Not Sure	No	No	Not Sure	Not Sure	Yes	Yes	Yes	Not Sure	No	Yes	Yes	Yes	No	Yes	120	38.9	Yes	No	LF SUSPECTED	SUSPECTED

TABLE 7: A Comparative summary of sample dataset

RESULTS	
INTELLIGENT HYBRID EXPERT SYSTEM RESULT (SUSPECTED CASE)	EXPERT OPINION BASED ON LAB RESULTS (SUSPECTED CASE)
16	17

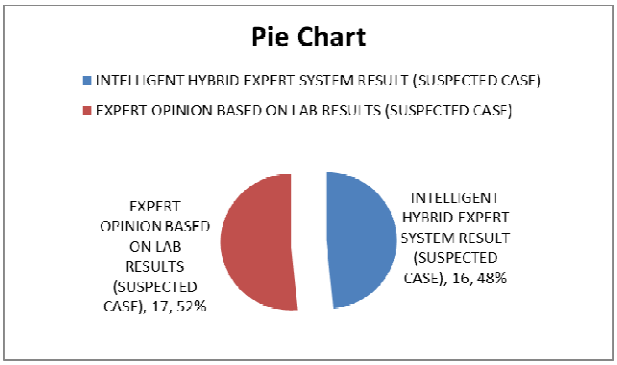


Figure 13: Graphical representation of Table 6

4.2 Calculation of classification accuracy

The classification of accuracy rate of our implemented intelligent hybrid expert system in analyzing and diagnosing the Lassa fever correctly from the given dataset as compared to the laboratory results is computed as follows:

Classification Accuracy Rate (CAR) is defined as

$$(CAR) = \frac{\text{Number of Training Instances Correctly Classified}}{\text{Number of Training Instances}} x\,100$$

$$(CAR) = \frac{16}{17} x\,100 = 94.12\%$$

Classification Error Rate (CER) is defined as:

$$(CER) = \frac{\text{Number of Training Instances Incorrectly Classified}}{\text{Number of Training Instances}} x\,100$$

$$(CER) = \frac{1}{17} x\,100 = 5.88\%$$

This accuracy evaluation shows that our implemented model is very close to the human expert opinion based on laboratory results.

V. CONCLUSION

The developed neuro-fuzzy CBR expert system is able to successfully and efficiently detect the presence or non-presence of Lassa fever cases from the given sample dataset captured. The fuzzy inference engine can also effectively classify the outcome from the neural network into mild, moderate and severe cases with no false positive alarm rate.

This application developed is a web application considering the audience for wide usage. In future, researchers should consider developing and deploying the product as a mobile application (Mobileapp) that can conveniently be executed from any smartphone with little memory and processor speed. This will enable large awareness, acceptability and usage of the application; and further reduce mortality rate.

ACKNOWLEDGMENT

We wish to express our gratitude to the Dr. Vincent Aiwuyo (Medical Doctor) and Mr. Felix Aitufe (Principal Medical Lab Scientist) at Ambrose Alli University Health Centre, Ekpoma, Edo State for sharing their time and knowledge on some issues pertaining Lassa fever.

The authors are also very grateful and indebted to Dr. Ephraim Ogbawi-Emovon (Director) and some key members of his team, namely; Dr. Donatus E. Adomeh (Ph.D., DDMLS), Mr. Aire Chris, FMLSCN, and Mr. Ikponmwosa Odia, ADMLS (Lab Manager) at, Irrua Specialist Teaching Hospital (ISTH), Institute of Lassa Fever Research and Control (ILFRC), Edo State for also sharing their very precious time, wonderful insight on Lassa fever trend of infection, varieties of signs and symptoms, control, degree of severity classification and possible treatment plan for suspected cases, and finally some of their published statistical data on Lassa fever trend.

REFERENCES

[1] Bowen, M. D., Rollin, P. E., Ksiazek, T. G., Hustad, H. L., Bausch, D. G., and Demby, A. H. (2000), Genetic diversity among Lassa virus strains. Journal of Virol, 74(15): 6992 - 7004.

[2] Ogbu, O., Ajuluchukwu, E and Uneke, C. J. (2007), Lassa fever in West Africa sub-region: An Overview, Journal of Vect Borne Dis. 44(1):1 - 11.

[3] Eze, K. C, Salami, T. A. T, Eze, I. C., Pogoson, A. E., Omordia, N., and Ugochukwu, M. O. (2010), High Lassa fever activity in Northern part of Edo state, Nigeria: Re-analysis of confirmatory test results, Afr Journal of Health Science, 17:52 - 56.

[4] Centre for Disease Control and Prevention. Lassa fever fact sheet. 2014. Available:http://www.cdc.gov/ncidod/dvrd/spb/mnpages/dispages/factsheets/lassa_fever_fact_sheet.pdf. Accessed January 2016. Google Scholar

[5] World Health Organisation. World Health Organisation fact sheet on Lassa fever. 2015; 10(11). Google Scholar

[6] Nasir, A. I., and Sani, M. F. (2015), Outbreak, pathogen containment and laboratory investigation of Lassa fever in Nigeria: How prepared are we?. International Journal of Tropical Disease and Health, 10(1):1 -10.

[7] Achinge, I. G., Kur, T. J., and Gyoh, K. S. (2013), Lassa fever outbreak in Makurdi, North Central Nigeria: what you need to know. IOSR Journal of Dental and Medical Sciences, 7(5):42 - 46.

[8] Sanjay, K. and Rajikishore, P (2015), Importance of Expert System Shell in Development of Expert System, International Journal of Innovative Research and Development, 4(3): 128 – 133.

[9] Vasighi, M. (2016), Artificial Neural Networks, Part 1: Biological Inspiration. Institute for Advanced Studies in Basic Science, DOI – 10.13140/RG.2.1.3083.6089.

[10] Wikipedia (2010), "Artificial Neural Network" retrieved from http//: en.Wikipedia.org/wiki/Artificial-neural-network.

[11] Aleksander I. and Morton H. (1998), "An introduction to neural computing" 2nd Edition Computer Science press.
[12] Chokshi, D. A., and Farley, T. A. (2012). The cost-effectiveness of environmental approaches to disease prevention, New England Journal of Medicine, 367(4):295-7. doi: 10.1056/NEJMp1206268
[13] Negnevitshy, M. (2005). Artificial Intelligence: A Guide to Intelligent Systems, Second Edition, Addison Wesley, p373.
[14] Vieira, J., Morgador, D. F., and Mota A. (2004). Neuro-Fuzzy Systems: A Survey, Vol. 3(2): 414 - 419.
[15] Somdatta, P., and Gour-Sundar-Mitra, T. (2013), A Proposed Neuro-Fuzzy Model for Adult Asthma Disease Diagnosis, Journal of Computer Science & Information Technology, 3: 191 - 205.
[16] Fatumo, S. A., Adetiba, E., and Onaolapo, J. O. (2013). Implementation of XpertMalTyph: An Expert System for Medical Diagnosis of the Complications of Malaria and Typhoid, IOSR Journal of Computer Engineering, 8(5): 34 – 40.
[17] Obi J. C., and Imianvan, A. A. (2011). Interactive Neuro-Fuzzy Expert System for Diagnosis of Leukemia, Global Journal of Computer Science and Technology, 11(12): 42 – 50.
[18] Tunmibi, S., Adeniji, O., Aregbesola, A., and Ayodeji, D. (2013). A Rule Based Expert System for Diagnosis of Fever, International Journal of Advanced Research, 1(7): 343 – 348.
[19] Patel, M., Patel, A., and Virparia, P. (2013), Rule Based Expert System for Viral Infection Diagnosis, International Journal of Advanced Research in Computer Science and Software Engineering, 3(5).
[20] Hossain, M. S., Khalid, M. S., Akter, S., and Dey, S. (2014). A belief rule-based expert system to diagnose influenza, In proceedings of Strategic Technology 9th International Forum 2014 IEEE
[21] Komal, R. H., and Vijay, S. G. (2014), Rule-Based Expert System for the Diagnosis of Memory Loss Diseases, International Journal of Innovative Science, Engineering & Technology, 1(3).
[22] Hambali, M. A., Akinyemi, A. A., and Luka, J. D. (2017). Expert System For Lassa Fever Diagnosis Using Rule Based Approach, Annals Computer Science Series, 15(2): 68 – 74.
[23] Osaseri, R. O, Onibere, E. A., and Usiobiafo, A. R. (2014). Fuzzy Expert Model for Diagnosis of Lassa fever, Journal of the Nigerian Association of Mathematical Physics, 27: 533 – 540.
[24] Imianvan, A. A., and Obi, J. C. (2012). Cognitive Neuro-Fuzzy System for Hypotension Control, Computer Engineering and Intelligent Systems, 3(6): 21- 31.
[25] Ephzibah, E. P., and Sundarapandian, V. (2012). A Neuro Fuzzy Expert System for Heart Disease Diagnosis, Computer Science & Engineering: An International Journal, 2(1): 17 – 23.
[26] Tom, J. J. and Anebo, N. P. (2018), A Neuro-Fuzzy Based Model for Diagnosis of Monkeypox Diseases, International Journal of Computer Science Trends and Technology, 6(2): 143 – 153.
[27] Shaabani, M. E., Banirostam, T and Hedayati, A. (2016), Implementation of Neuro Fuzzy System for Diagnosis of Multiple Sclerosis, International Journal of Computer Science and Network, 5(1): 157 – 164.
[28] Omotosho, A., Oluwatobi, A. E., Oluwaseun, O. R. (2018). A Neuro-Fuzzy System for the Classification of Cells as Cancerous or Non-Cancerous, International Journal of Medical Research and Health Sciences, 7(5): 155 – 166.
[29] Egwali, A. O and Obi, J. C. (2015), An Adaptive Neuro-Fuzzy Inference System for Diagnosis of EHF, The Pacific Journal of Science and Technology, 16(1): 251 – 261.
[30] Oladele, T. M., Okonji, C. D., Adekanmi, A., and Abiola, F .F. (2018). Neuro-Fuzzy Expert System for Diagnosis of Thyroid Diseases, Annale Computer Science Series, 16(2): 45 – 54.
[31] Goni, I., Ngene, C. U., Manga, I., and Auwal, N., and Sunday, J. C. (2018), Intelligent System for Diagnosing Tuberculosis Using Adaptive Neuro-Fuzzy, Asian Journal of Research in Computer Science, 2(1): 1 – 9.
[32] Gumpy, J. M., Goni, I., and Isa, M. (2018), Neuro-Fuzzy Approach for Diagnosing and Control of Tuberculosis. The International Journal of Computational Science, Information Technology and Control Engineering, 5(1): 1 – 10.
[33] Maskara, S., Kushwaha, A., and Bhardwaj, S. (2018), Adaptive Neuro Fuzzy Expert System for Disease Diagnosis, International Journal of Innovations in Engineering and Technology, 10(2): 121 – 123.
[34] Nagarajasri, M., and Padmavathamma, M. (2013), Threshold Neuro Fuzzy Expert System for Diagnosis of Breast Cancer, International Journal of Computer Applications, 66(8): 6 – 10.
[35] Manikandan, T., Bharathi, N., Sathish, M., and Asokan, V. (2017), Hybrid Neuro-Fuzzy System for Prediction of Lung Disease Based on the Observed Symptom Values, Journal of Chemical and Pharmaceutical Sciences, 8: 69 -76.
[36] Jaakkola, H., and Thalheim, B. (2011). "Architecture-driven modelling methodologies." In: Proceedings of the 2011 conference on Information Modelling and Knowledge Bases XXII.
[37] Clements, P. C. (1996). "A survey of architecture description languages." Proceedings of the 8th International workshop on software specification and design. IEEE Computer Society, 1996.
[38] Medvidovic, N., and Taylor, R. N. (2000). "A classification and comparison framework for software architecture description languages." Software Engineering, IEEE Transactions on 26.1 (2000): 70-93.
[39] Nnebe, S. E., Okoh, N. O. A., John-Otumu, A. M., and Oshoiribhor, E. O. (2019), A Neuro-Fuzzy Case Based Reasoning Framework for Detecting Lassa fever Based on Observed Symptoms. American Journal of Artificial Intelligence, 3(1): 9 - 16. doi: 10.11648/j.ajai.20190301.12
[40] Adewuyi, G. M., Fowotade, A. and Adewuyi, B. T. (2009). Lassa fever: Another Infectious Menace, AFRICAN Journal of Clinical and Experimental Microbiology, 10(3): 144-155. http://www.ajol.info/journals/ajcem
[41] Olayiwola, J. O., and Bakarey, A. S. (2017). Epidemiological Trends of Lassa fever Outbreaks and Insights for Future Control in Nigeria, International Journal of Tropical Disease & Health, 24(4): 1-14.
[42] Omeh, D. J., Achinge, G. I., and Echekwube, P. O. (2017). Lassa fever in West Africa: A Clinical and Epidemiological Review, Journal of Advances in Medicine and Medical Research, 24(6): 1-12.
[43] Dongo, A. E., Kesieme, E. B., Iyamu, C. E., Okokhere, P. O., Akhuemokhan, O. C., and Akpede, G. O. (2013). Lassa fever presenting as acute abdomen: a case series, Virology Journal, 10(123): 1 – 7.
[44] Richmond, J. K., and Baglole, D. J. (2003). Lassa fever: epidemiology, clinical features, and social consequencies. BMJ, 327:1271–1275.
[45] CDC Lassa fever fact sheet. CDC infectious disease information. Viral haemorrhagic fever. www.cdc.gov/ncidod/dvrd/spb/mnpage/dispages/ Lassaf.htm.
[46] Bausch, D. G. (2008). Viral haemorrhagic fevers. In Clinical Infectious disease. Edited by Schlossberg D. New York, NY: Cambridge University Press;1319–1332.
[47] Asogun, D. A., Adomeh, D. I,, Ehimuan, J., and Ikponmwonsa O. (2012). Molecular diagnostics of Lassa fever in Irrua Specialist Teaching Hospital, Nigeria: Lessons learnt from two years of laboratory operations. PLoS Trop Dis, 6(9):el1839. doi:10.1371/journal.pntd.0001839.
[48] McCormick J. B, King, I. J., Webb, P. A., Johnson, K. M., O'sullivan, R., Smith, E. S., Trippel, S,, and Tong, T. C. (1987): A case-control study of the clinical diagnosis and course of Lassa fever. J Infect Dis 1987, 155(3):445–455.
[49] Roseline O. O. and Osaseri E. I. (2016). *Soft Computing Approach for Diagnosis of Lassa fever*, World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering, 10(11)

Hybrid Approach for Selection of Academic Major Using Naïve Bayes

[1,4]Farouk Lawan Gambo[*], [2]Adamu Sani Yahaya, [3]Abdullah Abubakar [3]Khalid Haruna, ,[1]Aminu A. Aliyu and [2]Suhail M. Kamal.
[1]Department of Computer Science
Federal University Dutse, FUD
Jigawa, Nigeria[2]Department of Information technology
Bayero University Kano, BUK
Kano, Nigeria
[3]Department of Computer Science
Bayero University Kano, BUK
Kano, Nigeria
[4]Modibo Adama University of Technology, Yola
Adamawa, Nigeria

[*]farouk.gambo@fud.edu.ng and farouk4142@gmail.com

ABSTRACT

There is a need for proper selection of academic major for students' going into senior secondary schools; this right placement of students is important toward academic achievements. In the Nigerian educational system, students are subjected to qualifying exams into any of the secondary school classes. Though this approach is valid and reliable but is associated with some challenges as it does not capture student academic record. Predictive models have been proved to be essential toward achieving remarkable improvements in both productivity and proficiency in almost all human endeavors. In this paper, a conceptual framework is proposed for Prediction of student's academic major from their existing records using a hybrid Naïve Bayes classifiers.

Keywords:*Classification; Naïve Bayes; Predictive Model; Major Selection*

I. Introduction

Choosing the right academic major for junior secondary student into senior secondary school will assist both student and their teachers toward achieving an academic goal. There are certain guidelines or standard set by the national policy on education to be followed for placing students into the right academic major while entering senior secondary schools. Traditionally, students seeking admission into Senior Classes must have passed stipulated examination, i.e., BECE exams. They must pass the exam satisfactorily with emphasis on any of Science, Art, and Commercial related subjects. Though this method is valid and reliable, it is associated with some challenges as it does not carry along with the student's academic history in the decision. Another issue with this approach is that some parent influences the process by choosing that they want their son/daughter to do [1,2].

On the other hand, the use of predictive models has brought about remarkable improvements in productivity, and proficiency in almost all human endeavours, which is one way to go about the issues mentioned above. This can be done by developing a model that automatically takes JSS 1 - JSS 3 academic history into account to predict and recommend the right major. Predictive models have a precise goal of allowing us to predict the unknown values of attributes of a target given known values of other attributes. These predictions can be described as the process of studying the present and past states of an attribute in order to predict its future state by [3] while predictive modeling can be understood as a learning function that maps input set of vector to a scalar output [4]. There are different approaches toward building a predictive model for recommendation; these approaches are content-based filtering, collaborative filtering, and hybrid approaches.

This paper suggests that instead of the traditional way of selecting academic major, schools should not only restrict the admission process on the exam but to find a way of admitting student beyond the examination by looking into the student performance from JSS 1 to JSS 3. To this end, this paper adopts the concept on classification through developing a model using Naïve Bayesian to predict and make a recommendation of the right major of a student from their academic history right from JSS 1 to JSS 3.

II. REVIEW LITERATURE

A. Concept of Classification

Classification is one of the techniques used in machine learning for prediction. Classification provides predictive

modeling approach which allows prediction of the unknown from the known results found from different data [3]. There are several classifiers or classification model that can be used for prediction; Artificial Neural Network, Naïve Bayes, Decision Trees, and K-Nearest Neighbour. However, in this work, we adopt Naïve Bayes because it works as both supervised learning and statistical-based technique for classification [5]. It works based on Bayes' theorem through finding the probability of an event occurring given the probability of another event that has already occurred. It assumes a model which relies on the probability to calculate the uncertainty of future events in such a principled mechanized way through estimating the probabilities of the events. Such a mechanism has been widely used in prediction. Naïve Bayes classification is simple and particularly suited when the dimensionality of the input is high. Despite its simplicity, it can outperform more sophisticated classification. It provides perspective for understanding many learner algorithms and works on the assumptions that: is easy to construct, classifying categorical data, occurrences of an event (attributes) are independent and can be trained in a supervised manner [6]. The major advantage of Naïve Bayes in classification is its simplicity and its ability to approximate probabilities for a class on any given instance [7].

B. Approaches for Selection and Recommendation

i. Collaborative Filtering Approaches

. Collaborative filtering approaches build a model from a user's past behaviour (items previously purchased or selected and/or numerical ratings given to those items) [8].This model is then used to predict items (or ratings for items) that the user may have an interest in.

ii. Content-Based Filtering approaches

Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties [9]. In this study, we decide to use content-based filtering approaches by applying Bayesian Classifier to solve the issue of recommending major in our schools. This is because Bayesian Classifier is one of the most successful machine learning algorithms in many classification domains.

iii. Hybrid Approaches.

Usually, in the hybrid approach, content-based is often combined with collaborative filtering to build a hybrid recommendation algorithm [8]. The hybrid method integrates two or more scores produced by the two independent recommendation algorithms

C. **Naïve Bayes Theorem**

Let A be the training data set and X be n-dimensional attribute vector represented by $X = (X_1, X_2, \dots, X_n)$, which corresponds to the data set $A_1, A_2, \dots, A_n$ respectively. Suppose that there are n- classes: $C_1, C_2, \dots, C_n)$, with given a tuple X, the classifier will predict where X belongs to by considering the class having the highest probability (P) condition on X.

For example

X Belongs to class C_k if and only if

$$P(Ck|Y) > P(Ci|Y)$$
$$for\ 1 \leq i \leq n, i \neq k$$

D. Predictive and Recommendation Systems in Education**.**

The problem associated with result computing in the education sector is something that actively attracts the attention of many researchers. Lots of research efforts have been devoted to studying various emerging technologies in solving the problem in Recommendation Systems [10-13]. The Intelligent Online Academic Management System (IOAMS) is an intelligent web-based system, which offers a useful tool for the education sector. It is designed to provide academic counselling and monitor student progress [14]. Another research by [15] identify the problem of student academic result processing system and further design and implement a client-server distributed database system to eliminate the deficiencies inherent such system. [9] Also, having identified the problem of student result processing system, which is improper subject registration, late release of students' results, inaccuracy due to manual and tedious calculation and retrieval difficulties/inefficiency. He went ahead to developed automated student result management system, which is capable of computing of raw score and the storing student result as required. According to [16] problem in student result processing was solved by designing and implementing a software application, which meant to ease the processing of students' results in secondary schools. Among the studies reviewed [1,2] both focuses on the choice of science, Art and commercial among secondary school students and challenges encountered by the students when using computer-based test platform respectively. Based on the previous literature, we discovered that there is a need for the Recommendation system for the selection of major in the school. Therefore this paper proposes a system which will help not only on keeping student record, viewing student academic history but also generating recommendations for major.

II. Proposed Approach

In this paper, we proposed a technique toward building a model that would classify and predict the major of a student. The model classifies whether a student is to be in Science class, Art class or Commerce class. The proposed methodology of this research divides into the following steps which conform to general classification techniques; general data collection, Preprocessing, Classification, and evaluation of the model to infer knowledge. The figure below shows all the steps of the proposed technique.

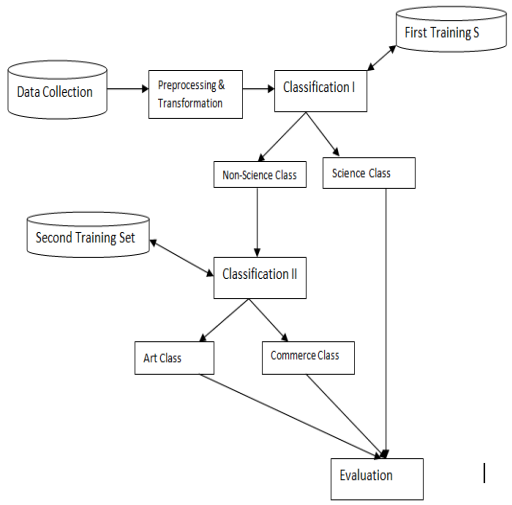


Fig. 6. The Hybrid Naïve Bayes Prediction Framework.

E. Procedure for the Selection of Academic Major

III. i. Data Collection

To solve the problems of deciding the right major of graduating junior secondary school students, this is the stage where results of students were collected from the manual archive of Delight Secondary school consisting of the scores by subjects of each student and stored in an academic history database. The data is then divided into two tables: the first table which contains predetermined science and Non-science subjects as data used for classification I, while the second table which contains predetermined Art and commerce subjects as data used for classification II.

IV. ii. Data Preprocessing and Transformation

Data preprocessing was done to remove noise and outliers that is students that are absent in some subject and those that scored very high or very low. In this stage, data must be transformed from analog to electronic record that is '1' is assigned to where student score above the average mark and '0' for a student that scores below the average mark in each subject. Having done with data collection and preprocessing, pre-processed data is stored in a first training set and second training set 2. These constructed datasets are to be used during classification process.

iii. Classification and Prediction

Classification is a technique used to assign a class of unseen records as correctly as possible by using a collection of records called a training dataset, where each tuple in the training set comprises a set of attributes, and then one of the attributes is called a class. While prediction is obtained by extracting rules and pattern in training data set and to use those extracted rule to predict the class of records whose class label is unidentified. And the objective of this step is to provide a classification model for the class elements, then devise a validation mechanism using test data set in order to determine the accuracy of the model (Kuar & Wasan, 2006)

In this phase, the dataset is divided into training and testing sets where one part is used to build the predictive model through training the system while few data extracted to compare with the result generated by the classifier (testing). Looking at the nature of attributes in our dataset, the conceptual work uses the classification techniques twice to build the two classifiers (Naïve Bayes) through; *Classification I* and *Classification II*. The first classifier classifies students into science or non-science major while the second classifier classifies into art or commerce class.

V. **Example** .

Suppose there is a new student with a set of attributes say X that we want to find his/her class label.

X = (*Mathematics=1, English=1, Basic Science=1, Basic technology=1, Religious Studies =0, Language Studies=1Social Studies= 0, Business Studies=0).*

If P(X | Science) and P(X | Non-Science_Class) are calculated using the two equations below.

$$P(\text{Science}/\text{Student}) = \frac{P(\text{Student}|S) * P(S)}{P(\text{Student}|S) * P(S) + P(\text{Student}|N) * P(N)}$$

$$P(\text{Non}-\text{Science\_Class}/\text{Student}) = \frac{P(\text{Student}|N) * P(N)}{P(\text{Student}|N) * P(N) + P(\text{Student}|S) * P(S)}$$

Thus, X to the classified to the class having the highest probability. Moreover, the same applied to P(X | Art_Class) and P(X | Commerce_Class)

F. **Model Evaluation.**

As we mentioned, we adopted a classification technique that builds a model that would classify and predict the academic major of a student. The model classifies weather a student is to be in Science class, Art class or Commerce class. The proposed methodology of this research divides into the following steps which conform to general classification techniques; general data collection, Pre-processing, Classification, and finally model evaluation.

This part explains how we hope to evaluate the model using most used evaluation measures (Precision, Recall and F1) which would require a partitioning of the dataset into a training data, a test data and a validation data. To ensure that these subsets are complimentary, multiple cross-validations would be performed to ensure that the variation between the different subsets is kept at a minimum.

By doing this, it will enable us to make comparisons with results published in similar and previous researches.

VI. Conclusion

In this paper, we proposed a framework that will extract paper from student academic history, transform the data and apply classification technique twice using naive Bayes classifiers to classify students into science Art

and Commerce class. If fully implemented, the approach will provide a better placement of student into the right academic major in senior secondary school and in the future we hope to develop an application that will implement the proposed conceptual framework.

References

[1] Christiana I Oriahi, PO Uhumuavbi, and LI Aguele. 2010. Choice of science and technology subjects among secondary school students. Journal of Social Sciences 22, 3 (2010), 191–198.

[2] N Patricia and NC Norbert. 2015. Challenges Encounter Using Cbt by 2015 Utme Candidate In Owerri Zone One Nigeria: Test Validity Implications. IOSR Journal of Research and Method in Education (IOSR-JRME) 5, 5 (2015), 28–35.

[3] Rokach, L. & O. Maimon (2010). Data mining and Knowledge Discovery Handbook. Second ed. 2010. *Springer Science and Bussiness*.

[4] Jothi, N., Rashid, N. A., & Husain, W. (2015). Data mining in healthcare - A Review. *Procedia Computer Science*, 72, 306–313

[5] Al-Radaideh, Q.A., & Al Nagi, E. (2012). Using data mining techniques to build a classification model for predicting employees performance. *International Journal of Advanced Computer Science and Applications,* 3(2), 144-151.

[6] Patil, R. R. (2014). Heart disease prediction system using Naive Bayes and Jelinek-mercer smoothing. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(5), 2278-1021.

[7] Kononenko, I. (1991). Semi-naive Bayesian classifier. In *Machine Learning—EWSL-91* (pp. 206-219). Springer Berlin/Heidelberg.

[8] Hosein Jafarkarimi, Alex Tze Hiang Sim, and Robab Saadatdoost. 2012. A naive recommendation model for large databases. International Journal of Information and Education Technology 2, 3 (2012), 216.

[9] Prem Melville and Vikas Sindhwani. 2011. Recommender systems. In Encyclopedia of machine learning. Springer, 829–838.

[10] Mohamed Hamada and Mohammed Hassan. 2018. Artificial Neural Networks and Particle Swarm Optimization Algorithms for Preference Prediction in Multi-Criteria Recommender Systems. In Informatics, Vol. 5. Multidisciplinary Digital

[11] Mohammed Hassan and Mohamed Hamada. 2017. A Computational Model for Improving the Accuracy of Multi-criteria Recommender Systems. In Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2017 IEEE 11th International Symposium on. IEEE, 114–119.

[12] Mohammed Hassan and Mohamed Hamada. 2017. Improving prediction accuracy of multi-criteria recommender systems using adaptive genetic algorithms. In Intelligent Systems Conference (IntelliSys), 2017. IEEE, 326–330.

[13] Mohammed Hassan and Mohamed Hamada. 2017. A Neural Networks Approach for Improving the Accuracy of Multi-Criteria Recommender Systems. Applied Sciences 7, 9 (2017), 868.

[14] I Ivanto, J Wang, and Fei Liu. 2003. Intelligent online academic management system. In International Conference on Web-Based Learning. Springer, 320–326

[15] Hosein Jafarkarimi, Alex Tze Hiang Sim, and Robab Saadatdoost. 2012. A naive recommendation model for large databases. International Journal of Information and Education Technology 2, 3 (2012), 216.

[16] Raymond J Mooney and Loriene Roy. 2000. Content-based book recommending using learning for text categorization. In Proceedings of the fifth ACM conference on Digital libraries. ACM, 195–204.

Adoption of E-commerce in Nigeria Challenges and Future Prospects

Abbas Muhammad Rabiu*
Department of Computer Science
Federal University Dutse (FUD)
Dutse, Nigeria
mambas86@fud.edu.ng

Abdulhamid Ibrahim
Nigerian Television Authority, (NTA)
Abuja, Nigeria
Ibrahi19@uni.coventry.ac.uk,

Mu'azu Dauda
Department of Computer Science
Federal University, Dutse (FUD)
Dutse Jigawa State
muazudauda@fud.edu.ng

Maryam Ibrahim Mukhtar*
Department of Software Engineering
Bayero University, Kano
Kano, Nigeria
mimukhtar.se@buk.edu.ng

Aisha Muhammad Abdullahi
Department of Computer Science
Federal University, Dutse (FUD)
Dutse, Jigawa State
am.abdullahi@fud.edu.ng

Yusuf Abdulazeez
Department of Computer Science
Federal University, Dutse (FUD)
Dutse, Jigawa State
ludbaw23@gmail.com

Abstract*—Nowadays, Electronic commerce has received attention in many parts of the world, because it has the potential to improve many aspects of business activities. Despite this, there have been recent studies carried out on Electronic commerce adoption in both developed and developing countries. Some of these studies have placed importance on economical and technological infrastructure in countries where E-commerce is to be used. Additional factors that are to be considered to successfully adopt E-commerce are the creation of public awareness, creation of trust amongst online stores and customers, geographical locations and other internal and external influences. This paper is aimed to find out the reason why E-commerce has not been fully adopted in Nigeria though some of the requirements needed are available. The paper will also look at other problems experienced by those countries that have fully benefitted from Electronic commerce activities and the benefit they get from it and to make a proper recommendation for a way forward to fully adopt e-commerce in Nigeria.

Keywords*—Adoption, Business, Countries, E-commerce, Nigeria, Problems.

I. INTRODUCTION

There have been advances in the field of Information and Communications Technology today. This and the emergence of the Internet have made a significant improvement in the way people carry out their business activities [1]. Additionally, a study [1], [7], also stated that E-commerce exists for over 30 years it originated from the electronic transmission of messages during Berling airlift that took place in 1948, it then moved to what was called Electronic Data Interchange (IDE). Different groups of industries came together and created electronic data used for the purchase, and transportation for their transactions [7], [9]. .Electronic commerce is the selling and buying of goods and services over the internet [2], [3]. In the world we live today, Electronic commerce has received attention in many parts of the world because it has the potential to improve many aspects of business activities. Despite this, there have been recent studies carried out on Electronic commerce adoption many developed and developing countries. Some of these studies have placed importance on economical and technological infrastructure in countries where E-commerce is to be used. Additional factors that are to be considered to successfully adopt E-commerce are the creation of public awareness, creation of trust amongst online stores and customers, geographical locations and other internal and external influences.

This paper aims to find out the reasons why E-commerce has not been fully adopted in Nigeria, even though some of these requirements are available. The paper will also look at other problems experienced by those countries that have fully benefitted from Electronic commerce activities and the benefit they got from it and therefore to make a proper recommendation for the possible ways to fully adopt e-commerce in Nigeria. There are different activities involved in E-commerce during the selling and buying of goods and other related services. Some of these activities are Business to Business (B2B), Consumer to Consumer (C2C). The former is involving the information and product buying through the internet while he later involved the business activities carried out between different business firms and their suppliers [5]. There a lot of benefits enjoyed by the business firms who choose E-commerce to buy and sell their goods and services such as gaining higher profits, low operational cost, and gain of customer's loyalty and easy shopping and advantages over rivals, greater reach to high number of customers as they can be accessed over the internet from all over the world [6]. Apart from all these benefits, there are challenges and problems associated with E-commerce as well. They include cybercrimes, delay in delivering the ordered goods, rejection of orders made by the customers especially in some countries like Nigeria. Website's problems related problems such as lack of proper update and maintenance and many more [6]. Other factors that prevent some organization from taking advantage of E-electronic Commerce are lack of proper expertise in some organizations, risks and security issues of carrying out transactions online, lack of trust especially in Nigeria, lack of proper knowledge of computers and inability of some management to adopt E-commerce [2] [6].

II. RELATED WORK

A. Origin of E-commerce and definition

Electronic commerce is defined as the selling and buying of goods and services over thousand of computers connected via a network called the internet [2], [3]. It is also referred to as the process of transferring information, buying, selling and exchanging goods and services through a computer network called the Internet [5]. Electronic commerce has been the area of Information technology that keeps on changing.

A study [7], defined Electronic Commerce as the way of maintaining Business relationship and carrying out business-related transactions through a network of computers called the internet. E-commerce has its main objectives as ensuring that online business is carried out in a better and faster way. It helps to ensure that the buying and selling of goods and services are carried out electronically over the network. This is electronically possible because of the availability of software applications that helps in carrying out business activities on many websites. These applications allow customer display and preview product available on stock, place an order, make payments online using credits cards or opt for payment on delivery for trusted and regular customers in such websites (A good example of this is jumia.com.ng where people order and purchase varieties of product online in Nigeria) [8].

A study [7], also stated that E-commerce exists for over 30 years it originated from the electronic transmission of messages during Berling airlift that took place in 1948, it then moved to what was called Electronic Data Interchange (IDE). Different groups of industries came together and created electronic data used for the purchase, and transportation for their transactions [7], [9]. Initially, before the emergence of the Internet, the IDE system was used to carry out transactions online but the running cost of IDE was higher due to the price of a private network. Therefore only a few giant companies were able to run it and they had to finance smaller companies to enable to them embrace and implement the use of the IDE system. But unlike today, the internet has redefined electronic commerce and many companies are adopting it [7], [10].

B. Benefits of E-commerce

Many companies in a different part of the world can adapt to E-electronic commerce using appropriate technology to carry out online transactions and these have enabled them to transform their business and thereby improving their economy [11]. This has allowed easy access to different information. Therefore, companies that are not willing to adapt to these new technologies may lose their customers to their competitors. Other companies such as Juma and Jiji in Nigeria have been using these technologies to create new ways of advertising and marketing their product and the results is the attraction of thousands of customers and profits making [11].

C. Strategies of Electronic commerce

The availability of the Internet and the increasing number of sites have improved on the efficiency in chain values, building good customer relationship, information disseminations, and revenue generations to industries [12]. This has provided a better way of carrying out business online. This technology has provided better channels to link customers to products and distributors. This type of stores are never closed, customers can make order and purchase product online at any place any time as long as they are connected to the internet. Another advantage of such online stores is that they have no crowd, have many products in stock and inventories, thereby generating higher profits and revenue compared to any physical stores [12]. With the rapid growth of E-commerce, different strategies are emerging at different points in times. Therefore, electronic commerce should be able to meet with these changes so as not to lose customers.

D. Maintenance of Customer's need

One of the responsibilities of companies using E-commerce as a mean of business is to maintain good customer relationships. Companies that are reluctant to maintaining customer's needs, they may end up losing such customers to their competitors [13]. It is however stated in [7] that keeping an existing customer is five to nine times cheaper than gaining new customers. Therefore, one of the most important strategies used to measure the value of the customer is Customer Life Value (CLV). This measuring is made using some important attributes of customer behavior such as frequent orders and purchases made by the customer, acceptance, and payment of delivered items or products purchased through the payment on delivery options. These are some of the historical data that can be easily collected from the websites [7].

E. Maximization of customer's Trust and Loyalty

Trust and loyalty should be created in the mind of the customers. This is very important if a company wants to maintain the customer's relationship. For companies that advocate for the position of their customers through listening to them and by sharing ideas and finding methods that will improve product and services. Companies that engage in doing these strategies, will increase profits and generate more revenue by meeting customers' expectations [7], [14].

F. Provision of good shopping Experience

Business firms should know the behavior and the potential needs of their customers as much as possible to improve their shopping experience through different channels. To improve on the customer shopping experience, companies should hire more workers to handle frequent demands, inquires and request especially during festivities and holidays [7], [15].

G. Online shopping and Customer's Interaction

Though, it is easy for business firms to reach as many customers as possible these days but is it is also difficult to interact with all of them. 70% multi-channel customers shopping online choose to buy their product online because the prices of these products are cheaper than their price in physical stores [7]. Both online and offline channels should be integrated by online retailers. They should use frequent

activities of customers, customers such as customers' profiles, their transaction history, by integrating all this information, companies are likely to capture as much as possible customer's data and make it available to different business store where it is difficult to get those information. These will help in producing a good quality relationship between online and offline channels and enable them to provide better shopping experiences for their customers [7],

H. Provision of Effective Communication to customers

According to [7] that it is good for companies using electronic commerce to provide a direct means of communication for their customers, these will enable customers to compare different prices of different products that are available on the websites. Provision of commission for frequent buying to customers and other incentives will make customers happy and comfortable to provide their confidential information during registration on the website. Retails should be able to keep the confidentiality of customers' information provided in the websites and they should allow customers to unsubscribe to from their website whenever they need to do so.

I. Maintenance of Business Websites

To improve business performances and to ensure the smooth running of business transactions, the website must be maintained whenever the need arises. Customers who experience difficulties and other problems in carrying out transactions on a particular website or don't get feedback from online retailers may decide to use different websites. Hence, live chat should be provided by the company to respond to customers' inquires.

.

J. Huge Investment in customer care service

Customer care service provided by the companies is a necessary activity to ensure the success of any online shopping mall. In the olden days, most of the companies considered customer service to be profit draining cost rather than necessary investment in their business as it is today. Due to the advancement in technology coupled with the internet, there has been an improvement in the provision of customer care services. By so doing, companies can maintain customers' loyalty and satisfaction. Customer care agents are supposed to be fair and humble in dealing with customers' complaints. They should respond as quickly as possible to customers' requests and inquiries. They also need to show some level of understanding to customers. These will no doubt help to increase customers' loyalty and satisfaction [7].

K. Availability of Computers and mobile phone

It is a well-known fact that the level of poverty in developing countries like Nigeria is very high. This makes the availability of computers very difficult especially in a rural area and small cities. Some of the people living in rural communities are low-income earners and jobless, hence this affects the ability of people to buy and own computer and mobile phones [16].

L. Lack of transaction trust among people

Fear in the mind of some people especially in African countries prevents them from carrying out online transactions. Therefore, companies have to find a way of securing their websites, they should improve on the quality of their products to ensure that they meet standard. They have to also ensure that payment using credits card are done safely without any problems [17].

.

M. Penetration of ICT Nigeria and Africa

The rate at which Information and Communication Technology and the Internet is growing has an exponential influence on the online business and the customers in general [17]. For African countries and Nigeria to enjoy this information revolution, it must engage in incorporating the ICT tremendously. In developing countries for example usage and owning of the email address are common among teaming populace. However, in Nigeria, it is very common to find even the so-called educated person without having an email address which is the first thing to use to register with any website [17]. There have been called by policy-makers in Nigeria to link up its ICT with the global network society. To ensure this is a success, the National Communications Commission (NCC) in Nigeria allows for the broadcasting of private electronic media. This can be used to increase the growth of the economy and business activities in Nigeria and Africa at large [17], [19].

N. Penetration of Ecommerce in Nigeria

It is noticeable that the availability and internet access in Nigeria is increasing everyday especially in the rural areas. This has given rise to growing E-commerce and business activities in Nigeria. Many countries in the world continue to focus on the Internet for growth and development. It was found that in the UK only E-commerce generated about 100 billion pounds every year which accounts for 7.2% of Britain's gross domestic product. It was also estimated by the Internet World Stats that 59, 466,249 which represents over 29.5 % of the Nigerian population used the internet [20]. In spite of all these, most of the Nigerian communities do not use the Internet broad-band because it is deemed to be unreliably expensive, too slow. Many of the Internet users in Nigeria prefer to use mobile phones due to their mobility and reliability. This makes the usage of E-commerce in Nigeria very low especially among the rural communities [27], [28], [29]. The reason for this was found to be as a results of lack of proper electricity supply as the major factor, cyber-crimes, lack of trust among Nigerians, lack of awareness, lack of innovations, resistance to changes especially amongst the rural communities, lack of good infrastructures and lack of government support and proper policy to address cyber-crimes related activities among others [28], [34],[35], [36]. Several strategies have been established to promote E-commerce utilization in Nigeria such as cashless policy; however, customers do not very much appreciate them [29].

III. METHODOLOGY

Quantitative research is a type of research used in this paper. It has the aim of determining the relationship that exists between the dependent and independent variables [21]. It helps the researcher to examine the nature of a particular study from the data that has been collected and analyzed [22].

The data that have been collected was used to evaluate the results using graphical and statically approach in this paper.

Type of Data Used

The data used for this research comes from both primary and secondary sources. And accurate data is used to ensure that the results and findings are both reliable and valid.

Respondents used for the survey

Nigerians communities living in Jigawa and Kano State respectively are the respondents used for the survey. The respondents are all educated people with different levels of education and background. This is because educated people are the ones mostly using electronic media to buy goods and services. For example, Jumia is the most trusted and dominant online shop in Nigeria which is patronized by some of these people. The researchers believed that these classes of people will be in a better position to respond to question related to Electronic Commerce activities as they are involved in doing it.

IV. RESULTS AND DATA ANALYSIS

A. *Response 1: Respondents age distribution*

Table 1 Age distributions of respondents

Ages	17-20	21-26	26-30	Above30	Total
Number	8	18	5	6	37
Percentage (%)	23	49.3	15.5	14.1	100

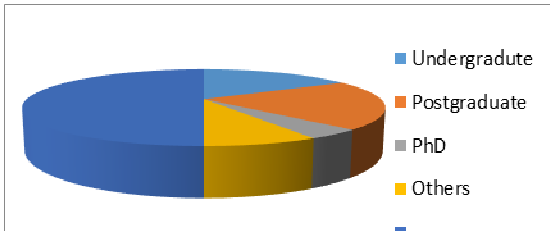


Fig. 1: Showing age distributions of respondents

Fig. 1 shows that 8 people, representing a 23% fall between the ages of 17 & 20. While those between the age of 21 and 26 are 18 in number, this represents 49.3% of the total population. While those between the age of 26 and 30 representing 15.5% were 5, the last group fall between the ages of 30 and above represents 14.1% of the total population.

B. *Response 2: Level of Education of the respondents*

Table 2: Educational level of the respondents

Age	Undergrad.	Postgrad.	PhD	Others	Total
Distribution	14	13	4	6	37
Percentage (%)	31	42.6	9.8	15.6	100

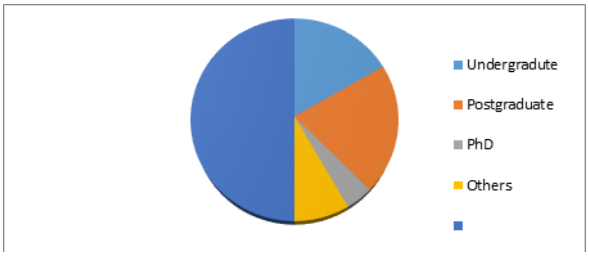


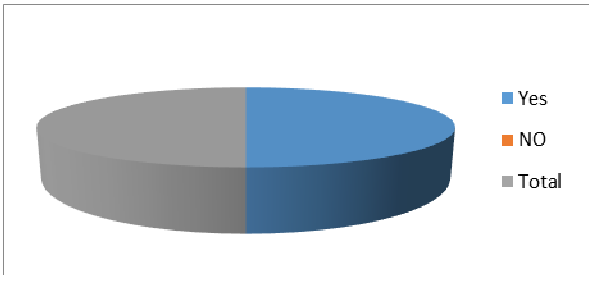
Fig 2: Showing level of education of the respondents

From Table 2 and Fig. 2 above, it can be seen that 14 representing 31% of the respondents were undergraduate. 13 people representing 42.6% were postgraduate while 4 respondents representing 9.8% were Ph.D. holders, other categories were six in number representing 15.6% of the total population.

Response 3

Table 3: Level of awareness of the respondents

Responses	Yes	NO	Total
Number	37	0	37
Percentage	100%	0%	100%



F

The Table 3 and fig. 3 above shows that 37% representing 100% gave positive response ('YES') response while 0% of the respondents answered ('NO') to the question.

Response 4

Table 4: level of usage of E-commerce

Respondents	Yes	NO	Total
Number	30	7	37
Percentage (%)	86	14	100

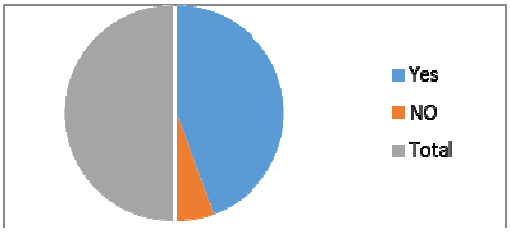


Fig 4: Showing the level of usage of E-commerce

From Fig. 4 and Table 4 above show that 7 people representing 14% of the respondents answered (YES) to the question when they were asked about the level of usage of E-commerce. 30 people representing 86% of the respondents chose (NO) as their answers.

Response 5

Table 5: Technological infrastructural impact

Responses	S/agree	Agree	Neutral	Disagree	S/Disagree	Total
Distribution	20	9	1	6	1	37
Percentage (%)	50	30	1	18	1	100

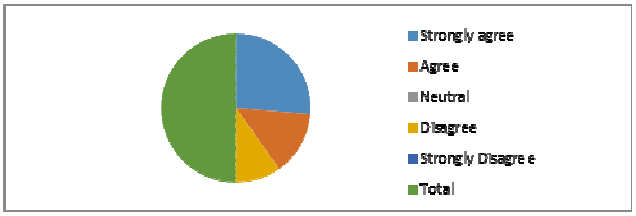


Fig. 5: Showing the impact of technological infrastructure

From Fig. 5 above, it is seen that 50%, Representing 20 people strongly agreed on that lack of technological infrastructure has an impact on E-commerce. 9 people, representing 30% of the total population agreed on the impact. One person remains neutral about the impact, six people disagreed and one person strongly disagrees with the effect of technological infrastructure on the adoption of E-commerce.

F. Response 6

Table 6: Respondents' view on the costs of Implementation

Responses	S/Agree	Agree	Neutral	Disagree	S/Disagree	Total
Number	14	12	1	6	4	37
Percentage (%)	37.84	32.43	2.7	16.22	10.81	100

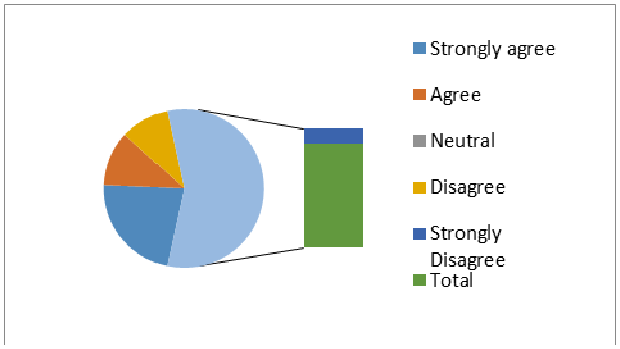


Fig 6: Showing responses on the costs of Implementation

From Fig. 6 and Table 6, it is can be seen that 14 persons representing 37.84% strongly disagreed that the cost of implementation affects E-commerce adoption in Nigeria. 12 persons representing 32.43% agreed with the effects, 1 person representing 2.7% remained neutral, 6 people representing 16.22 percents disagreed with the earlier view and 4 people representing 10.81% strongly disagreed that the cost of implementation effects E-commerce adoption Nigeria.

G. Response 7

Table 7: Showing the importance of user's confidentiality

Responses	S/agree	Agree	Neut	Disagree	S/Disagree	Total
Number	18	14	3	1	1	37
Percentage (%)	48.65	37.84	8.10	2.70	2.70	100

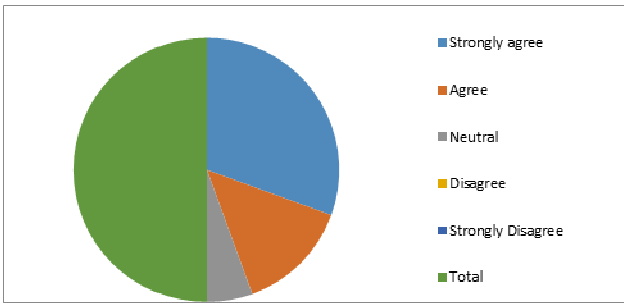


Fig 7: Showing the importance of user's confidentiality

From Fig. 7 and Table 7, it is also clear that 18 persons representing 48.65% strongly disagreed that the confidentiality of users affects the adoption of E-commerce in Nigeria. 14 persons representing 37.84% agreed with the effects, 3 persons representing 8.4% remained neutral, 1 person representing 12.70 percent disagreed with the earlier view and finally, 1 person representing 2.70% strongly disagreed that that lack of confidentiality of users' information affects E-commerce adoption Nigeria.

H. Response 8

Table 8: Effects of fraud on the adoption of E-commerce

Responses	S/agree	Agree	Neut	Disagree	S/Disagree	Total
Number	14	13	1	6	3	37
Percentage (%)	37.84	35.12	2.7	16.22	8.12	100

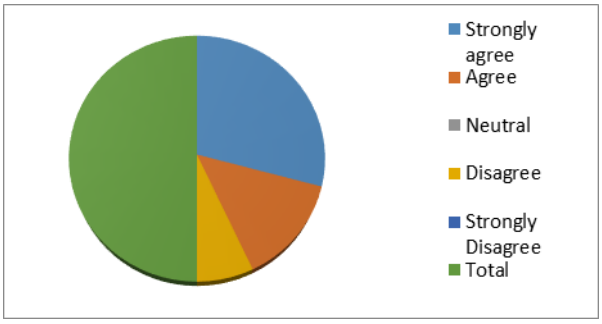


Fig. 8: Showing effects of online fraud on E-commerce

From Fig. 8 and Table 8 above, it is noticeable that 37.84%, Representing 14 people strongly agreed that online fraud affects E-commerce adoption in Nigeria. 13 people, representing 35.12% of the total population agreed with the effects. One person representing 2.7% of the total population remained neutral about the effects, six people representing 16.22% disagreed and 3 people strongly disagreed with the effects of the online fraud on the success of E-commence in Nigeria.

I. Response 9

Table 9: Showing effects of low Internet access

Responses	S/ agree	Agree	Neutral	Disagree	S/Dis agree	Total
Number	10	12	1	9	5	37
Percentage (%)	67.57	32.43	2.70	24.32	13.51	100

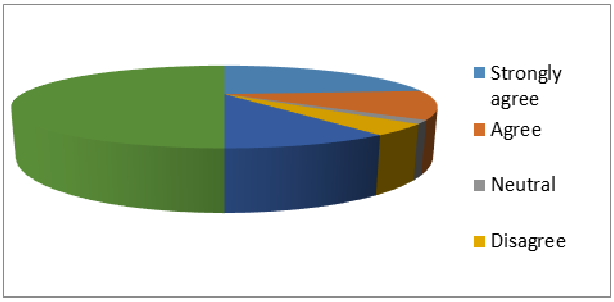


Fig.9: Showing effects of low Internet access

Fig 9 and Table 9 above show answers given by different respondents on the effects of low Internet access on the adoption of E-commerce in Nigeria. 67.57%, representing 10 people strongly agreed that low Internet penetration has effects on the E-commerce adoption in Nigeria. 1 person representing 2.70% of the total population was neutral about the effects. Nine people representing 13.51% disagreed with the negative effects of low Internet access on the adoption of E-commerce in Nigeria.

J. Response 10

Table 10: Effects of quality of merchandise

Responses	S/agree	Agree	Neutral	Disagree	S/Disagree	Total
Number	15	10	0	5	7	37
Percentage(%)	40.54	27.03	0	13.51	18.91	100

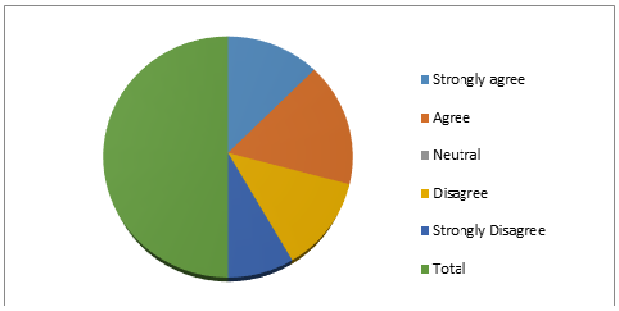


Fig. 10: Effects of quality of merchandise on E-commerce

Fig. 10 and Table 10 above show the answers given by different respondents on the effects of quality of merchandise on E-commerce adoption in Nigeria. 40.54%, representing 15 people strongly agreed that uncertainty of merchandize quality affects E-commerce adoption in Nigeria. 10 people representing 27.03% of the total population agreed that the uncertainty about the quality of product purchased online prevents people from buying online. Nobody was neutral about the effect. 5 people representing 18.91% disagreed while 7 people agreed that uncertainty of the quality of the product sold online effects the adoption of E-commerce in Nigeria.

V. DISCUSSION:

A. Effects of Infrastructural technology

From the results and analysis carried out so far, it can be seen that a higher percentage of the people responded that infrastructural technology significantly affects the adoption of E-commerce in Nigeria. Therefore, by following the study [2] which stated that for E-commerce to be fully adopted, technological infrastructure and access to computers by the customers must be ensured.

B. Effects of the cost of implementation:

It can be observed from the responses that the cost of implementation hinders the success of E-commerce in Nigeria [22]. To make this a success, users of the Internet in Nigeria and Africa at large need to have cheaper Internet connectivity as it is the case with some countries like India.

C. Lack of confidentiality of users

From the results and analysis carried, it can be observed that a higher percentage of the people responded that uncertainties over the confidentiality of the information they give online prevent them from buying products online. Some people are afraid of making their information available online due to current cyber-security issues, kidnapping and other security issues in Nigeria. This is in accordance with the study [14] which stated that the confidentiality of users must be in place for E-commerce to be accepted by the people [14].

D. Negative effects of online Fraud on E-commerce

By following the study [23] and [24], online fraud posed a threat to the successes of E-commerce adoption in Nigeria. This has been confirmed by the respondents with the majority of them agreeing on the negative effects of fraud on the success of E-commerce.

E. Effect of low Internet Penetration in Nigeria

A Significant number of respondents from the survey carried out agreed that low Internet penetration in Nigeria has been a problem hindering the success of E-commerce adoption in Nigeria. However, the government can help in bridging this gap by providing immense support through funding to provide Internet connectivity in schools and other important places.

F. Effects of Merchandise Quality on E-commerce

To reduce these effects, online marketers such as Jumia in Nigeria use a return policy to enable customers to return defective products within stipulated time. Using this policy will reduce the level of uncertainties which prevent people from buying online [25]. Other problems preventing people from buying products and services online are lack of good incomes because most of the people living in rural communities are poor and they lack proper education to use computers and Internet [12]. Others problems are lack of availability of good payment channels such as credits card which are used to make payment online.

G. *Lack of trust amongst citizens of Nigeria and Africa*

Lack of trust amongst the citizen of developing countries posed a great threat to the successful adoption of E-commerce in Nigeria. For this to be a success, trustworthy transactions need to be carried out to serve as an example for E-commerce to be accepted by the Nigerian citizens.

H. *Sufficient computer knowledge*

From the survey carried out, most of the respondents agreed that the lack of proper knowledge of computers affects E-commerce adoption in Nigeria. According to [26] lack of proper computer and ICT knowledge in developing countries are two of the problems hindering the success of E-commerce in such countries.

I. *Successful adoption of E-commerce in Nigeria*

To successfully adopt E-commerce in Nigeria we should follow similar ways and frame work used by some developing countries like Bangladesh and Botswana. Because Nigeria like some of these countries faces similar challenges. Some of these challenges were addressed in the study [27]. However, for purpose of this paper two aspects would be proposed. Firstly, the Nigerian government through Nigerian Communication Commissions (NCC) should spare head and put in place Government to Business policy (G2B) as described in the study [36] to encourage E-commerce activities amongst private sectors to maximize profits. The Nigerian government is the economic driver, like any other developing country, its policies will significantly impact on the provision of new policies such as E-commerce activities in the country. Secondly, the creation of Agro based related activities will significantly help people living in rural communities as it is related to agriculture. This involves the provision of capacity building, logistic supports, coordination, information and corporation and awareness creation amongst the people living in the rural areas. It has been discussed earlier that poverty is one of the problems hindering the success of E-commerce activities in Nigeria [20], [37].

V. CONCLUSION

Having deeply explored different research that has been carried out to ensure the successful adoption of E-commerce is achieved. This paper makes comparisons with some situation in the developing countries that are similar to that of Nigeria. A survey was generated based on this relationship. From the analysis carried out in this paper, most of the findings are consistent with the results generated from previous studies in this subject. Hence, this enables the researchers in this paper to propose some frame-works that are similar to two countries; Botswana and Bangladesh. This is because Nigeria has many things in common with these countries and a similar approach can be used to ensure the successful adoption of E-commerce in Nigeria. Other problems that hinder E-commerce adoption in Nigeria discussed earlier such as access to the Internet, improved public awareness, trustworthy transactions, and technological infrastructure are seen as the tools that would help provide successful E-commerce adoption in Nigeria if they are put in place.

References

[1] L. Aghaunor, and X. Fotoh, "Factors affecting Ecommerce adoption in Nigeria Banks", 2006, pp 34-56.

[2] J.E. Lawrence, and U.A. Tar, " Barriers to Ecommerce in Developing Countries", *Information, Security and Justice,* vol. 3, No. 1. 2010, pp. 23-25.

[3] K. Charles and Ayo," The Prospects of E-Commerce Implementation in Nigeria", *Journal of Internet Banking& Commerce,* Vol. 11, Issue 3, 2016, pp3.

S. [4]Kurnia, T . , Maiorescu, and S . V . Mirescu" The Premises and the Evolution of Electronic Commerce" *Journal of Knowledge Management, Economics and Information technology*, vol. 4, 2011, pp 4-7.

[5] A. Ahmed etal., " Potential E-Commerce Adoption Strategies for Libyan Organization", *International Journal of Information and Communication Technology Research,* Vol. 1, No. 7, 2011, PP5-7.

[6] J. Fisher, et al., "Electronic Commerce Benefits, Challenges and Success: Factors in the Australian Banking and Finance Industry", 2012, pp3.

[7] V. Zwass,"Electronic Commerce and Organizational Innovation: Aspects and Opportunities "*International Journal of Electronic Commerce.*2018. pp5

[8] J. Cooperetal.," What Prevente Commerce Adoption in Developing Countries?: A Socio-Cultural Perspective" *In:17th Blede Commerce Confrence, Slovenia,* June 2004, pp 3-7.

[9] J.L. King, T., Mc Gann, K. Lyytinen,"Globalization of E- Commerce: Growth and Impacts in the United States of America, "*Case Western Reserve University, USA. Sprouts: Working Papers on Information Systems*2012, pp2-5.

[10] O. Kirwan,"An Action Research Case Study of the Facilitators and Inhibitors of E-Commerce Adoption", *International Business Research*, Vol.2, No.1,2009, pp4-9.

[11] T. Maiorescu, andS . V . Mirescu, " The Premises and The Evolution of Electronic Commerce" *Journal of Knowledge Management, Economics and Information technology.* 2019, Vol2. pp 4-6

[12] Y.I Harande, " Information Services for Rural Community Development in Nigeria", Available at: http://www.webpages.uidaho.edu/~mbolin/harande.htmLast accessed20thOctober, 2018.

[13] L.K. Geller, " CRM: What does it mean?" *Target Marketing, Philadelphia,* August 2016, pp 23-25.

[14] M. Bressler, " Internet CRM must have human touch" *Marketing News*, Chicago, October 2012.

[15] A.D. Smith, "Exploring online dating and customer relationship management" *Online Information Review, Bradford,*2012, pp.18-34.

[16] J. Fisher et al., "Electronic Commerce Benefits, Challenges and Success Factors in the Australian Banking and Finance Industry" Sept., 2017, pp 34-40.

[17] P.I. Akpan," Basic needs to globalization: Are ICTs the Missing link? "*Information technology for Development.* 2013, pp 261-274.

[19] I. Apulu, and . Latham, " Information and

Communication Technology Adoption: Challenges for Nigerian SMEs". *TMC Academic Journal.* Vol. 4, Issue 2, 2016, pp 64-80.
[20] A. Al-Nahidet.," A Generic Frame work for Implementing Electronic Commerce in Developing Countries", *International Journal of Information and Communication Technology.* Vol. 1, Issue 2. 2017, pp 4-5.
[21] W.L Neuman,." Social Research Methods: Qualitative and Quantitative Approaches" Boston: Allyn and Bacon, 2017, pp3-7.
[22] A. Ikpefan, and S . Ojeka, " Electronic Commerce, Automation and Online Banking in Nigeria: Challenges and Benefits". *School for Doctorate Studies (EuropeanUnion) Journal.*2018, pp 5-8.
[23]Akintolaetal.," Appraising Nigeria Readiness for Ecommerce Towards: Achieving Vision 20:2020".*International Journal of Research and Reviews in Applied Science*. Volume 9, Issue 2, pp 6-8
[24] A. Bola, and O . R , Ehimen, " Cyber-crime in Nigeria". *Business Intelligence Journal*, Vol 2, 2017,pp 5-9
[25] B. Anckar, " Drivers and Inhibitors of E-commerce Adoption: Exploring the Rationality of Consumer Behavior in the Electronic Marketplace". 2012, pp3-7.
[26] Adewoyeetal., "Business-to-Consumer E-commerce in Nigeria: Prospects and Challenges". *African Journal of Business Management*, Vol. 5, Issue 1, 2016, pp 5109–5117,
[27] O.D. Faloye, "The adoption of e-commerce in small businesses: an empirical evidence from retail sector in Nigeria." *Journal of Business and Retail Management Research*, vol. 8, Issue (2), 2014, pp 54-64.
[28] B.E. Akanbi, T.S. Akintunde, "E-Commerce Adoption And Small Medium Scale Enterprises Performance In Nigeria." *European Journal Of Management And Marketing Studies*, vol. 3, Issue 1, 2018, pp.10-24.
[29] H.U. Khan, and S. Uwemi, "Possible impact of e-commerce strategies on the utilization of e-commerce in Nigeria." *International Journal of Business Innovation and Research,* vol 15, Issue 2, 2018, pp.231-246.
[30] I.Ekanem, G.E. Abiade, "Factors influencing the use of E-commerce by Small Enterprises in Nigeria." *International Journal of ICT Research in Africa and the Middle East (IJICTRAME)*, vol 7, Issue 1, 2018, pp.37-53
[31] A.A. Taiwo et al., "Nexus Between Users' Perceived Risk, Perceived Legislation And E Commerce Acceptance: Implications For E-Commerce Adoption In Southwest Nigeria." *International Journal of Mechanical Engineering and Technology (IJMET),* vol. 10, Issue 1, 2019, pp.2021-2031.
[32] A. Ogbo, C.C. Ugwu, J. Enemuo, and W.I. Ukpere, "E-Commerce as a Strategy for Sustainable Value Creation among Selected Traditional Open Market Retailers in Enugu State, Nigeria." *Sustainability,* vol. 11, Issue 16, 2019, pp.4360.
[33]E. Agwu, and P.J, Peter, "Drivers and Inhibitors to E-Commerce Adoption among SMEs in Nigeria." *Journal of Emerging Trends in Computing and Information Sciences*, vol 5, Issue 3, 2014, pp.192-195.
[34]E. Agwu, and P.J, Peter, "Empirical Study of Barriers to Electronic Commerce Adoption by Small and Medium Scale Businesses in Nigeria." *International Journal of Innovation in the Digital Economy*, vol 6, Issue 2, 2015, pp.1-19.
[35]T.A. Akanbi, "An Investigative Study of Challenges Facing Nigerian Small and Medium Scale Enterprises in Adoption of E-Commerce Technology." IJAME, vol. 5, Issue 1, 2016, pp.22-31.
[36]H.U. Khan, and S. Uwemi, "What are e-commerce possible challenges in developing countries: a case study of Nigeria." *International Journal of Business and System Research*, vol. 12, Issue 4, 2018, pp.454-486.

Disturbance Events Analysis and Partial System Collapse Hunting on the Nigerian Power System Using Frequency Disturbance Recorder (FDR)

S. K. Mohammed, U. O. Aliyu, G. A. Bakare, Y. S. Haruna, M. Ahmed and I. T. Balami
Department of Electrical and Electronics Engineering, Faculty of Engineering and Engineering Technology, Abubakar Tafawa Balewa University, Bauchi, Nigeria

mskumo@atbu.edu.ng, gabakare@atbu.edu.ng, uoaliyu@atbu.edu.ng, ysharuna@atbu.edu.ng, amohammed3@atbu.edu.ng, and tbishaku@atbu.edu.ng

Abstract*—Frequency is a very important parameter in an electrical power system's control, protection and stability. It must therefore be maintained very close to the nominal value of 50 Hz, as any disturbance in substantial quantity, be it loss of load or generation causes frequency deviation and subsequent propagation throughout the system. In this paper, the frequency data of Nigerian power system recorded at Bauchi using Frequency Disturbance Recorder (FDR) was used to detect and analyze disturbance events and partial system collapse that propagated from their various occurrence locations to Bauchi through electromechanical wave propagation. The effects of system's load-generation imbalances on the frequency were graphically shown and the detected disturbance events were analyzed and validated by the operators' data, subsequently the system load-frequency sensitivity coefficient, β was determined based on the loss of 140 MW generation unit at Shiroro to be 34.15MW/0.1Hz indicating improvement against 29MW/0.1Hz obtained in 2006.

Keywords*— Frequency, Electromechanical Wave, Disturbance, Loss of Load-Generation, Frequency Deviation, Wave Propagation.

I. INTRODUCTION

The synchronized frequency measurement network is centered around the notion which believes that the system's frequency is always constant during steady state operation, although fluctuates when the power grid is affected by substantial disturbance events [1]–[5]. In the event of improper coordination of control, protection and communication, the grid could be susceptible to anomalies including system collapse [6], disturbances and human operation errors [7], [8]. Power system status awareness lends a hand in avoiding blackouts, the absence of which results in inevitable system blackouts [8], [9].

As stated, the frequency within the vast grid is the same under steady state operation, therefore its monitoring offers an understanding on the features of the system dynamics. The system generation-demand disparity, control actions and disturbance events have a direct consequence on its frequency [10]. Disturbance events can be caused by the system oscillations, generator trip, loss of transmission line and loss of load. In the event of generator trip, frequency of the system drops, fluctuates for a while and stabilizes at a lower post-disturbance value [11], [12]. The post disturbance frequency differs from its earlier value before the disturbance by

$$\Delta f = f_{ss} - f \quad \ldots (1)$$

Where Δf is the frequency deviation, f_{SS} is the steady state frequency after the disturbance event, and f is the averaged pre-disturbance frequency. The extent of generation loss, P_G is determined from (2).

$$P_G = \beta \Delta f \quad \ldots (2)$$

where β is the load-frequency sensitivity coefficient [13], [14].

Loss of load can be observed in the frequency only when a significant amount of load is lost. The amount of load lost P_L can be estimated from (3).

$$P_L = \beta \Delta f \quad \ldots (3)$$

The quantity of line flow deviation is proportional to the closeness of the lost load. In a heavily meshed network loss of a transmission line commonly results in line flows rearrangement around the vicinity of the trip.

The change in time, Δt is given by

$$\Delta t = t_{ss} - t \quad \ldots (4)$$

where t_{ss} is the time to attain steady-state after the disturbance and t is the averaged time before the disturbance.

The occurrence of undamped oscillations poses concern in power system operations, signifying the likelihood of probable system instabilities. In this situation, oscillations are detectable in all measured variables: frequency, voltage and line flow. At times such event last about a minute.

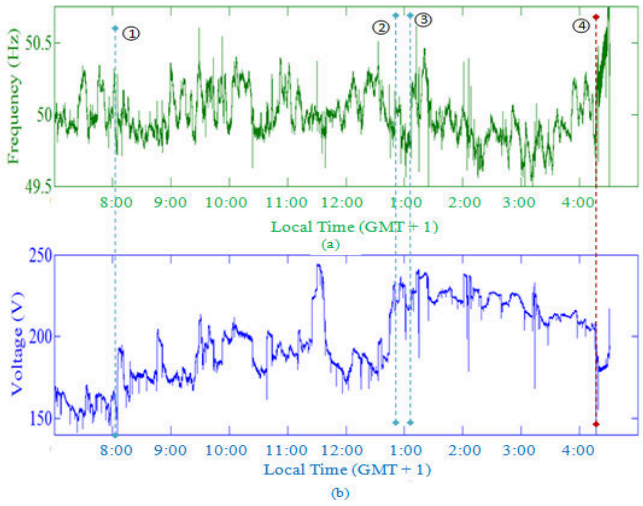
II. METHODOLOGY

An FDR was set up as in [7], [10] for the dynamic frequency and voltage recordings. It was used to record the system frequency and voltage of the Power Holdings Company of Nigeria (PHCN) on February 02. The data obtained was preprocessed as in [10]. Disturbance events information for the

same day was obtained from the PHCN operators to help in validating the events detected in the data.

III. RESULTS, DISCUSSION AND ANALYSIS

The Fig. 1 shows the plot of frequency and voltage against time for the data recorded on 02 February, with its associated disturbance events indicated in the circled numbers. The event 1 indicates the disturbances between 8:03 to 8:08 a.m. with the major events of a 135 MW Shiroro generation unit shut down and consequently a 55 MW load shed on Kaduna-Kano 132 kV line at Mando/Zaria at 8:04:48 and 8:05:07 a.m. respectively. The second event of the day is the event 2 between 12:56 and 1:00 p.m. whose main event was a 140 MW Shiroro generation unit deloading itself at 12:58 p.m. Event 3 covers the period between 1:03 and 1:12 p.m. with major events at 1:03:35 and 1:09:20 p.m. corresponding to Sapele ST2 55 MW generation trip and a 121 MW load shed respectively. Finally, a partial system collapse affecting the whole North-Eastern region occurred at 4:22:36 p.m. as shown on event 4 indicated by red broken lines. Fig. 1 is elaborated and discussed in the subsequent figures.



Disturbance events of February 02 (a) Frequency (b) Voltage

Discussion of the major disturbance events of February 02

Loss of generation disturbance event is characterized with decrease in system frequency. Likewise, loss of load event is characterized with a frequency rise as can be seen in Fig. 2. In which events 1 and 3 indicate sudden load connection to the grid or a loss of remote generation while events 2 and 4 indicate a loss of load or load shed from the grid. The frequency and voltage deviations caused by the disturbances are always proportional to their magnitudes and proximity.

The events of Fig. 3 are also elaborated in the same vein with the major event recorded at 12:58:35 p.m. between events 5 and 6. This was a loss of 140 MW generation unit at Shiroro. The frequency dynamics are shown in the inscribed window.

The frequency variation is determined as

$\Delta f = f_1 - f_2$

$= 50.07 - 49.66 = 0.41$ Hz

The time change at the frequency variation is

$\Delta t = t_2 - t_1$

$= 12:58:45 - 12:58:35 = 10$ sec.

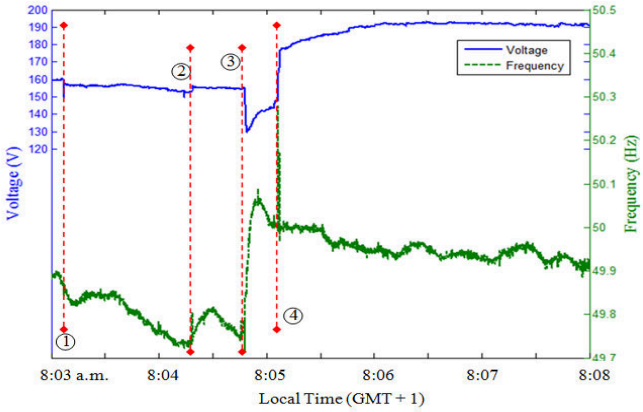
Generated power lost, $P_G = 140$ MW

The frequency sensitivity coefficient, β is calculated as

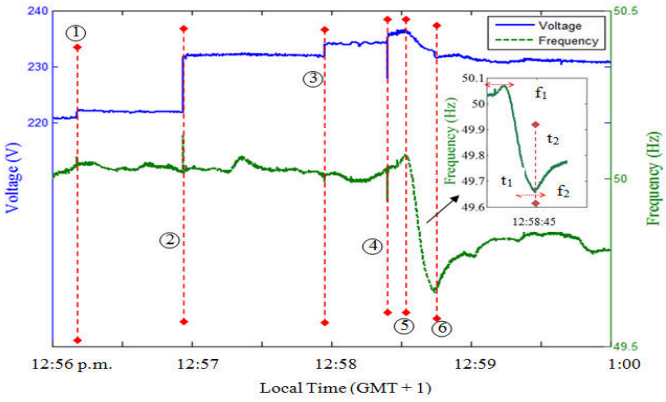
$\beta = 140$ MW/ 0.41 Hz = 34.15 MW/ 0.1 Hz

This value indicates improvement against the value 29 MW/0.1 Hz obtained in 2006.

The frequency declined at the rate of 4.1 mHz /0.1 sec [15].



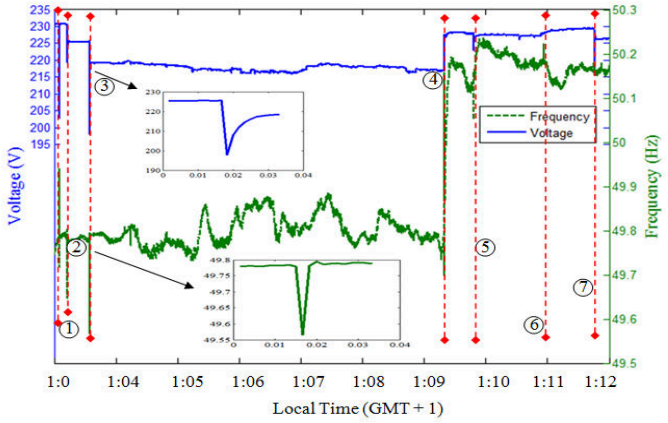
FDR recordings captured between 8:03 a.m. and 8:08 a. m.



FDR recordings captured between 12:56 p.m. and 1:00 p. m.

The major events of Fig. 4 are the tripping of Sapele ST2 55 MW generation unit at 1:03:35 p.m. and a 330 kV circuit (121 MW) load shed at 1:09:20 p.m. The former event is

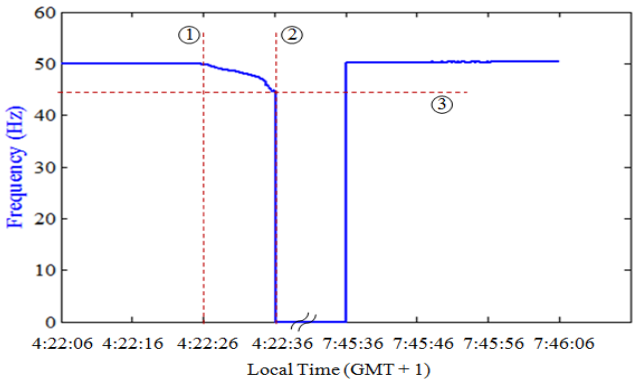
expanded and shown as in the inscribed windows with system stiffness of 25 MW/ 0.1 Hz. The other marked events are justified by their variations in both frequencies and voltages as a sign of disturbance. Situations that show variations in only the voltage or frequency are local events such as connecting or shedding a load locally at the substation.



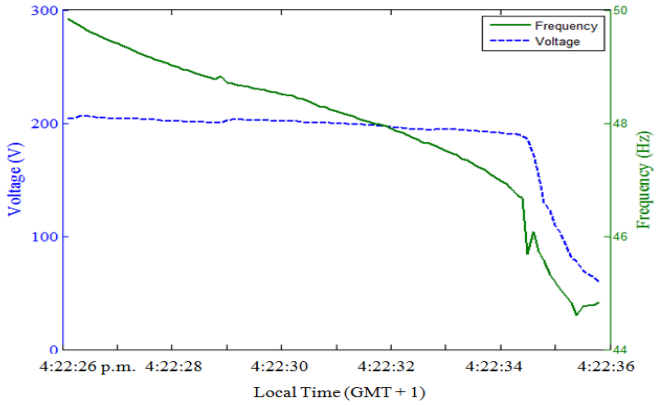
FDR recordings captured between 1:03 p.m. and 1:12 p. m.

Partial system collapse of wednesday 02 February

It can be observed from Fig. 5 that the system frequency had been around the nominal frequency of 49.89 Hz until at 4:22:26 p.m. when it started declining steadily to a value of 44.61 Hz at 4:22:36 p.m. when the system eventually collapsed with a frequency decay rate of 52.8 mHz/ 0.1 sec or 528 mHz per second. After the collapse the system remained down until 7:45:36 p.m. when there was restoration. Fig. 6 gives the elaboration of the frequency and voltage as they declined during the collapse. The voltage had been steady even though the frequency started decreasing until at the point of the collapse when it suddenly dropped to 60 V from about 200 V.



FDR recordings dynamics showing partial system collapse and restoration between 4:22:26 p.m. and 4:22:26 p. m.



Details of the FDR recorded partial system collapse

IV. SUMMARY AND CONCLUSSION

It can be observed that FDR data can be used to detect and analyze disturbance events such as loss of generation, loss of load, loss of transmission lines and a system collapse in an electrical power system. This is due to the fact that the system frequency characterizes its electromechanical properties. A loss of generation causes a drop-in frequency while loss of load or load shed causes increase in system frequency.

References

[1] Z. Lin et al., "Application of Wide Area Measurement Systems to Islanding Detection of Bulk Power Systems," vol. 28, no. 2, pp. 2006–2015, 2013.

[2] V. Salehi and P. Mehr, "Development and verification of control and protection strategies in hybrid ac / dc power systems for smart grid applications". A dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy, Miami, Florida, 2012.

[3] Y. L. Y. Liu, "A US-Wide Power Systems Frequency Monitoring Network," *2006 IEEE PES Power Syst. Conf. Expo.*, pp. 1–8, 2006.

[4] Y. Liu, "A US-Wide Power Systems Frequency Monitoring Network," *Power Syst. Conf. Expo. 2006. PSCE '06. 2006 IEEE PES*, pp. 159–166, 2006.

[5] Y. Zhang *et al.*, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 159–167, 2010.

[6] I. Adaji and A. G. Raji "Prevention of Voltage Collapse on the Nigerian 330 kV Grid Network Using Under-Frequency Load Shedding Scheme," pp. 533–537, IEEE PES-IAS PowerAfrica Prevention 2017.

[7] Z. Zhong *et al.*, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1914–1921, 2005

[8] E. Oleka, S. Ndubisi, and G. L. Lebby, "Synchrophasor for Enhancing Operational Reliability of Nigerian Electric Power Grid," pp. 5–10, 2017.

[9] N. N. Sabrina and H. A. Halim, “Security Enhancement of Power Systems with Smart Grid Implementation,” 2012.

[10] L. Vanfretti *et al.*, “Estimation of the Nigerian Power System Electromechanical Modes using FDR Measurements,” in *the 9th International Conference on Power System Operations and Planning (ISCPSOP)*, 2010.

[11] L. Vanfretti, U. Aliyu, J. H. Chow, and J. a. Momoh, “System frequency monitoring in the nigerian power system,” *2009 IEEE Power Energy Soc. Gen. Meet. PES ’09*, pp. 1–6, 2009.

[12] M. Farrokhabadi and L. Vanfretti, “State-of-the-art of topology processors for EMS and PMU applications and their limitations,” *IECON Proc. (Industrial Electron. Conf.*, pp. 1422–1427, 2012.

[13] J. H. Chow, A. Chakrabortty, L. Vanfretti, and M. Arcak, “Estimation of radial power system transfer path dynamic parameters using synchronized phasor data,” *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 564–571, 2008.

[14] L. Vanfretti, J. H. Chow, S. Sarawgi, D. Ellis, and B. Fardanesh, “A framework for estimation of power systems based on synchronized phasor measurement data,” *2009 IEEE Power Energy Soc. Gen. Meet.*, 2009.

[15] L. Vanfretti and F. Milano, “Triggering the deep learning approach in power system courses using Free and Open Source Software,” *IEEE Power Energy Soc. Gen. Meet.*, 2011

EFFECT OF YOUTUBE-VIDEOEMBEDDED INSTRUCTION ON STUDENTS' ACADEMIC ACHIEVEMENT IN AUTOMOTIVE TECHNOLOGY EDUCATION IN TERTIARY INSTITUTIONS OF NORTH-EASTERN NIGERIA

Isma'ila Y. Shehu
Department of Vocational and Technology Education, Faculty of Technology Education, Abubakar Tafawa Balewa University,Bauchi
Nigeria
iyshehu@gmail.com

Umar Abubakar
Department of Vocational and Technology Education, Faculty of Technology Education, Abubakar Tafawa Balewa University, Bauchi
Nigeria
umarmaintainor@gmail.com

Aminu M. Kawu
NDE/MDGs Skills Acquisition Centre
Minchika, Adamawa
Nigeria
aminukawu6@gmail.com

Bala Sa'idu
Department of Vocational and Technology Education, Faculty of Technology Education, Abubakar Tafawa Balewa University, Bauchi
Nigeria
balasaidu58@gmail.com

Abstract*—The study determined the effect of YouTube-VideoEmbedded Instruction (YVEI) on students' academic achievement in automotive technology education in tertiary institutions of north-eastern Nigeria.Quasi experimental research design (pretest and posttest) was employed in the study. The instrument used for data collection was Automotive Suspension System Academic Achievement Test (ASSAAT) validated by three experts in the area. The reliability of the instrument was established using KR-21 and 0.81 co-efficient was obtained. The data was analyzed using mean, standard deviation and t-test. Findings of the study revealed that students in the experimental group had higher academic achievement than the control group. Based on this finding, it was concluded that YouTube-Video Embedded Instruction will assist in improving students' academic achievement in Automotive Technology Education in tertiary institutions of Nigeria.Therefore, it is recommended that YouTube-Video Embedded Instruction should be integrated into the teaching of Automotive Technology Education courses in tertiary institutions of north-eastern Nigeria. This study has an implication on providing an opportunity of relating theories acquired in classroom with practical in an easy manner which will in turn improve students' academic achievements. Lastly, it was suggested that this study should be replicated in other areas of technology education.

Keywords*—YouTube -Video; Embedded Instruction; Automotive Technology Education; Academic Achievement

I. INTRODUCTION

Automotive technology is one of the areas of technical and vocational education that contributes immensely to economic development of many countries. It is devastating to notice that in many developing countries like Nigeria many students in tertiary institutions perform poorly in automotive technology practical. That is to say, they could not relate the theory acquired in the classroom to the real-life situation when it comes to repairs and maintenance of automotive vehicle components. This is further confirmed by [1] who stated that automobile technology education graduates display low level of skills in carrying out practical skills. As such, for them to be proficient in the field they need to improve much in both theoretical and practical aspects. Although, several factors such as; lack of equipment, poor attitude; inappropriate (traditional) method of instruction and inconducive learning environment may be attributable to this situation. Review of some researches revealed that this situation may not be unconnected with the use of conventional method of teaching; which at times is very inappropriate for teaching automotive technology education courses in the tertiary institutions [2,3,4,]. In addition, [5,6] contended that the use of conventional method does not contribute much in teaching automotive courses as well as improving students' academic achievement which in turn makes the students become passive rather than being active in the process [7]. This could perhaps be the reason why difficult concepts, principles of operation, procedure of diagnosing faults, repair and maintenance could not be explained/performed effectively by students in many institutions of Nigeria.

In light of the above, a study conducted by [8] revealed that students taught with instructional media perform better than those taught with conventional teaching methods during academic achievement test. It follows that the use of videos during instruction have advantages in teaching as the media, help students retain what they have learnt, hence improve students' academic achievement. As such, use of social media technology like YouTube would assist in addressing most of the problems as well as improving students' academic achievement in tertiary institutions of Nigeria.

Therefore, this study aimed at determining the effect of YouTube-video embedded instruction on students' academic achievement in Automotive Technology Education in tertiary institutions of north-eastern Nigeria.

II. OBJECTIVES OF THE STUDY

The objectives of this study are to:

1. Determine the difference between students' academic achievement in experimental group and that of control group

in automobile technology in the pre-test in tertiary institutions of north-eastern Nigeria.

2. Determine the difference between students' academic achievement in experimental group and that of control group in automobile suspension system in the post-test in tertiary institutions of north-eastern Nigeria.

III. METHODOLOGY

Quasi experimental research design (pretest and posttest design) was employed for this study as this design does not disrupt the original class setting organized by schools there by having the opportunity to use intact class. This is a type of experimental design that does not include random assignment of individuals to groups [9]. According to [10] the design of this study can be illustrated in table I below:

TABLE I. QUASI EXPERIMENTAL (PRETEST AND POSTTEST DESIGN)

Group	Pretest	Treatment	Posttest
Experimental	T_1	X	T_2
Control	T_1		T_2

Where: $\mathbf{T_1}$and $\mathbf{T_2}$: indicate pretest and posttest respectively

X : is the treatment that will be given to the experimental group only

The geographical area of the study was North eastern region of Nigeria where the tertiary institutions involved in the study are located.

IV. SAMPLE AND SAMPLING TECHNIQUE

Simple random sampling technique was used to select the experimental and control groups respectively. This was employed to ensure that each tertiary institution in North eastern Nigeria had an equal chance of being selected. However, all students in each of the two groups selected as the experimental and control group were involved in the study as sample because of the nature of the design of this study. Based on the design, intact class was used so as not to disrupt the existing class setting.

V. INSTRUMENT FOR DATA COLLECTION

Automotive suspension system academic achievement test (ASSAAT) developed by the researcher was used as an instrument for data collection and was validated by three experts in the area. The reliability of ASSAT was established using Kuder-Richardson method (KR-21) where 0.81 reliability coefficient was obtained.

VI. TREATMENT/INTERVENTION

The experimental group was taught using YouTube-Video embedded instruction while the control group was taught using conventional method of teaching (Lecture Method). For the experimental group, the YouTube videos of Automotive Suspension system downloaded from YouTube blog with the consent of the blogger. This YouTube video was used in teaching the topic in a form of instruction called YouTube-Video embedded instruction and used in this study. In the course of applying the treatment, research assistants were selected and act as facilitators where they play the YouTube video to students on the lesson whilst it is paused at every presentation steps to discuss the content with the students so as to make clarifications on the areas that students did not understand or the language/terms that are not clear to them before moving to the next steps. This process was proceeds as such until the end of the instruction. The control group on the other hand were taught using conventional method of teaching. Both the experimental and the control groups were met once a week for a period of 2 hours for 3 weeks and this process took place according to normal class period as schedule in the course time table.Furthermore, the topic: automotive suspension system was selected from curriculum (Nigerian certificate in education minimum standard 2012) in the course TED 214 (Auto braking, suspension and steering system) and was divided into smaller teachable units. Four Lesson plans and one lesson note were developed based on content of the course and the two methods of instructions (YouTube-Video embedded instruction and Lecture Method).

VII. METHOD OF DATA COLLECTION

The data was collected by the researcher with the aid of research assistants.

VIII. METHOD OF DATA ANALYSIS

The statistical tools used for the analysis of data include: mean, standard deviation and independent sample t-test.

IX. RESULTS

The result in Table II showed the difference between students' academic achievement in experimental group and that of control group in automobile suspension system at the pre-test level in tertiary institutions of north-eastern Nigeria. The table revealed that the pre-test mean score of students in the experimental group was 9.82 and standard deviation was 3.40 while the control group obtained the mean score of 10.30 and standard deviation of 3.32. In order to find the magnitude of the difference between these two groups, effect size (ES) was calculated where 0.005 eta square was obtained which indicated a very small difference. This therefore, indicated that there was no significant difference between students' academic achievement in the experimental group and that of the control group in automobile suspension system at this level. This is further confirmed by the independent sample t-test which shows that the p-value is greater than the confidence level (.169>0.05). As a result, the null hypothesis was accepted which revealed that the academic achievement of both experimental and control groups was almost the same at this level.

TABLE II. MEAN, STANDARD DEVIATION AND T-TEST RESULT OF STUDENTS IN BOTH THE EXPERIMENTAL AND THE CONTROL GROUP IN PRE-TEST LEVEL

Group	Test	N	Mean	SD	ES	df	T	p
Exp.	Pretest	162	9.82	3.40				
					.005	36	1.32	.169
Cont.	Posttest	205	10.3	3.32				

Shown in Table III, is the difference between students' academic achievement in experimental group and that of control group in automobile suspension system during the post-test level in tertiary institutions of north-eastern Nigeria. The Table indicates that the mean post-test scores of students in the experimental group was18.79 and standard deviation was 13.92 while the control group mean score was 15.81 and standard deviation was 5.63. In order to find the magnitude of the difference between these two groups, effect size was calculated where 0.02 eta square was obtained. This therefore indicated that there was a significant difference between students' academic achievement in experimental group and that of control group in automobile suspension system in the post-test. This is further confirmed by the independent sample t-test which shows that the p-value is less than the confidence level (.006<0.05). As a result, the null hypothesis was rejected which showed that the academic achievement of students in the experimental group was higher than that of the control group.

TABLE III. MEAN, STANDARD DEVIATION AND INDEPENDENT SAMPLE T-TEST OF STUDENTS IN BOTH THE EXPERIMENTAL AND CONTROL GROUPS IN THE POST-TEST LEVEL

Group	Test	N	Mean	SD	ES	df	t	p
Exp.	Pretest	162	18.79	13.92				
					0.02	36	2.78	.00
Cont.	Posttest	205	15.81	5.63				

X. DISCUSSION OF FINDINGS

The finding in relation to research objective 1 revealed that the students' academic achievement in automotive suspension system in both the experimental and the control groups were almost similar at the pretest level. This was evident in the hypothesis tested in table 2 where it showed that the p-value was greater than the confidence level (0.169>0.05). Consequently, there was no significant difference between students' academic achievement in automotive suspension system in the experimental and that of the control groups in this level. This finding was consistent with[11]who conducted a study on the effect of multimedia technique on technical college students' academic achievement in auto-mechanics in Benue State and reported that there was no significant difference in achievement score of auto mechanic students in the experimental and control groups in the pretest. This was as well in harmony with the findings of [12]who conducted a study and revealed that there was no significant difference in the mean performance of students between the experimental and the control groups in the pretest. Hence, the students of both the groups had nearly the same scores in the pretest. This finding may be attributed to the fact that the knowledge of their initial equivalence (homogeneity variance) was almost the same.

The finding in relation to objective two exhibited that the students' academic achievement in automotive suspension system in the experimental group was higher than that of control group in the post-test. This was confirmed by the research hypothesis tested in table 3 where it revealed that the p-value was less than the confidence level (.006<0.05). This indicated that there was a significant difference between students' academic achievement in experimental group and that of control group in automobile suspension system in the post-test level. This is associated with the finding of [2] who conducted a study on the effect multimedia instruction on students' performance and found out that there was a significant difference in the posttest scores of students between the experimental and the control groups. However, this finding contradicted the finding of the study conducted by[12] which disclosed that there was no significant difference between the mean academic achievement of students taught in the experimental group than that taught in the control group in the post-test but yet showed higher academic achievement in the experimental group than the control group. Another finding that corresponds to the finding of this study was the finding of the study carried out by[6] on the effect of models on students' academic achievement and interest who found out that there was a significant difference between the mean achievement scores of students' in the experimental and the control groups. Based on this finding, it was established that students taught automotive suspension system using YouTube multimedia-enhanced instruction had better academic achievement than students taught using lecture method. This also agreed with findings of [8]on their study that revealed students taught with video assisted instruction tended to be superior than those taught with lecture method.

XI. IMPLICATION OF THE STUDY

The findings of this study will have tremendous implications for automobile technology education teachers, students, researchers and curriculum planners. The implication of this study to automobile technology education teachers is that they will have a new instructional method to teach skills in place of teaching methods that are deficient in learning practical concepts. Another implication of this study is that, with the use of YouTube video-embedded instruction automobile students can have the opportunity to relate practical skills with the theory acquired in classroom and also retain concepts easily which will in turn improve their academic achievements.

Another Implication of this study is that, it will pave a way for other researchers by replicating this research work in various field of study thereby having the opportunity to adapt the use of YouTube video-embedded instruction in learning practical related skills. Lastly, with the adaptation of YouTube video-embedded instruction, curriculum planners will have more insight on the kind of instructional media to incorporate in the learning environment that will help students in relating theory and practical acquired in school with the work environment and also become specialist after graduation.

XII. CONCLUSION

After the findings of this study, it was concluded that there was a significant difference in academic achievement between the students taught with YouTube video-embedded instruction and those taught using conventional methods. Hence, YouTube video-embedded instruction is more credible to use in teaching courses such as automotive suspension system that are abstract in nature than lecture method.

XIII. RECOMMENDATIONS

1. The use of YouTube video-embeddedinstruction should be adopted in teaching suspension system and other automotive courses in tertiary institutions of north-eastern Nigeria so as to improve the academic achievements of students.
2. Federal government of Nigeria should provide more instructional gadgets on YouTube video-embeddedinstruction to support automobile and other TVET teachers in teaching and learning delivery.
3. All automobile teachers should shift away from the use of conventional teaching methods to YouTube video-embeddedrelated instruction in order to lessen the weaknesses associated with conventional method of teaching

XIV. SUGGESTION FOR FURTHER STUDIES

1. Researchers in TVET and other field of study whose curriculum is practical in nature should replicate this study to promote understanding of instructional objectives and improve students' academic achievements.
2. YouTube Video embedded instruction on students' academic achievement and attitude should also be conducted.
3. Researches on YouTube video-embedded related instruction such as daily motion, vidmate and other social media sites should also be conducted to find their effectiveness in teaching and learning

References

[1] A. N Abidiak and A. A Godwin ''Automobile Technology Skills Acquisition and Self-Employment Needs Satisfaction of Technical College Graduates in Akwa Ibom State of Nigeria''. *African Journal of Educational Assessors, vol. 3*, pp.121-130, 2017

[2] M. U. Cyril, "Effects of Multimedia Instruction on Retention and Achievement of Basic Machining Skills in Mechanical Craft Practice". *International Journal of Education and Information Technology,* vol. 2, pp.1-7, 2016

[3] Y. S. Isma'ila, J. D. Enemali, M. H. Muhammad and M. S Nordin, "Students' Psychosocial Perception of Automobile Technology Learning Environment and Attitudinal Outcomes in North-eastern Nigeria Tertiary Institutions". *Procedia-Social and Behavioral Sciences,* vol. 204, pp.136-142, 2015.

[4] A. Rufai, Y. S. Isma'ila, G. A. Gabriel, M. El-Mahmood, N. Sabi'u and D. Adedokun, "Work Skills Required for Employment by Motor Vehicle Mechanics Graduates of Technical Colleges for Sustainable National Development". Paper presented at the 30th National Annual Conference, Bauchi, 2017.

[5] I. P. Ogundola, "*Comparative Effect of Guided and Structured Inquiry Techniques on the Performance of Technical College Students in Motor Vehicle Mechanic Work in Ekiti State",* 2012.

[6] A. O. Oyenuga, "*Effect of Models on Interest and Academic Achievement of Auto-Mechanics Students in Technical Colleges in Lagos-State",* 2016.

[7] A. Ahmad, M. K. Nordin, D.F. Ali, D. F and A. Nabil, "Conducting Hands-on Task in Vocational Education: Teaching Method in Automotive Courses". *Journal of Technical Education and Training,* vol. 7*, pp.* 23-34, 2015

[8] A. U. Nwanekezi and N. E. Kalu, "Effect of Multimedia on Primary School Pupils' Retention and Interest in Basic Science Concepts". *An International Multidisciplinary Journal, Ethiopia,* vol. 6, pp. 206-214, 2012.

[9] J. W. Creswell, "Research Design: Quantitative, Qualitative and Mixed Method Approaches". United Kingdom: SAGE Publications Ltd, 2014.

[10] J. D. Enemali, "*Education for Training and Tndustrialization"* Ibadan: Stirling-Horden Publishers Ltd, 2010

[11] A. D. Tyav, "*Effect of Multimedia Technique on Technical College Students' Achievement in Auto-Mechanics in Benue State, Nigeria".* (M. Ed), University of Nigeria Nsukka, 2016.

[12] S. A. Olufemi, "*Effect of Cognitive mind mapping on the Achievementof Electrical and Electronic Trades Students in Technical Colleges in Ogun State".* (M. Ed Thesis), University of Nigeria, Nsukka, 2013.

honeyCAPTCHA: AN ENHANCED INTRUSION DETECTION MODEL

M. Abdullahi,
Department of Computer Science
Ahmadu Bello University
Zaria- Nigeria
abnaum03@gmail.com

S. Aliyu,
Department of Computer Science
Ahmadu Bello University
Zaria- Nigeria
salis_001@yahoo.com

S. B. Junaidu
Department of Computer Science
Ahmadu Bello University
Zaria- Nigeria
abuyusra@gmail.com

Abstract - In this paper, we proposed a honeyCAPTCHA Intrusion Detection System capable of detecting both known and unknown intelligent and non-intelligent bots. The system was designed as an alternative to the CAPTCHA-BASED IDS which suffers some setbacks due to the treatment of CAPTCHA as IDS which is a popular IPS that creates some level of misclassification of bots and human. CAPTCHA BASED IDS, after evaluated with some prominent IDS metrics gave a Detection Rate of 0.74 and False Positive Rate of 0.36. The proposed model seeks to eliminate CAPTCHA entirely from the user interface by replacing it with an obscured field using JavaScript that will redirect the detected bots to a honeyCAPTCHA which then classifiesit as intelligent when the CAPTCHA is solved, otherwise non-intelligent. Our proposed model will be measured using Intrusion Detection Capability (CID), an intuitive and appealing metric that is grounded by information theory, in addition to Detection rate and False Positive rate used for the CAPTCHA BASED IDS.

Keywords—Intrusion Detection System (IDS), Intrusion Prevention System (IPS), Completely Automated Turing Test to Tell Computer and Human Apart (CAPTCHA), Honeypot, Intrusion Detection Capability (CID).

I. INTRODUCTION

Computer security is a field in IT that focuses on the protection of both computer hardware and software resource. Internet as a major tool in IT needs to be secure for its impact on our daily life. Hence, security on the internet becomes an appealing area of research. There is no doubt that Artificial Intelligent comes with a huge success in our technological advancement. Although, with the abuse of Artificial Intelligent technology, it is becoming a monster which stand to be a threat to its development. The situation of cyber-attack in the past was of concern for only offices or the Government, but today, cyber-attack is a general concern to all as it can trigger future war and political instability [1]. There is no defined feature that qualifies anyone to fall a victim of cyber-crime apart from being on the internet. Legitimate users of the internet can be attacked by web-bots in many ways, some of which include: social engineering, Malvertising, ransomware, phishing and spyphishing, malware, sql injection etc. While attacking users of the internet, bots cause a severe harm to the victims ranging from losing victims files, losing computer control, hardware destruction, probably victim's life. About 4.5 million identities were stolen in 2017 approximately more than the internet users. Cyber-criminals will continuously target identities and steal credential of internet users in 2018 [1].

Some of the measures used to technically protect the computer and internet resources against those attacks include:

- Authentication: ensuring that users and computer are who they claimed to be by establishing proof of identity.
- Encryption: a process of encoding a message or information in such a way that only authorized parties can access it and those who are not authorized cannot.
- Firewall: a network security system that monitors and controls all incoming and outgoing network traffic base on advanced and defined set of security rules.
- Intrusion Detection System (IDS): a process used to identify intrusion. It is classified into one of two methodologies: anomaly detection or misuse detection.

- Intrusion Prevention System: an advance IDS that is capable of preventing a known detection intrusion. One of the prominent IPS is CAPTCHA

- Honeypot: a decoy and sacrificed system that is used to lure the intruders, monitor and log their activities. Honeypot detect both known and unknown intrusions

The CAPTCHA BASED Intrusion detection model was designed by [11] with a dummy CAPTCHA. The system works in a way that a CAPTCHA will be displayed to the user with an instruction written like "DO NOT TYPE ANYTHING IN THIS TEXTBOX", human user is tasked to read, comprehend and abide by the instruction. While an intelligent bot which targets the system will solve the CAPTCHA and pass in with the zeal of bypassing the security. The aim of their research is to identify those possible intrusions from a spyware capable of breaking the CAPTCHA. The design considers CAPTCHA a popular IPS as an IDS, to lure an intelligent spyware into a dummy page and which then retrieves their IP address for identification and subsequently blocking them from using the system.

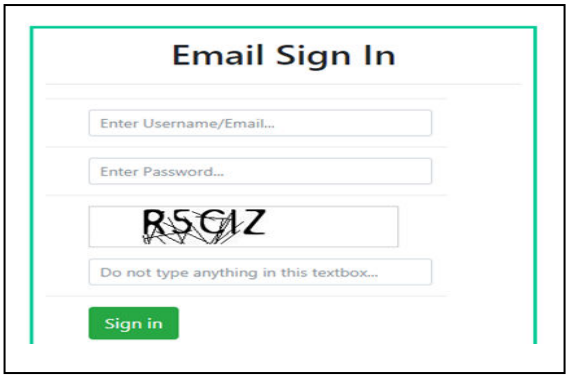


Figure 1: Login Page for the CAPTCHA BASED IDS [11]

CAPTCHA BASED IDS was evaluated with the most basic and commonly known IDS evaluation metrics [12]. The two major metrics used, are the Detection rate, False Positive rate. In their study, it was found that the detection rate (DR) is 0.74 while the False positive rate is 0.34 which is considered too alarming.

Despite efforts to combine Honeypot and CAPTCHA to tackle security challenges, some limitations associated with their works include:

- It is language dependent, which may discourage users and hence, increasing error rates.
- CAPTCHA is usually disliked that most users don't want to see them.
- Increase false positive alerts. If a user mistakenly tries to fill in the text field, it will be considered a robot due to that interaction with the system.
- CAPTCHA, though an IPS, was used as an IDS. Intruders that ignore them can be authenticated as a genuine user (false negative).
- Only intelligent bots are considered for retrieval in their system, underrating the non-intelligent bots.
- Detection rate and False positive rate might not adequately proof better IDS between compared IDS's.
- In general, the trick can be easily uncovered and that may cause more harm to the system.

In this work, we propose a model that addresses the above limitations.

II. RELATED WORK

[2] reviewed some of the traditional IDS that were used previously which include; Haystack developed to help Air force security officers detect misuse of the Unisys 1100/2200 mainframe used at air force base routine for routine unclassified but sensitive data processing and the MIDAS that provides a real time intrusion and misuse detection for the national computer security center network mainframe, Dockmaster, a honeywell DPS 8/70 multics computer system etc.

But generally, traditional IDS are passive in such a way that they only report the detected anomalies or misuse to an administrator or some preventive techniques like, firewalls for update prevention. Traditional IDS focus mainly on how to detect attacks not to prevent them [3]

[4] reviewed most of the soft computing techniques used in IDS development. They comprise the biologically-inspired techniques like the genetic algorithm and some major machine learning tools like Fuzzy logic, support vector, artificial neural network. On their conclusion, they emphasized that soft techniques application in IDS and IPS will optimally improve systems' security.

The passiveness of IDS and the increase in false alerts brings the suggestion on how to improve the state of security. [4] Suggested the combination of host IDS and Network IDS to overcome their challenges though the architecture will be complicated.

[5] sees Intrusion Prevention Systems (IPS) as more advanced version of Intrusion Detection Systems that provides powerful protection by blocking intrusion attempts. IPS in general is considered IDS with advance response and prevention.

[6] Reviewed different CAPTCHAs and categorized them into three i.e. the visual, non-visual and hybrid. After their evaluation, they suggested some alternatives base on some given criteria that will be considered when prioritizing the selection and implementation of CAPTCHAs. The criteria consist of the cost, efficiency, robustness and usability. They identified security and usability as the major barrier in CAPTCHA deployment. The suggested alternatives are: Administrative, Interactive and cheating bots which were tested by the same criteria with the CAPCTHA and found an improvement in both the usability and security of their system.

There is no single solution that will be used to catch all type of spam. So, using multiple methods will help since spammers are creative in getting around spam blocking tools [7]

After a comparative analysis of different CAPTCHAs and their alternative, [8] arrived at a conclusion with a suggestion that, honeypots and CAPTCHA have their respective weaknesses and drawbacks that makes them independently less effective, but, by integrating and removing the weaknesses, they will form a viable defense.

Honeypots stands to be an outstanding technique which performs optimally in detecting both known and unknown bots. There are many ways that CAPTCHA and honeypots can be used simultaneously to boost system security.

[9] Used Entropy and Bayesian Classifiers to classify Chat bots on one of the most popular and large commercial chat networks, Yahoo! Chat. The entropy classifier is to identify new chat bots and then, add them to the Chat bot corpus based on certain metrics: message sizes and inter message delay. While, the Bayesian use the bots and human corpora to learn text pattern of bots and humans and then quickly classify bots base on this pattern**.**

[10] used Naïve Bayes to classify email spam tested on a different dataset, where they concluded that the algorithms quality performance is based on the dataset used. Dataset with few instances of email and attributes gives good performance for the Naïve Bayes Classifiers. The performance on the datasets is evaluated based on accuracy, recall, precision and F-measure. They concluded that the Naive Bayes classifier gets the highest precision in the percentage of spam messages blocked, especially when the dataset is collected from a single e-mail account.

III. THE PROPOSED MODEL

honeyCAPTCHA will be designed to ease user access to the verifiable resources in the internet at the same time increasing the security robustness of the system. In that regard, the proposed system comes with a unique future of allowing a genuine user to have access to the resources by only supplying his username and password without any further verification, while on the other hand obscuring access to a detected attacker. This happens through the modification of the existing system by removing the CAPTCHA test entirely from the user's view, while a hidden detector will redirect the attacker to a fruitless and cognitive honeyCAPTCHA. This is to lure the attacker away from the actual system resources. After the design of the system, it will be implemented on a real website to attract visitors from all over the world. This is to identify the uniqueness and a global treatment of the system. The system robustness will be considered best on the confusion metrics of the IP addresses of the visitors taken from the two systems which will be relatively compared to summarize the performance of our system. Although, the existing system only considers collecting the intelligent bots capable of breaking the CAPTCHA test. Our system will in addition, consider non-intelligent bots. We also consider how robust our system is, as compared to that of the existing system by considering the percentage of the bots' penetration of our system to that of the existing system.

The Intrusion Detection Capability (CID) metricwas also introduced to measure the usability and the security of the system.

IV. EVALUATION METRICS

The system will be measured using the ranking evaluation metrics of IDS such as the Detection rate and the False positive rate that stands to be the most popular evaluation metrics used in IDS. Another IDS

metrics that is unique and comprises almost all the known evaluation criteria, grounded in Information Theory known as the Intrusion Detection Capability (CID) will be used in the evaluation of our system. This is to bridge the gap that may possibly be created by using the above-mentioned ranking metrics.

Those metrics are pulled out from the popular confusion matrix table, which is a matrix that always represents the true and false result of classification.

Actual	**Predicted**	
	Attack	Normal
Attack	TP	FN
Normal	FP	TN

A vivid explanation of the metrics to be used is as follows:

- Detection rate (DR) is the ratio between the number of correctly detected attacks and the number of the attacks

$$Detection\ rate\ (DR) = \frac{Correctly\ detected\ attacks}{Number\ of\ attacks}$$

$$DR=\frac{TP}{TP+FN} \quad (1)$$

- False positive rate (FPR): It is the ratio between the number of normal instances detected as attack and the total number of normal instances.

$$FPR = \frac{Number\ of\ normal\ instances\ detected\ as\ attacks}{Total\ number\ of\ normal\ instances}$$

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

- Intrusion Detection Capability (CID): Is simply the ratio of mutual information between the IDS Input and Output to the entropy of the Input. The mutual information measures the amount of uncertainty of the input resolved by knowing the IDS output, which is normalize by the original uncertainty of the input (entropy). . All the previous metrics like TP, FP, PPV, NPV and B are incorporated in the unified Intrusion Detection Capability (CID).

$$CID = \frac{I(X;Y)}{H(X)}$$

$$CID = \frac{H(X) - H(X|Y)}{H(X)} \quad (3)$$

All values of ranking metrics lie in the range from 0 to 1 [12]. The higher the value, the better is the IDS n terms of CID and DR, where as the FPR is the reverse.

REFERENCES

[1]. Joseph, C. (2018). Symantec Internet Security Threat Report 2018: The Top Takeaways [Blogpost]. Retrieved from: https://thycotic.com/company/blog/2018/04/symantec-internet-security-threat-report-2018/ It is a snapshot of the page as itappeared on 26 Aug 2018,06:22:50 GMT

[2]. B., Mukherjee, L., Heberlein, K., Levitt, (1994) Network Intrusion Detection, IEEE network 1994-Ieeexplore.ieee.org

[3]. Mohammad A. F. & Syed S. H.(2010). Towards Cyber Defense: Research in IntrusionDetection andIntrusion Prevention Systems, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.7

[4]. Chirag, M., Dhiren P., Hiren P., Bhavesh B., Avi P., Muttukrishnan R. (2013). A survey of intrusion detection techniques in Cloud,Centre for Cyber Security Sciences, City University London EC1V 0HB Email: R.Muttukrishnan@city.ac.uk

[5]. Eslam M. H., Aida O. A., & Ahmed I. S. (2010). A hybrid intrusion prevention system (hips) for web database security.International Journal of Engineering Science and Technology, Vol. 2 (7), 2010,2745-2762

[6]. Mohammad, M., & MohammadReza, K. (2014). CAPTCHA and its Alternatives: A Review. SECURITY AND COMMUNICATION NETWORKS Security Comm. Networks 2015; 8:2135–2156.

[7]. Kyrnin, J. (2018). 6 Modern Solutions to Protect Web Forms from Spam,LIFEWIRE. Retrieved from: *https://www.lifewire.com › How To › Web Design &Dev › Basics*

[8]. Parita, C., CHintan, T., & Manish, S.(2016), A Review Paper on Analysis of Decisive and Non-intrusive Technique to Combat Form Spam, *International Journal ofInnovative Research in Computer and Communication Engineering (An ISO 3297:2007 Certified Organization)* Vol.4, Issue 3.

[9]. Steven G., Mengjun X., Zhenyu Wu., & Haining W. (2011). Humans and Bots in Internet Chat:Measurement, Analysis, and Automated Classification. *, IEEE/ACM TRANSACTIONS ON NETWORKING*, , VOL. 19, NO. 5.

[10]. Rusland, N., F., Norfaradilla Wahid, N., Kasim, S. & Hafit, H. (2017). Analysis of Naive Bayes Algorithm for Email Spam. *International Research and Innovation Summit*

[11]. Boukare, S.,&Abubakar,H.(2018). Acaptcha – based intrusion detection model**.**International Journal of Software Engineering & Applications (IJSEA), Vol.9, No.1. Chat: Measurement, Analysis, and Automated Classification. IEEE/ACM

[12]. Gulshan Kumar. (2014) Evaluation Metrics for Intrusion Detection Systems - A Study. A n International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 11, November-2014, pg.11-17. ISSN: 2321-8363

Blockchain: Panacea for Corrupt Practices in Developing Countries

Abdurrashid Ibrahim Sanka and Ray C.C.Cheung
Department of Electronics Engineering
City University of Hong Kong
Hong Kong SAR, China
iasanka2-c@my.cityu.edu.hk, r.cheung@cityu.edu.hk

***Abstract*—Blockchain technology is one of the most exciting emerging technologies and has the capability to digitally revolutionize many systems for better security, autonomy, transparency, auditability, speed, reduced cost and efficiency. Blockchain can be regarded as a revolutionized distributed ledger technology which uses temper proof database in conjunction with consensus algorithms and cryptography. On the other hand, corruption has been a major problem in developing countries like Nigeria that bedevils their progress and development. In this paper, we investigate corrupt practices in developing countries with Nigeria as a case study. We study some blockchain use cases in some countries like Estonia and Georgia. We propose a blockchain framework to curb corrupt practices related to public funds embezzlement as the biggest corruption act. All transactions made by government agencies are stored in the blockchain shared by anti-corruption agencies, police, judiciaries and other related offices at the same time. We also propose the blockchain architecture and consensus suitable for the framework. Blockchain technology is immune to the influence of central authorities and intermediaries. It is secure against tempering and provides transparency and auditability. Hence using blockchain technology will help to tackle corrupt practices for betterment of developing countries such as Nigeria.**

Keywords—Blockchain, Cryptography, Consensus, Corrupt practices, and Developing Countries.

I. INTRODUCTION

Blockchain technology is a disruptive emerging technology that can revolutionize many aspects of our lives and many industries. It provides data security, autonomy, transparency, auditability, speed, reduced cost and efficiency to systems [1]– [4]. The technology emerged from Bitcoin as its first use case. Bitcoin is a peer to peer payment system invented using the blockchain by Satoshi Nakamoto in 2008 in his trial to solve the Europe's economic crises [5]. Bitcoin is now the most successful cryptocurrency. Beside the cryptocurrencies, there are several other applications of blockchain in healthcare, banking, insurance, identity management and so on [6]. There are also increasing blockchain adoptions by many companies such as IBM, Oracle and Microsoft as well as many countries such as Estonia, Georgia, Russia, UK and Singapore. United Kingdom government office for science released a report describing the blockchain as highly disruptive and urges the UK government to invest in the technology for its immense benefits. The report testified that blockchain secures data and provides transparency and traceability which are so beneficial for public services [7]. Blockchain removes central authorities and middle parties thus, it is very suitable for removing corruption in systems with much human interference.

On the other hand, developing countries such as Nigeria have been blessed with various human and natural resources that can be well harnessed to solve the problems faced by the countries. These countries have the potential to grow, develop and even compete with some of the developed nations. However, incessant and massive acts of corruption from different angles hamper and impede the growth of these nations. The corrupt practices are predominantly in the government sectors and become norm in civil services. However, some corrupt practices also come from marketers as well as the civilians.

Annually, corruption causes African countries to lose $140 billion through few corrupt individuals [8]. More than US$400 billion was looted from Nigeria alone by corrupt leaders. This amount was said to be six times in value over the expenses of the resources used to rebuild western Europe after World War II [9]. Unemployment rate has been raising due to the corrupt attitude of the leaders to secure jobs for the vehement citizens. Many government officials have piled ghost workers (fake workers) receiving their illegal salaries, denying offers for legitimate workers and affecting the efficiency of the work space. Electoral corruption deprived people their right to vote for the leaders they want. Results are being tampered by thugs or the corrupt electoral officers.

"We must fight corruption before corruption kills us" is a popular quote of the current president of Nigeria [10]. In 2016, the former Britain's prime minister, David Cameron described Nigeria and Afghanistan as 'fantastically corrupt' nations [11]. Despite efforts made by several governments, the fight against corruption in these countries seems futile, yielding no or ineffective results. Hence there is urgent need to embrace information technology (IT) to tackle corrupt practices for the development of corrupt countries such as Nigeria.

Current solutions failed because of the failure to evict the massive influence of humans. Blockchain technology is immune to the influence of central authorities and intermediaries. It secures data against tempering and provides transparency and auditability. Therefore, using blockchain will help to solve or minimize corrupt practices for betterment of developing the countries like Nigeria. In this paper, we investigate corrupt practices in Nigerian public service as a case study. We propose the use of the powerful emerging technology 'Blockchain Technology' to tackle

corruption in these countries. We propose a blockchain framework for tackling public funds embezzlement which is the biggest corruption related to many corrupt practices such as bribery and illegal award of contracts. We propose the blockchain structure and consensus for the framework. Our contributions are hence summarized as follows:

- We investigate corrupt practices in developing countries with Nigeria as a case study.
- We review blockchain technology.
- We propose the use of blockchain technology to tackle corruption. A blockchain framework was hence proposed for tackling public funds embezzlement which the biggest corruption related to many corrupt practices.
- We propose the blockchain structure and consensus for the framework.

The rest of the paper is organized as follows: Section II gives the background of blockchain technology and related work. Section III discusses blockchain use cases in countries other than Nigeria. Section IV is the proposed framework while conclusion is made in section V.

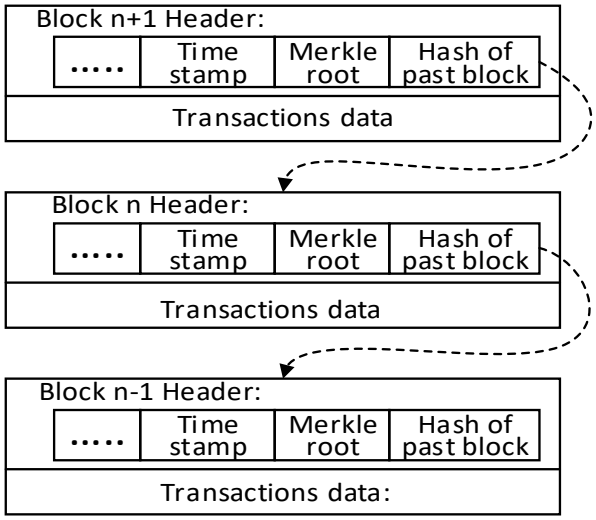
II. BACKGROUND OF BLOCKCHAIN TECHNOLOGY AND RELATED WORKS

A. Blockchain Structure

Blockchain is a distributed ledger or database secured against tempering and modification. Once a record gets into the ledger no one can change or temper with it. The records in blockchain are connected and stored on computers of the participants in the blockchain network (nodes) in consecutive order of data blocks . Each block contains the hash of the previous block. Hence, modification of the data is readily detected since the hash value will change from the stored value in the next block. An adversary has to modify all the records (thousands of blocks) of the computer and also the records in most of the computers on the network which is almost impossible with sufficient number of participants. Transaction is another term referring to the unit of record. The records in blockchain are usually termed as transactions.

The nodes on blockchain network update and maintain the ledger following the consensus agreement defined by the consensus algorithm of the network. The consensus algorithms define how the nodes reach agreement on who to create new blocks among other rules related to the blockchain network. Example of consensus algorithms include the proof of work (PoW), proof of stake(PoS) and Byzantine fault tolerant (BFT) consensus. Physical identities such as names, photos or passport numbers are not used in blockchain for privacy protection. Therefore hashed and scrambled digits are used as blockchain addresses serving as the identity of the users. Blockchain technology underpins cryptocurrencies such as Bitcoin, Ethereum and several other applications. It works without central authorities and intermediaries. It secures data using cryptography and the consensus [12]–[14].

Fig. 1 describes the structure of blockchain as in Bitcoin. Each block consists of the block header followed by transactions data. The block header contains the previous



block hash, Merkle root, timestamp, Merkle root of the transactions among others [6], [15].

B. Benefits and Applications of Blockchain

Blockchain technology comes with many benefits. Because of these benefits, many enterprises want to use blockchain for improving their systems. R3 is an

Fig. 7. Bitcoin blockchain structure

international consortium consisting of over 200 firms most of which are banks. R3 leverages blockchain technology and use it for enhancing business enterprises. Corda is the blockchain platform used it to enhance businesses [16], [17]. Blockchain provides data security, immutable records, privacy, speed and efficiency. It saves costs of systems as well as provides traceability, auditability and transparency. It removes central authorities and intermediaries thus allowing systems to be autonomous and decentralized [18].

Nowadays, there are increasing applications of blockchain with cryptocurrencies such as Bitcoin and Ethereum as the most prominent. Other applications are also found in non-cryptocurrency systems. These include the healthcare, smart contracts, insurance, banking, identity management, land registry, Internet of things (IoT), supply chain, voting systems and more [19]–[21]. Blockchain has been used in big companies such as IBM, Oracle, and Microsoft. It is also used by Gem and Healthbank for illegal drugs protection. Bitshares, NASDAQ and Coinsetters have been using blockchain for stock marketing. Diamond certification records is been recorded on blockchain by Everledger [21], [22].

C. Cryptography used in Blockchain

Cryptography and consensus algorithms are the backbone of blockchain technology. Satoshi Nakamoto intelligently utilised existing cryptographic principles such as the Merkle root and digital signatures for blockchain which was first used in Bitcoin. Those principles are maintained for other blockchain instances [23]. Blockchain technology uses hash functions such as SHA-256 which is the major function used in the proof of work mining. The SHA-256 hash function is also used to create and sign transactions in conjunction with digital signature schemes. RACE Integrity

Primitives Evaluation Message Digest (RIPEMD160) hash function is used together with the SHA-256 hash function to create addresses in Bitcoin. Blockchain also uses the Elliptic curve digital signature algorithm (ECDSA) to create, sign and verify transactions. Merkle tree is used to construct the hash of all transactions in a block (Merkle root) [24], [25].

D. Types of blockchain

Blockchain is classified as public, private and consortium [26]. The Public blockchain is regarded as permissionless while the private and consortium blockchain are grouped as permissioned blockchain [14].

1) Public (Permissionless) blockchain:

Any person (node) in the world can join public blockchains such as Bitcoin. The participants enjoy equal right in the consensus agreement. All the participants can read, verify and add new valid blocks. Bitcoin is an example of public blockchain. In public blockchain, the nodes have to compete in order to be eligible for adding new block (mining). Issues in public blockchain include privacy, 51%attack and selfish mining [12], [27].

2) Private (permissioned) blockchain:

Unlike the public blockchain, the participants of private blockchain are known and selected. Only the authorized user is allowed to join, read, verify, add new blocks or participate in the consensus agreement. This blockchain is centralized and mostly used within single organizations and companies like banks for accountability and more security. Private blockchain have lesser issues and threats compared to the public one [27], [12].

3) Consortium (Permissioned) blockchain:

Consortium blockchain is partially centralized. In this kind of blockchain, some of the users (nodes) are pre-selected to validate new blocks as well as participate in the consensus agreement. The rest of the nodes permissioned to join the network are only allowed read permission. A typical example is a network of say 20 nodes but some selected 5 nodes must sign each block for it to be valid. Consortium blockchain is used by group of organizations acting as a consortium. Examples of consortium blockchain are R3 Corda and Hyperledger [12], [26], [27].

E. Hyperledger Fabric

Hyperledger Fabric is open source blockchain platform initially contributed by IBM and Digital Asset companies in 2015 to the Hyperledger project under Linux foundation. It is a modular, consortium blockchain that allows business enterprises share data in form of transactions. It supports the creation and running of blockchain networks for better performance, reduced costs, security and scalability. Research conducted at IBM shows that Hyperledger Fabric can support over 3500 transactions per seconds unlike Bitcoin which supports 3-4 transactions per second [28], [29].

F. Related Works

Several anti-corruption agencies, measures and programs were set up by many countries, non-governmental organisations (NGOs), African Union (AN), International Monetary Fund (IMF) and united nation (UN). In Nigeria, the Economic and Financial Crime Commission (EFCC) and the Independent Corrupt Practices and Other Related Offenses Commission (ICPC) were set up to curb corruption in the countries. However due to large human influence in the activities of these anti-corrupt agencies, the efficacy at which the corruption is fought is always undermined. Using blockchain technology to run some of the anti-corruption programs and systems will drastically reduce the human influence and make the system more efficiency and reliable. Imagine that EFCC, ICPC CBN, Police, NGOs all share the same blockchain database containing all the transactions carried by government agencies and parastatals. With the blockchain database they trace and audit any transaction while the blockchain does not allow alteration of data.

Large body of literature exists on corruption. Ijewereme [30] review corruption and its harmful effects on the progress of Nigeria. The paper related corruption in Nigeria to many factors such as tribalism, pressure by the society and nepotism. Some measures were recommended for combating the corruption in the country. Obuah [8] reviews corruption in Nigeria and efforts put by the government of Nigeria to tackle the corruption. The activities of EFCC was also examined. Hope Sr. [31] studies corruption in Africa with focus on Swaziland, Nigeria and Kenya. Other related works include [32]–[34]. In contrast, these bunch of studies do not use technological approach in tackling the corrupt practices. Our paper propose the use of blockchain technology which has much potential in tackling some of the corrupt practices when appropriately used.

Adamu et al. [35] suggested posting the financial record of government agencies on websites to curve corruption in the government agencies. However, the management and control of the agencies have the full control of the websites, therefore they can censor what they want as well as edit or delete when they wish. In contrast, we propose blockchain technology in which records cannot be changed and is autonomous that is to say the agencies do not have full control of the data. Kshetri et al [36] mentioned (in few lines) in IEEE magazine that developing countries could benefit from blockchain technology to fight corruption. However no detailed plan or procedure was given. In contrast, we study the corrupt practices and provide a detailed description of using blockchain to tackle the embezzlement corruption. We propose a detailed blockchain framework, its structure and consensus for fighting the corruption.

III. BLOCKCHAIN USE CASES IN OTHER COUNTRIES

Currently, there are several blockchain use cases from many countries. Singapore's government has been using the blockchain to secure banks against invoice fraud. This prevents customers to duplicate invoices as it happened when almost $200 million were lost by the Standard Chartered due to the same act [37].

Georgia is the first country that stores land titles on Bitcoin blockchain. Since 2016, Bitfurry has been working with Georgia for the blockchain public land registry. The project was expanded in 2017 after successful tests to enable other land services like sells, mortgage, new land title and notary [38]. Factom company had aimed to build similar

blockchain land registry for Honduras government. However because of the issues with their existing non-digital land registry, the project seems to be halted [39].

Estonia has been using blockchain since 2012. In March, 2016 the Estonian government uploaded its 1 million health care records on blockchain [40]. More important use case is their blockchain integrated e-Identity. Every Estonian has a digital identity and can use digital signature to access public e-services. Nowadays Estonia uses blockchain for their registries, health, legislative, commercial and security systems. Keyless Signature Infrastructure (KSI) is an Estonian blockchain based integrity monitoring technology against network advanced persistent threats (APTs) and developed by Guardtime (Estonian company) for network security. In September 2016, the American Defence Advanced Research Projects Agency (DARPA) awarded $1.8 million contract to Guardtime and Galois to verify the KSI system [41]. Other countries using blockchain include Russia, USA, UK, Sweden and Australia.

IV. Corrupt Practices In Developing Countries

Corruption is the major setback that impedes developments in developing countries mostly African nnations such as Nigeria and some Asian countries such as India. Although there are variety of definitions to corruption, it is defined as illegal use of someone's power or position for personal gain (to himself or others) [30], [42]. Corruption deters development by deviating resources, projects, power and capital from its real beneficiaries (mostly the poor) to the few illegitimate individuals. It undermines real investments and businesses. Corrupt officers loot money making infrastructural developments infeasible and subjecting the country into economic hardship. Contracts awarded by corruption use to be partially or not implemented. Poor quality infrastructures are built due to the corruption involved in their execution endangering the lives of the people. Corruption also lead to inflation and harbour bad leadership. Poor education, social and economic growth are all harvests of corrupt practices.

Corruption manifests in several forms most of which are spelt by the law and have prescribed sanctions. Corrupt practices in Nigeria are ubiquitous. They include bribery, public funds embezzlement, black mail, nepotism, extortion, electoral corruption, ghost working. Corruption is widely practised in judiciary, land registry, fuel subsidy, Bureau de change (BDCs), contracts awards, budgets, energy sector, education, custom services and immigration. Most government office holders In Nigeria are involved in corruption in one way or the other [8]. Some of these corrupt practices are discussed as follows:

A. Public funds embezzlement and bribery:

Public funds embezzlement and bribery have been the major acts of corruption in developing countries like Nigeria. Bribe is now a common practice across all government agencies and parastatals. In fact, officers who refuse to collude in bribery face various challenges and threats from the other workers due to their lack of interest to cooperate in doing the corrupt acts. Executive offices and their nominees handling the affairs of ministries and agencies loot and spend money recklessly. They allocate fake contracts and take kickbacks from contractors. Inflating contracts and kickback acceptance have become common norm among the political office holders in Nigeria. Recently in the end of 2018 a video clip was released showing a popular governor of a state accepting millions of dollars as kickback for contracts. Even though the governor denied the allegations, the journalist of Daily Nigerian magazine confirmed the authenticity of the video clips [43].

The former secretary to the Nigerian government is currently facing prosecution by EFCC on corruption charge of over 500 million naira one of which involves 258 Million naira contract for grass-cutting [44]. Since 2006, the money looted by the past and present leaders of Nigeria was estimated to 212 billion. Recently in July 2019, the acting EFCC executive chairman, Ibrahim Magu disclosed that Nigeria annually loses more than 50 billion naira to illegal cash flows due to corruption[45]. Public funds embezzlement is the biggest corruption resulting from several forms of corrupt practices. Therefore it must be stopped or minimized to the barest minimum for the development of any country.

B. Ghost workers corruption

Ghost working refers to the practice of inflating the pay role of an institution, agency or ministry with employees that do not exist in reality and then taken away their salaries. This usually happens in collusion with the administrators of the institution [30]. Ghost workers bring a lot of problems to country's development and heavily drains its economy. It brings higher rate of unemployment, encourages corruption and deteriorate the entire work space. With ghost workers, the existing employees tend to be overloaded which reduces their productivity and the system's efficiency. Nigerian government losses over 100 billion naira annually to ghost workers. In 2013, over 45,000 ghost workers were discovered by in the pay roll of Federal government alone. Unfortunately, these cases are still rarely investigated [46].

C. Electoral corruption:

Electoral corruption involves range of corrupt practices done in order to influence the legitimacy of elections. These include bribing electoral officers, votes buying, using thugs to victimise voters or alteration of results at the collation centres. Electoral corruption has been around in Nigerian system since post-independence around 1960s. The Nigerian electoral commission, INEC has been severally accused of involvement in massive election riggings including the just concluded 2019 election [47].

Other forms of corruption are also practised in land ownership, fuel subsidy, budgets, immigration, custom and others.

V. BLOCKCHAIN FRAMEWORK FOR PUBLIC FUNDS EMBEZZLEMENT PREVENTION

The framework proposes the use of a single shared record (blockchain) for keeping the transaction records of all the government agencies, ministries, institutions and other parastatals. The records are shared and seen simultaneously by other institutions that are responsible for fighting, discovering and punishing corrupt practices.

As a disruptive technology, blockchain is capable of addressing the major corrupt practices listed in III. The immunity of the blockchain provides absolute security to

data against tempering. Once the data is written on to the blockchain, no one can change or temper with the record. This feature will facilitate fighting corruption by providing temper proof record so that shared records could not be changed at later time. Traceability and transparency of blockchain will make all transaction records traceable and transparent. All parties involved in the blockchain network could be able to see all transaction as well as trace them. This feature will make auditing of public services easy and transparent. When properly used, blockchain will dispose any government agency doing corruption since all the transactions will be monitored and seeing by other institutions including anti-corruption agencies. Blockchain dispenses with central authority. In blockchain, the effect of central authority is negligible and much minimized. This will reduce corrupt practices by the central and middle men since the records will be added and managed autonomously, censorship and modification of records is prevented. These are some of the advantages of blockchain over other existing shared database that could be used.

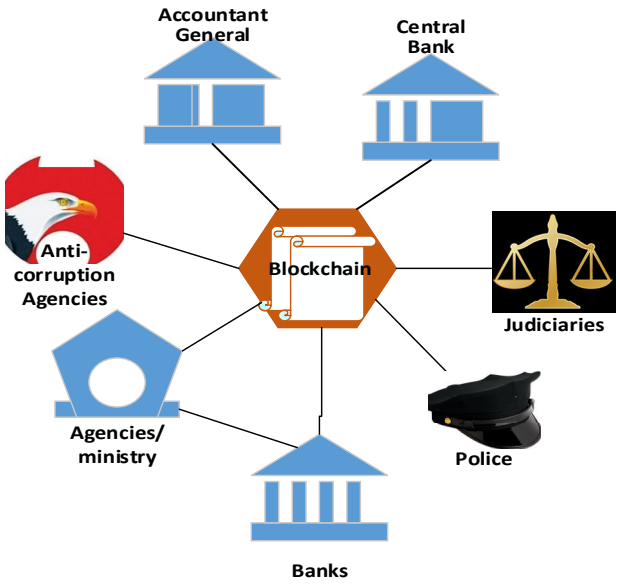
A. The framework architecture

The framework consists of the blockchain shared database at the heart of the system for recording all the transactions made by the government agencies, ministries, institutions and parastatals. The record is shared and monitored by all the participants in this network. Each participant can view, query and trace any transaction he want to audit or check. Remember transactions cannot be modified or deleted once they are recorded in the blockchain. The participants of the network include the central bank of the countries and all anti-corruption agencies. Anti-corruption agencies here comprises of the governmental and non-governmental or international agencies working for fighting corrupt practices. Examples of such agencies in Nigeria are the EFCC and ICPC. Other participants in the network include the police, office of the accountant general of the federation, judiciaries (supreme court) and the banks. Fig. 2 shows the architecture of this framework.

Government agencies put their transaction records onto the blockchain through their banks. For every transaction made by the agencies, the bank creates a record of the transaction and uploads to the blockchain. The actual payment for the transaction is withheld until the transaction record gets into the blockchain. Only the banks create and send the transaction records for the blockchain. The other participants in the network can only check, monitor, trace and participate in the consensus of the network.

B. Proposed Blockchain Type and Structure

We propose the use of a consortium (permissioned) blockchain due to the privacy of the records against the general public. Public blockchain is not suitable in this case because the records are not meant for the general public consumption. Public blockchain also have more scalability, security and consensus issues. In this framework non-proof of work consensus is proposed to avoid the huge energy waste in the proof of work. On the other hand private blockchain is more centralized and suitable for single organization. As our framework consists of various entities from different organizations, consortium blockchain is more suitable as it is less centralized and secured.



Regarding the blockchain structure, we propose the structure shown in Fig. 3. This structure shows the composition of each block in the blockchain. A block consists of block header and transactions data. The header content proposed are the hash of the previous block, timestamp, nonce, Merkle root and the block creator ID. We introduce the block creator ID to identify the node that create each block for traceability. The transaction data proposed also consists of the amount, sender's address, receiver's address, bank's digital signature, bank, date/time and remarks. These are the content in each single transaction. They are selected to give detail information on the transaction for auditing, monitoring and the tracing.

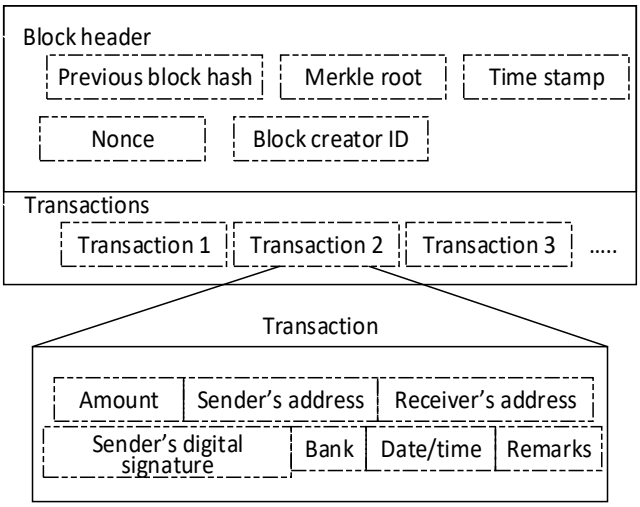
Fig. 8. The framework architecture

Hyperledger Fabric is one of the suitable consortium blockchain that could be used for this purpose. Its block structure could be leveraged and also used for this framework. However if government is capable of creating its own consortium blockchain based on this framework, that may be more preferred.

C. Proposed consensus mechanism

Non-proof of work consensus is proposed for this frame-work to avoid the high energy consumption and other mining issues in the proof of work as used in Bitcoin and Ethereum. Since all the nodes are known in this scenario, consensus mechanisms such as Byzantine Fault Tolerant (BFT), practical BFT (PBFT) or delegated BFT could be used. The Hyperledger fabric blockchain suggested for this framework currently implements a simplified Practical Byzantine Fault Tolerant consensus (SBFT) in addition its Kafka and SOLO ordering services. Kafka is voting-based crash fault tolerant only (CFT) with no tolerance against malicious nodes while SOLO is centralized and hence recommended only for testing. However, the SBFT consensus has both crash fault tolerance and malicious node tolerance (up to 1/3 Byzantine malicious nodes) [6], [48]– [50].

The activity diagram in Fig. 4 describes the consensus in Hyperledger fabric blockchain. It comprises of three steps that is proposal of transactions, ordering service and



validation with commit of blocks into the blockchain. The participants are classified into client, peers, endorsers and ordering services. The consensus goes as follows:

In the first step, client nodes create and sign transaction proposals which get endorsed by some set of peers (network participants) called endorsers. Endorsers endorse transactions by running smart contract which checks the authenticity (signatures and the validity) of the transaction based on the endorsement policy and generates the transaction endorsement as the proposal response. The endorsement executes the transactions and records its results without appending to the blockchain at this time. The client receives the endorsed transaction proposals, packs them as the

Fig. 9. The proposed blockchain structure

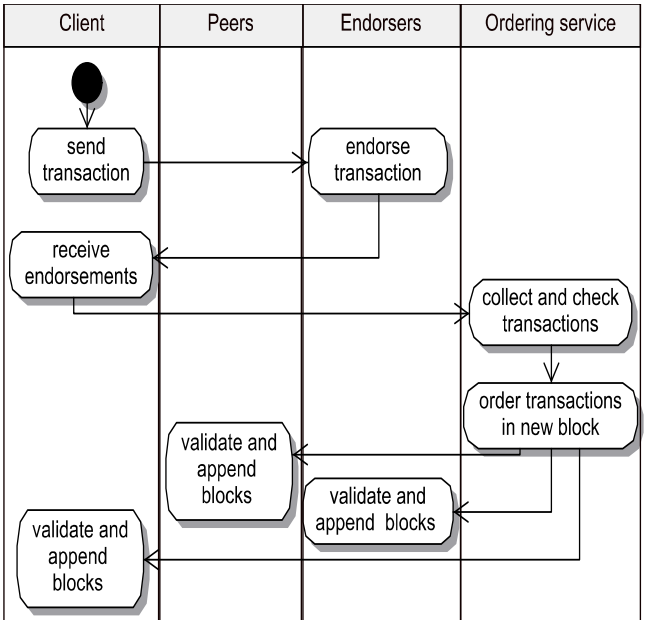


Fig. 10. Consensus in Hyperledger Fabric

endorsed transaction and sends to the ordering service. In the second step, the orderer is the node elected to order, verify, put transactions into new block and broadcast to all the peers for validation and adding to the blockchain in the last step. Finally, the block containing these transactions is added into blockchain as the final record [28], [51], [52].

D. Security of the system

The system is secured against tempering and modification. The nodes and transactions are cryptographically authenticated using digital signatures and access control is guaranteed by the ordering service, hence the framework is secured against masquerade, and man-in-the middle attacks. The underlying assumption for this framework is a secure channel. We assume the participants are secure and also communicate under a secure channel (channels using some security protocols such as SSL and IPsec) to avoid attacks such as denial of service attack (DoS). It is a sole responsibility of the government of the agencies to use secure medium for communications in this framework.

The perceived challenge in the adoption of blockchain technology in developing countries is the provision of reliable and fast internet connectivity. Most developing countries do not have sufficient and reliable internet connection. Therefore, any government that may wish to adopt this framework should provide reliable and fast internet connectivity to the participating entities.

VI. CONCLUSION

Blockchain technology is an emerging technology capable of disrupting many systems and allowing them to improve in terms of data security, speed, efficiency, immutable record, autonomy and reduced cost. On the other hand, corruption has been impeding developing countries such as Nigeria thus deterring them from development. We study corruption in Nigeria as a case study and propose the use of blockchain technology to tackle the corruption. We propose blockchain framework and structure for preventing public funds embezzlement which is the biggest corruption. We believe that blockchain technology can help so much in fighting corruption when harness properly to usher development in developing countries.

REFERENCES

[14] M. Nofer, P. Gomber, O. Hinz, and D. Schiereck, "Blockchain," Business & Information Systems Engineering, vol. 59, no. 3, pp. 183–187, 2017.

[15] M. Iansiti and K. R. Lakhani, "The truth about blockchain," Harvard Business Review, vol. 95, no. 1, pp. 118–127, 2017.

[16] L. Mearian, "What is blockchain? The most disruptive tech in decades," Computerworld, pp. 1–8, 2017.

[17] S. Huh, S. Cho, and S. Kim, "Managing IoT devices using blockchain platform," International Conference on Advanced Communication Tech- nology, ICACT, pp. 464–467, 2017.

[18] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," www.bitcoin.org, p. 9, 2008.

[19] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017, pp. 557–564, 2017.

[20] Government Office for Science, "Distributed Ledger Technology: beyond block chain," tech. rep., 2016.

[21] E. Obuah, "Combating corruption in a failed state: the nigerian economic and financial crimes commission (efcc)," Journal of Sustainable Development in Africa, vol. 12, no. 1, pp. 27–53, 2010.

[22] A. Azeez, "Endangering good governance for sustainable democracy: the continuing struggle against corruption in nigeria," Jourrnal in

Research, Peace, Gender and Development, vol. 1, no. 11, pp. 307–314, 2011.

[23] S. Monjola, “we must kill corruption before it kills us, apc candidate says,” 2015.

[24] BBC, “David cameron calls nigeria and afghanistan ’fantastically corrupt’,” 2016.

[25] I.-C. Lin and T.-C. Liao, “A Survey of Blockchain Security Issues and Challenges,” International Journal of Network Security, vol. 1919, no. 55, pp. 653–659, 2017.

[26] M. Swan, Blockchain : blueprint for a new economy. Cambridge: O’Reilly Media, Incorporated, 1 ed., 2015.

[27] C. Cachin and M. Vukolic´, “Blockchain Consensus Protocols in the Wild,” arXiv preprint arXiv:1707.01873, 2017.

[28] S. S. Gupta, Blockchain. John Wiley & Sons, Inc, 2017.

[29] R. G. Brown, J. Carlyle, I. Grigg, and M. Hearn, “Corda: an introduction,” R3 CEV, August, vol. 1, p. 15, 2016.

[30] R3, “Industry-wide collaboration is in our dna,” https://www.r3.com /about/ , 2019.

[31] L. Cocco, A. Pinna, and M. Marchesi, “Banking on blockchain: Costs savings thanks to the blockchain technology,” Future Internet, vol. 9, no. 3, pp. 1–21, 2017.

[32] D. Romano, G. Schmid, D. Romano, and G. Schmid, “Beyond Bitcoin: A Critical Look at Blockchain-Based Systems,” Cryptography, vol. 1, p. 15, sep 2017.

[33] O. Firica, “Blockchain technology: Promises and realities of the year 2017,” Quality - Access to Success, vol. 18, no. October, pp. 51–58, 2017.

[34] M. Crosby, “BlockChain Technology: Beyond Bitcoin,” Applied Inno- vation Review Issue, no. 2, 2016.

[35] M. Mettler, “Blockchain technology in healthcare: The revolution starts here,” 2016 IEEE 18th International Conference on e-Health Network- ing, Applications and Services, Healthcom 2016, pp. 16–18, 2016.

[36] A. Narayanan and J. Clark, “Bitcoin’s academic pedigree,” Communi- cations of the ACM, vol. 60, no. 12, pp. 36–45, 2017.

[37] D. L. K. Chuen, Handbook of digital currency: Bitcoin, innovation, financial instruments, and big data. Academic Press, 2015.

[38] A. Kaushik, A. Choudhary, C. Ektare, D. Thomas, and S. Akram, “Blockchainliterature survey,” in 2017 2nd IEEE International Confer- ence on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 2145–2148, IEEE, 2017.

[39] V. Buterin, “On Public and Private Blockchains,” 2015.

[40] H. Vranken, “Sustainability of bitcoin and blockchains,” Current Opin- ion in Environmental Sustainability, vol. 28, pp. 1–9, 2017.

[41] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, et al., “Hyperledger fabric: a distributed operating system for permissioned blockchains,” in Proceedings of the Thirteenth EuroSys Conference, p. 30, ACM, 2018.

[42] L. Foundation, “Hyperledger fabric,” 2019.

[43] O. B. Ijewereme, “Anatomy of corruption in the nigerian public sector: Theoretical perspectives and some empirical explanations,” Sage Open, vol. 5, no. 2, p. 2158244015581188, 2015.

[44] K. R. Hope Sr, Corruption and Governance in Africa: Swaziland, Kenya, Nigeria. Springer, 2017.

[45] P. O. Okolo, “Corruption and post 2015 development agenda in nigeria,” Global Journal of Humanities-Social Science: F Political Science, vol. 16, no. 2, pp. 7–18, 2016.

[46] S. B. Werner, “New directions in the study of administrative corruption,” in Classics Of Administrative Ethics, pp. 191–208, Routledge, 2018.

[47] K. Siddiqui, “Corruption and economic mismanagement in developing countries,” The World Financial Review, 2019.

[48] H. Adamu, A. Shatimah, and M. Aliyu, “Role of information and communication technology as a means of tackling corruption in nigeria,” International Journal of Innovative Research and Creative Technology, vol. 1, no. 5, pp. 478–480, 2016.

[49] N. Kshetri, “1 Blockchain’s roles in meeting key supply chain management objectives,” International Journal of Information Management, vol. 39, no. June 2017, pp. 80–89, 2018.

[50] Basu Medha, “Singapore Government builds blockchain system to protect banks,” 2016.

[51] L. Shin, “The First Government To Secure Land Titles On The Bitcoin Blockchain Expands Project,” 2017.

[52] Reason Foundation, “Can Blockchain Technology Reduce Third-World Poverty? - Hit & Run : Reason.com,” 2016.

[53] B. insider, “Estonia is using the technology behind bitcoin to secure 1 million health records,” 2016.

[54] Guardtime, “DARPA Contract to Formally Verify Blockchain-based Integrity Monitoring System,” 2018.

[55] S. Carvalho, Engaging with fragile states: an IEG review of World Bank support to low-income countries under stress. The World Bank, 2006.

[56] PremiumTimes, “Updated: Kano governor ganduje caught on video receiving dollars from suspected contractors,” 2018.

[57] Punch, “Grass-cutting scandal: Court adjourns trial of babachir lawal, five others,” 2019.

[58] EFCC, “Nigeria loses 50billion dollars to corruption annually - magu,” 2019.

[59] C. Okekeocha, “A case study of corruption and public accountability in nigeria,” 2013.

[60] O. Samuel, A. Collins, and R. Atama, “Political corruption and nigerias external image: The case of goodluck jonathans administration (2009- 2015),” International Journal of Governance and Development, vol. 6, no. 1, pp. 85–89, 2019.

[61] I. Bashir, Mastering blockchain: Distributed ledger technology, decen- tralization, and smart contracts explained. Packt Publishing Ltd, 2018.

[62] C. Cachin, “Architecture of the hyperledger blockchain fabric,” in Work- shop on distributed cryptocurrencies and consensus ledgers, vol. 310, p. 4, 2016.

[63] E. Androulaki, C. Cachin, A. De Caro, A. Kind, and M. Osborne, “Cryptography and protocols in hyperledger fabric,” in Real-World Cryptography Conference, 2017.

[64] M. Valenta and P. Sandner, “Comparison of ethereum, hyperledger fabric and corda,” [ebook] Frankfurt School, Blockchain Center, 2017.

[65] L. Foundation, “Hyperledger architecture, volume 1.introduction to hyperledger business blockchain design philosophy and consensus,” 2017.

Robust H-Infinity Control for Magnetic Levitation System

Ahmad Abubakar
Gipsa Lab University Grenoble, Alpes University Grenoble, Alpes
Alpes, France
ahmad.abubakar@gipsa-lab.grenoble-inp.fr

Ibrahim Kabiru Dahiru
Gipsa Lab University Grenoble, Alpes University Grenoble, Alpes
Alpes, France
kabiru-dahiru.ibrahim@gipsa-lab.grenoble-inp.fr

Sulaiman Haruna Sulaiman
Department of Electrical Engineering
Ahmadu Bello University
Zaria, Nigeria
shsulaiman@abu.edu.ng

Abdullahi Bala Kunya
Department of Electrical Engineering
Ahmadu Bello University
Zaria, Nigeria
abkunya@abu.edu.ng

***Abstract*— This paper studies Quancer Magnetic levitation plant and proposed a robust control for the system using H-infinity approach. The performance and robustness of the proposed method was compared with the existing Proportional Integral Velocity (PIV) + Feed-forward (FF) control designed by the manufacturer. The performance analysis of the existing PIV+FF controller was made to build the desired specification template. A dynamic state feedback controller satisfying H-infinity performance, that is numerically solved using Riccati approach was designed. Simulation results obtained, shows superior robustness to uncertainties of the proposed control over the existing control with sufficient tracking performance.**

Keywords*— H-Infinity approach, Quancer Magnetic levitation, robustness, state feedback.

I. INTRODUCTION

Magnetic Levitation can be viewed as a system in which an object is levitated without any support with the help of magnetic field. It is operated by overcoming the gravitational force on an object by applying counteracting magnetic field [1].

A Magnetic Levitation (MAGLEV) is a highly advanced technological innovation that captured the interest of engineers due to its vast area of application especially in industries, mechatronics, electrical and electronics, automobiles, electric trains, frictionless bearing, vibration isolation of sensitive machinery, etc. A common point of its application is the lack of contact with any other part, hence, eliminating wearing and tearing away of parts due to friction. These consequently increases the efficiency of the designed system. Maglev systems are currently used in developed countries for applications such as bearing and high-speed trains [2].

Among the recent research in magnetic levitation systems, an effective control approach used in [1] is made up of a proportional-plus-integral-plus-velocity (PIV) plus feedforward action, which suspend the steel ball at equilibrium point by magnetic current control using the measurements of the ball's vertical position. In simulation and experimental validation, the proposed approach which give more emphasis on the tracking performance shows a good tracking of the ball's position within a maximum settling time of 0.9secs and overshoot of 15%, with zero steady state tracking error.

Subsequently, to investigate the effect of disturbance on tracking performance, a Linear Quadratic Regulator (LQR) based control approach was proposed in [3], which is optimally tuned with a Proportional Integral and Derivative (PID) controller. The selection of the weighting matrices is based on the natural frequency and damping ratio of the closed loop. Experimental result obtained proved that the proposed control strategy is effective in stabilizing the ball and also in rejecting the disturbance presence. To further improve the control of magnetic levitation system, a Gain scheduling approach is proposed in [4] using redundant description representation. So that, the steel ball floats stably within some variational range by scheduling an equilibrium point. In simulation, performance comparison of the proposed method and that of [3] was done, results proved an effectiveness of the Gain scheduling approach for large parameter variation of the system.

In [5], the non-linearities present in magnetic levitation system was control with Inertial delay state feedback observer, which gives lumped uncertainty and unavailable states of the system. Effective trajectory tracking was achieved. Along the line, controlling a magnetic levitation system using the Linear Parameter Varying (LPV) approach was proposed [6]. Synthesis and analysis of a feedback tracking controller were performed and compared with parameter schedule observer, via both simulation and experiment. Absolute tracking performance were achieved with robustness to uncertainties.

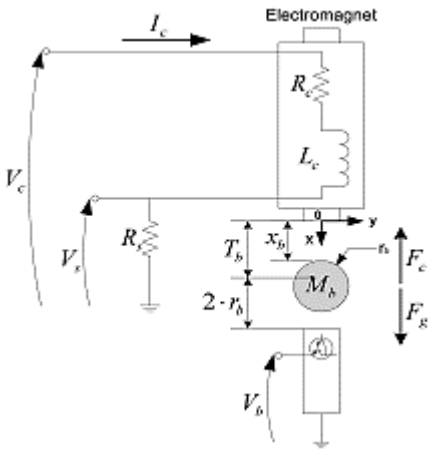


Fig. 1: Quancer Maglev System [1]

Finally, in this work, H-infinity control approach is proposed due to its exceptional robustness [7]. A more experimentally valid Maglev system model in [6] was adopted, which take into account the ball position offset and additional inductance due to ball position. In addition to model parameters in [1]. Synthesis and analysis of the tracking controller used in [1] were performed and compared with the proposed H-infinity control via simulation, to examine the robustness of the responses to uncertainties.

II. SYSTEM DYNAMICS MODEL

The dynamics of the system as demonstrated in fig. 1 is divided into two parts, the electrical and the mechanical part (guided by equations of motion).

MAGLEV SYSTEM PARAMETERS [1]

Symbol	*Description*	*Value*
g	Acceleration due to gravity	$9.81ms^{-2}$
μ_a	Absolute viscosity of air	$18\mu Nsm^{-2}$
M_b	Mass of ball	$18g$
r_b	Radius of ball	$12.7mm$
b_v	Viscous resistance	$4.309\mu Ns^{-1}$
$x_{b\,max}$	Maximum ball travel	$14mm$
R_c	Electromagnet coil resistance	10Ω
R_s	Current sensing resistor	1Ω
L_c	Electromagnet inductance	$412.6mH$
L_b	Additional inductance from the ball position	$31.7mH$
N_c	Number of turns in the coil	2460
r_c	Radius of magnetic core	$8mm$
l_c	Length of coil	$82.6mm$
α	Ball position offset	$10.1mm$

A. Electrical Subsystem (Current Control)

Applying the Kirchhoff's law to the circuit in fig. 1, the sum of series voltages in the loop is used to describe the dynamics of the electrical network. The mathematical model can be described as in [3]:

$$V = i_c R_s + i_c R_c + L_c \frac{di_c}{dt} \quad (1)$$

R_s, R_c, L_s are defined in table 1, while V *and* i_c represent supply voltage and coil current respectively.

B. Equation of Motion (Ball Position)

For the levitation of the ball, magnetic induced force acts in opposite direction to the gravitational force on the ball, hence this can be represented by newton's second law of motion.

Applying the newton's second law of motion, noting that the net force depends on the gravitational and electromagnetic forces, the dynamics of the system is described as in [3]:

$$mg - \frac{L_b \alpha i_c^2}{2x_b^2} = m\ddot{x}_b \quad (2)$$

Where x_b is the ball position.

C. State representation of the Overall Model

The magnetic levitation plant has 3 states given by the position of the ball, the velocity of the ball and the coil current [3].

Let the space variables be defined by $x_1 = x_b$, $x_2 = \dot{x}_b = \dot{x}_1$ and $x_3 = i_c$.
The non-linear model can be described as follows

$$\dot{x}_1 = x_2 \quad (3)$$

$$\dot{x}_2 = g - \frac{L_b \alpha x_3^2}{2m_b x_1^3} \quad (4)$$

$$\dot{x}_3 = -\frac{(R_s + R_c)}{L_c} x_3 + \frac{V}{L_c} \quad (5)$$

Linearized model in matrix form,

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{L_b \alpha x_{30}^2}{2m_b x_{10}^3} x_1 & 0 & -\frac{L_b \alpha x_{30}^2}{m_b x_{10}^2} \\ 0 & 0 & -\frac{(R_s + R_c)}{L_c} \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{L_c} \end{bmatrix} u(t) \quad (6)$$

Where the state vector $x = [x_1, x_2, x_3]^T$ with x_1 representing the ball position, x_2 the ball velocity, x_3 the coil current and $u(t)$ is the control input.

Remark: L_b *and* α are constants found experimentally.

III. EXISTING CONTROLLER

The PIV + feedforward controller was designed for position control of the ball. The control configuration is as shown in fig. 2. The coil current control transfer function is set as 1 assuming that it is not necessary to track the desired current.

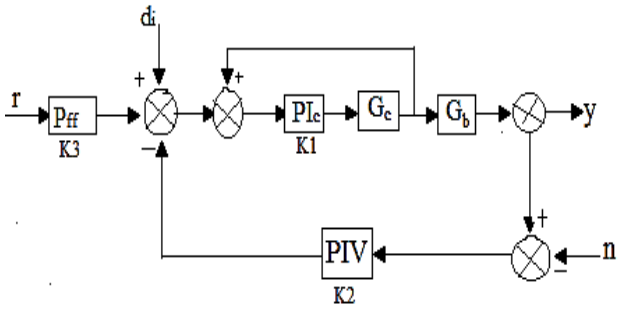


Fig. 2: Closed loop configuration- PIV + feedforward control

The PIV controller was designed with the specification listed below [1]:

- Maximum Overshoot $\leq 20\%$
- Steady state $ess = 0$
- Maximum settling time less than 1sec
- Equilibrium position: $x_{b0} = 6\,mm$

The various controller gains for the PIV controller [1] are $K_p = -252.6, K_i = -248.1$, and $K_v = -4.3$, for the PI current controller [1], $K_p = 789$, $K_i = 23.3$ and for the Feedforward P_{ff} controller, $K_{ff} = 142.9$.

A. Frequency domain Analysis

The sensitivity functions were derived from fig. 2 as follows:

For easy computation, lets $K_1 = \frac{K_i + K_p s}{s}$, $K_2 = \frac{K_i + K_p s - K_v s^2}{s^2}$, $K_3 = K_{ff}$ and $G = G_c G_b$.

Then,

$$S = \frac{1}{1+GK_2} \tag{7}$$

$$T: \begin{cases} T_r = \frac{GK_1 + GK_3}{1+GK_2} \\ T_n = \frac{-GK_2}{1+GK_2} \end{cases} \tag{8}$$

$$KS: \begin{cases} K_{sr} = \frac{K_1 + K_3}{1+GK_2} \\ K_{sn} = \frac{K_2}{1+GK_2} \end{cases} \tag{9}$$

$$SG = \frac{G}{1+GK_2} \tag{10}$$

The sigma plot for the sensitivity functions in (7) to (10) are shown in fig 3.

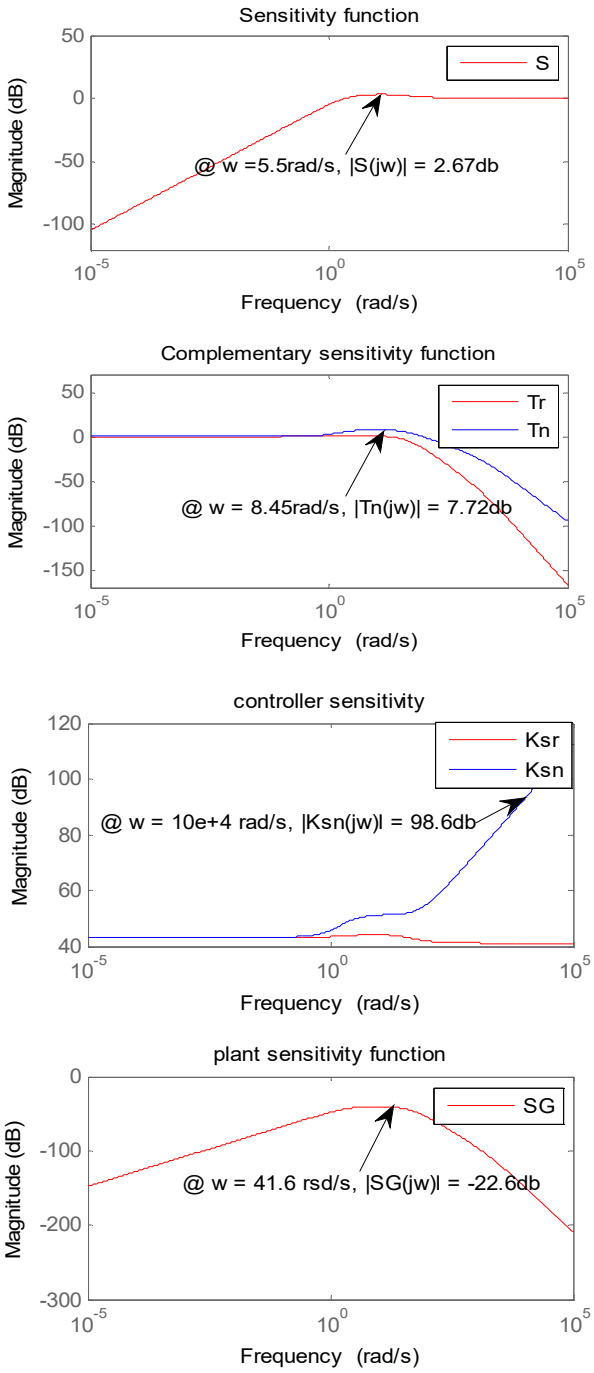


Fig. 3: Sensitivity functions for PIV + feedforward system control (a, b, c and d)

a) Performance Analysis

The position of the ball is viewed as a constant reference signal; thus, low frequency will be our region of interest, such that, the trade-offs will be applied within the frequency region $\omega_s \ll \omega c \ll \omega_T$.

- **S:** Fig .2a shows the response of the ball position to reference position, the system meets the desired performance specifications i.e. At low frequency, $|S(j\omega)|$ is low with a steady state error in tracking of approximately **–105db** (0.1%) at

$w \approx 0\,rad/s$. Thus, it can be concluded that there is a good tracking performance with acceptable error.

- **T:** Fig .2b gives the response of the ball position to noise injected at the output, because noise signals are typically of high frequencies, then, it can be said that the system meet the desired performance specifications i.e. at high frequency, $|T(j\omega)|$ is low. Moreover , $|T_r(j\omega)|_{w=0} \approx 1$ which highlights perfect reference tracking. Thus, measured ball position is less sensitive to uncertainties at output.
- **KS:** Fig. 2c gives the response of the coil current to reference position and noise. the system deviates from the desired performance specifications. i.e. at high frequency, $|KS(j\omega)|$ is high. Where $|KS_r(j\omega)| = 40db$ at all frequencies and $|KS_n(j\omega)|$ increase with increase in frequencies. Thus, the control input is very sensitive to noise.
- **SG:** Fig .2d gives the response of the ball position to disturbance at the input, because disturbances are typically of low frequencies, the system meet the desired specification i.e. at low frequency, $|SG(j\omega)|$ is low. $|SG(j\omega)|_\infty = -22.6db$, which is less than 0db. Thus, it can be concluded that there is an appreciable disturbance attenuation.

b) Robustness Analysis

For the system to be robust to uncertainties, Module margins; $\max|S(j\omega)| \leq 6db$ and $\max|T(j\omega)| \leq 3.5db$ must be met [3]. From fig. (2a) and fig. (2b), It can be seen that, the values of $|S(j\omega)|_\infty = 2.67db$ and $|T_n(j\omega)|_\infty = 7.72db$ respectively. Thus, the system robustness is not guaranteed because one of the module margins is not satisfied.

B. Time domain Analysis

In this section, simulations scenario with reference to the ball position of 6mm from the coil is performed on the non-linear system. Fig. 4 illustrates the action of the PIV + feedforward control design on the system response.

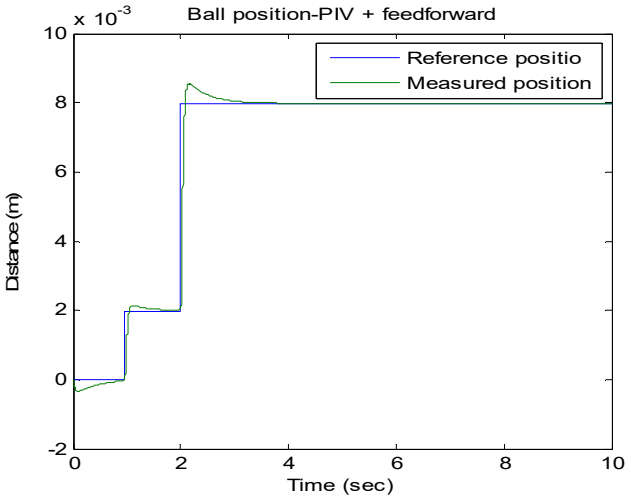


Fig. 4: Distance between coil and steel ball (from 2-8mm) - system response to PIV + feedforward controller

a) Performance Analysis

- The plot in fig. 4 shows a good tracking of reference input with maximum overshoot of about 25%, rise time of about 50msec, settling time of about 1 second and decay ratio of 0.3.
- Steady state error is indeed zero.
- No input saturation.

b) Robustness Analysis

In simulation, 25% loss of control due to uncertainties was tested. An increase in overshoot, settling time and steady state error were experienced as shown in fig.5. Thus, it can be concluded that the system is less robust to uncertainties.

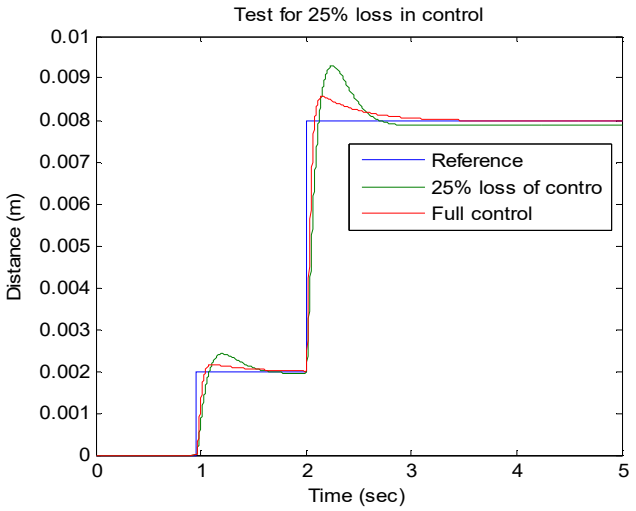


Fig. 5: Distance between coil and steel ball (from 2-8mm) – Loss in control due to uncertainties

IV. FORMULATION OF PROPOSED H-INFINITY CONTROL PROBLEM

A. Performance Specifications

Checking the magnetic levitation test-bed, it is therefore desirable to build a template for the desired performance requirement, so as to obtained the desired internal sensitivity functions for absolute performance and robustness. Weights for the Sensitivity functions were built from the desirable specification and some characteristics adapted from the existing controller.

For the design of the H-infinity control, one need to consider the use of two templates $W_e(s)$ and $W_u(s)$ for tracking and Actuator constraints respectively, with structures given by [3]:

$$\frac{1}{W_e(s)} = \frac{s+\omega_b\varepsilon}{\frac{s}{M_s}+\omega_b} \quad (11)$$

$$\frac{1}{W_u(s)} = \frac{\varepsilon_1 s+\omega_{bc}}{s+\frac{\omega_{bc}}{M_u}} \quad (12)$$

Specifications for weight, W_e:

- Peak value of W_e, $M_s = 2$, to ensure sufficient module margin.
- Desired steady state error in tracking, $\varepsilon = 0.01\%$

- Bandwidth, $w_b = 1.16\,rad/s$ of the PIV + feedforward control [1].

Specification for Weight, W_u:

- Saturation limit, M_u can be taken from the PIV + feedforward control [1] with $ks = \frac{\Delta_u}{\Delta_r} \approx \frac{2}{2\times10^{-3}} \approx 1000$. Hence, $M_u = 1000$, to avoid saturation of coil
- Desired steady state error in noise attenuation $e_1 = 0.001\%$.
- Bandwidth, w_{bu} will be taken as $w_T = 56.8rad/s$ with rise time of 40msec [1].

Remark: It is desirable for the Ks_r and Ks_n to roll off in high frequencies with $W^{-1}{}_u$, to give a fast roll off and more gain for control action.

B. H-Infinity controller design

In order to achieve perfect tracking with better uncertainties rejection (disturbance and noise), A two degree of freedom controller was selected and designed based on the need to control electromagnetic field force due to current and the position of the ball using two independently controllers.

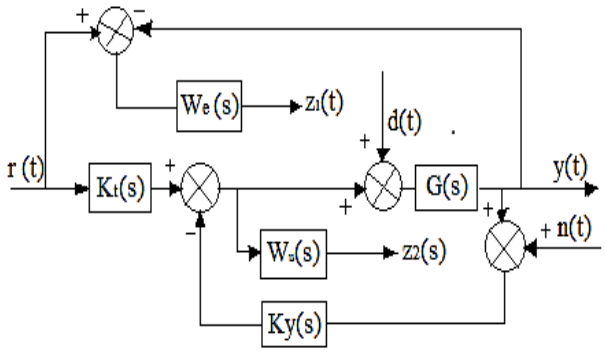


Fig. 6: Closed loop Configuration- H-infinity control

The closed-loop interconnection structure together with the weighting functions shown in fig. 6, to get the generalized plant *P* for the mixed sensitivity H-infinity control problem,

The variables were defined as follows:
$w(t) = [r(t),\ d_i, d_y, n]$ are the exogenous input signals
$u(t) = [u(t)]$, is the control inputs signal.
$y(t) = [y(t)]$ is the signal measurement output.
$z(t) = [z_1(t), z_2(t)]$ are the controlled output signals.

The state representation of the generalized plant is described as:

$$P: \begin{bmatrix} \dot{x} \\ z \\ y \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} x \\ u \\ w \end{bmatrix} \quad (13)$$

The closed loop system is deducted as:

$$\begin{bmatrix} z_1 \\ z_2 \\ r - y \end{bmatrix} = \begin{bmatrix} W_e & -W_e G & -W_e & -W_e G \\ 0 & 0 & 0 & -W_u \\ 1 & G & 1 & -G \end{bmatrix} \begin{bmatrix} r \\ u \\ d \end{bmatrix} \quad (14)$$

$$T_{zw} = F_l(P, K) = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21} \quad (15)$$

The optimization problem is to design a controller K(s) that minimize l_2 norm equivalent H_∞norm from the output z to exogeneous input w, such that the closed-loop system yield:

$$\|T_{zw}(s)\|_\infty \leq \gamma \quad (16)$$

Where γ is the maximum possible attenuation.

And the controller K yield:

$$K: \begin{bmatrix} x_k' \\ u \end{bmatrix} = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} \begin{bmatrix} x_k \\ y \end{bmatrix} \quad (17)$$

Remark: the controller K(s) is composed of two part: a feedforward control $K_r(s)$ and a feedback control $K_y(s)$. The control law generates an action $u(t)$ based on two inputs $r(t)$ and $y(t)+n(t)$.

C. Frequency response Analysis

The sensitivity functions were derived as follows:

Output performance:

$$y = \frac{Gd_i}{1+GK_2} + \frac{GK_1 r}{1+GK_2} + \frac{GK_n}{1+GK_2} \quad (18)$$

Input performance:

$$u = \frac{K_1 r}{1+GK_2} - \frac{GK_2 d_i}{1+GK_2} - \frac{K_2 n}{1+GK_2} \quad (19)$$

Then,

$$S = \frac{1}{1+GK_2} \quad (20)$$

$$T: \begin{cases} T_r = \frac{GK_1}{1+GK_2} \\ T_n = \frac{-GK_2}{1+GK_2} \end{cases} \quad (21)$$

$$KS: \begin{cases} K_{sr} = \frac{K_1}{1+GK_2} \\ K_{sn} = \frac{K_2}{1+GK_2} \end{cases} \quad (22)$$

$$SG = \frac{G}{1+GK_2} \quad (23)$$

The sigma plot for the weighting functions and sensitivity functions are shown in fig. 7. for the H-infinity problems with minimum attenuation ($\gamma = 0.5091$).

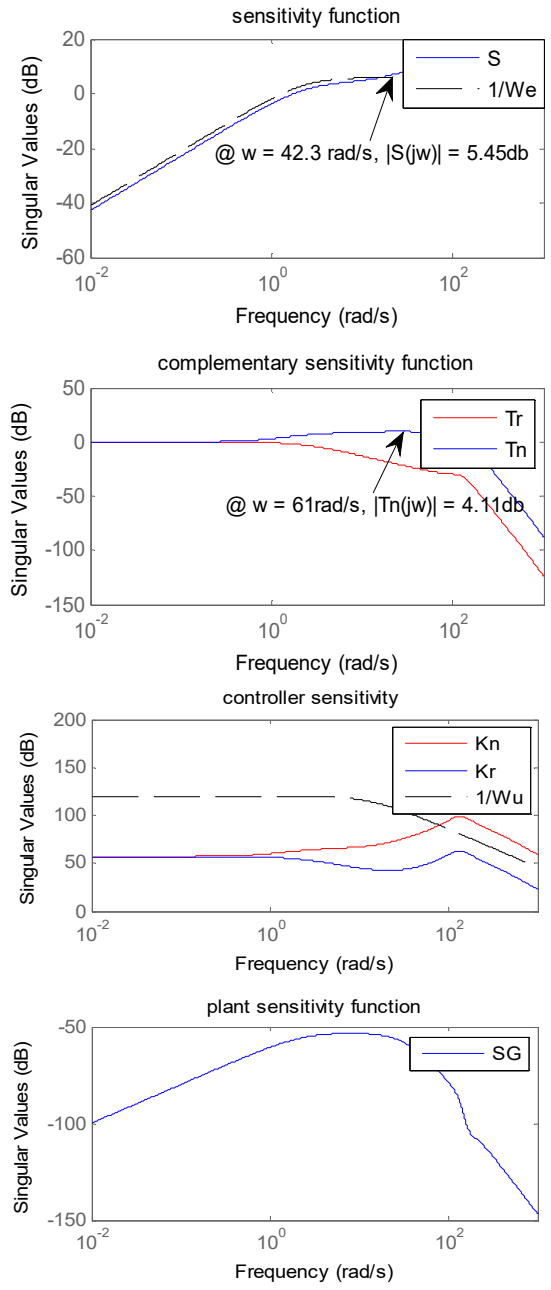


Fig. 7: Sensitivity function for H-∞ control system (a, b, c and d)

a) Performance Analysis

In comparison with frequency response analysis in (III-A)

- **S and SG:** when compared with analysis of fig. 2a, the system meets the desired performance specifications i.e. from fig. 7a, at low frequency, $|S(j\omega)|$ is small with a steady state error in tracking of approximately **–50db** (1%) at $w \approx 0\, rad/s$. Furthermore, from fig. 7d, at low frequency, | **SG**(jω)| is low, which is identical to that of fig. 2d. Thus, it can be concluded that there is a good tracking performance and good input disturbance attenuation.
- **T:** when compared with analysis of fig. 2b, one can say the system meet the desired performance specifications i.e. from fig. 7b at high frequency, $|T(j\omega)|$ is low, $|T_r(j\omega)|_{w=0} = 1$ which emphasis perfect reference tracking.
- **KS:** when compared with analysis of fig. 2c, both plots of $KS_r(j\omega)$ and $KS_n(j\omega)$ meet the performance specifications from fig. 7c i.e. low at high frequencies, which shows great noise rejection at high frequencies.

b) Robustness Analysis:

For the system to be robust to uncertainties, Module margins; $\max|S(j\omega)| \leq 6db$ and $\max|T(j\omega)| \leq 3.5db$ must be satisfied [3]. From fig. (7a) and fig. (7b), It can be seen that, the values of $|S(j\omega)|_\infty = 5.45db$ and $|T_n(j\omega)|_\infty = 4.11db$ respectively. Thus, the system is indeed robust to uncertainties, all module margins satisfied.

D. Time domain analysis

In this section, the same simulations scenario as in section III-B is repeated, with reference to ball position of 6mm from the coil. Fig. 8 illustrates the action of the H-infinity control design on the system response.

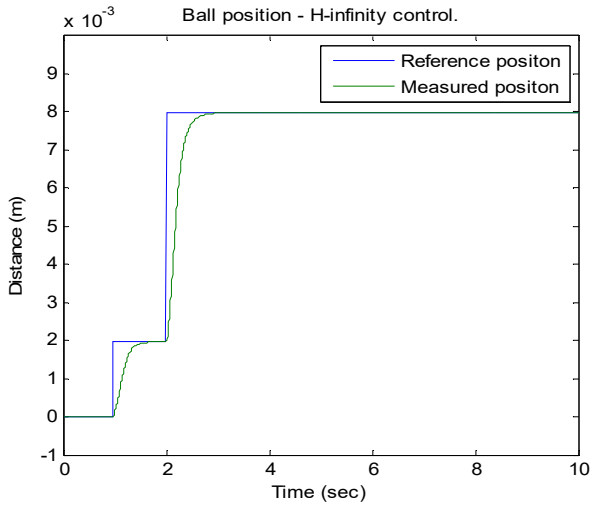


Fig. 8: Distance between coil and steel ball (from 2-8mm) – system response to H-Infinity Controller

a) Performance Analysis:

- The plot in fig. 8 shows an excellent tracking of reference input with no overshoot, rise time about 500msec and settling time of about 1 seconds, this indeed shows sufficient tracking performance.
- Steady state error is indeed zero.
- No input saturation.

b) Robustness Analysis:

In simulation, 25% loss of control due to uncertainties was tested. There was little effect experience on the response, rather a minimum of 50% loss of current control can only be noticeable on the response. as shown in fig.9. Thus, it can be concluded that the proposed control has an excellent tracking performance and well robustness to uncertainties.

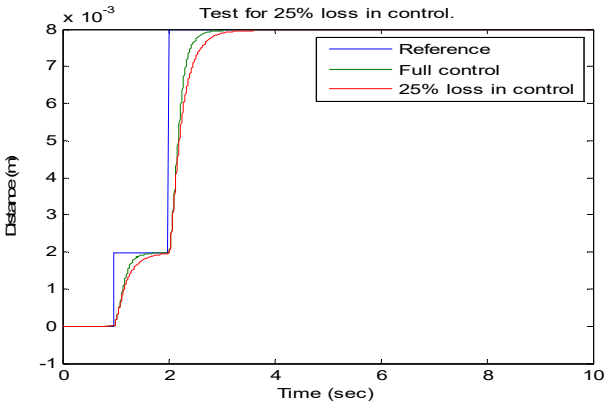


Fig. 9: Distance between coil and steel ball (from 2-8mm) – Loss in control due to uncertainties

V. CONCLUSION

In this paper, the overall objective was to designed an efficient controller based on H-infinity approach for magnetic levitation system. The task of this control is to improve the stability of the steel ball floating at equilibrium point via measurements of the ball's vertical position from the coil.

Performance analysis of the existing control was done, to verified its tracking performance and robustness to uncertainties. Simulation results obtained, confirm overall system performance but very sensitive to uncertainties.

The frequency and time response comparison of the designed H-infinity controller to that the existing PIV+feedforward controller was done. Results obtained shows significant performance and robustness against uncertainties of the proposed control. In addition, this proposed method gives a chance to include all kind of uncertainties in the control design, thus, able to withstand large scale of disturbances.

Hence, the designed H-infinity controller for position control of MAGLEV was successfully achieved and tested on the non-linear system model both on simulations and experimental validation.

REFERENCES

[1] Quanser Inc., "Magnetic levitation plant user manual,", 2012.

[2] Alexey A. Bobtsov , Anton A. Pyrkin, Romeo S. Ortega, Alexey A. Vedyakov. "A state observer for sensorless control of magnetic levitation systems." Automatica, vol. 97, Nov. 2018, pp. 263–70

[3] Vinodh Kumar, E., and Jovitha Jerome. "LQR based optimal tuning of PID controller for trajectory tracking of magnetic levitation system." Procedia Engineering,vol. 64,2013 pp 254-264.

[4] Tatsuro Kumada, Gan Chen and Isao Takami. "Gain scheduling control for magnetic levitation device using redundant descriptor representation." 2014 4th Australian Control Conference (AUCC), IEEE, 2014, pp. 199–204.

[5] Neha Singru, Divyesh Ginoya, P. D. Shendge, S.B. Phadke "A state-feedback control approach via inertial delay observer for Magnetic levitation system." 2015 International Conference on Industrial Instrumentation and Control (ICIC), IEEE, 2015, pp. 1670–74.R.

[6] Ryan J. McCloy, Jose A. De Dona and Maria M. Seron. "Control of a maglev system using the LPV framework: A tutorial from modelling to experimental implementation." IFAC-PapersOnLine, vol. 51, no. 26, 2018, pp. 100–05.

[7] Sename, O. "Robust control of MIMO systems".[powerpoint slides], 2018. retrieved from https://www.gipsa-lab.grenoble-inp.fr/~O. Sename

[8] H. I. Ali, "H-Infinity model reference controller design for magnetic levitation system." Engineering and Technology Journal, vol. 36, no. 1, Jan. 2018.

a. # WEB-BASED CATARACT DETECTION SYSTEM USING DEEP CONVOLUTIONAL NEURAL NETWORK

Musa Yusuf *Computer Science Department* *Bingham University,* Karu, Nigeria yusufmusa@binghamuni.edu.ng	Samuel Theophilous *Computer Science Department* *Bingham University,* Karu, Nigeria s.theophilus@binghamuni.edu.ng	Jadesola Adejoke Emmanuel *Computer Science Department* *Bingham University,* Karu, Nigeria emmanueladejoke@yahoo.co.uk
	Annah B. Hassan *Computer Science Department* *Bingham University,* Karu, Nigeria annebjk@gmail.com	

Abstract— ***The alarming cases of cataract within the last decade and the projection of cataract cases within the next few decades call for urgent intervention by early diagnosis. Formal ways of detecting cataract such as physical examination, tests and diagnosis suffered from subjectivity and errors due to professional bound. Hence the need for automation process. Some works have been done on Computer Aided Diagnosis (CAD) of cataract with tools such as Expert systems, which are limited to their knowledgebase thus has the tendencies of inaccuracies in result interpretation. Early detection of cataract is key to enabling quick intervention and treatment. This paper presents a web-based Computer Aided Diagnostic system for cataract detection leveraging of Deep Convolutional Neural Network (DCNN) that can be used by any nonprofessional outside the clinic environment. The systems model was trained on a data set of 100 eye images using transfer learning. The images were retrieved from google image search results of "normal human eyes" and "human eye cataract". Features are extracted by utilizing google inception model t model and transfer the extracted features using Transfer learning to train the proposed model. The proposed model gained the ability to classify eye images into "Normal" and "Cataractous". The system was designed to take images as inputs and achieved a Sensitivity of 69%, a Specificity of 86%, Precision of 86%, F-Score of 56% and AUC of 84.56%. Its accuracy score was 78% which was influenced using the model trained during the ImageNet image classification using deep convolutional neural network***

Keywords—*: **Computer Aided Design, Convolutional Neural Network, Cataract detection, Deep Learning.***

b. ## I. INTRODUCTION

US department of health and human services, National eye institutes define cataract as a clouding of the lens in the eye that affect most vision [1]. Lens is the clear part of the eye that helps in focusing light to the retina. Retina is the light sensitive tissue at the back of the eye, Cataract has been noted by National Eye Institute in the United States as one the leading causes of blindness resulting from eye diseases in the world [2]. The human eye lens gets cloudy which results to the production of blurry images by the eye in a situation. This eye defect at its latter stage leads to blindness but can be corrected through eye surgery by an Ophthalmologist to replace the natural lens with a synthetic lens. [3] stated that statistics from the Nigerian National Blindness and Visual Impairment Survey shows that 42 out of 1000 adults aged 40 and above are blind and that eye diseases like Cataract blind 2 out of 3 Nigerians. As at 2011, about 750,000 Nigerians were blinded by cataract. This shows that cataract could affect more individuals in future if it is managed properly. The need to find effective means of reducing the blindness rate is unarguably important Ease of Use.

The paper proposes a CAD software that would be easy to assess and reliable enough to perform medical diagnosis. Previous researches over the years have tried creating medical diagnostic tools using techniques such as the use of Expert Systems [4]. The problems with some of the methods used is the restriction it gives the system. For instance, an Expert System is software that works through logical decision-making based on a knowledge base extracted from an Expert in the specific field. Such a system cannot make complex analysis or computation. Such a system can only be assessed by the authenticity of the information in its knowledge base.

To develop a system that is not limited by such restrictions, machine leaning methods, was implemented which were used in training a Convolutional Neural Networks (ConvNets) model that was used in the classification of cataract images. Machine learning is a branch of computer science, which gives the computer the ability to learn without the need to explicitly program the knowledge. Its learning algorithm types have been classified into various categories based on the way the learning takes place [5].

c. ## II. BACKGROUND

Recently researchers have developed much interest in using machine learning in medical diagnosis, computer aided design and other fields such as computer vision. These researches were initially started as theoretical pieces but have recently found practical applications. In this section, some charts presenting ground-truth on cataract generated from research statistics collected by NEI in 2010 are presented in Fig 1[2].

Fig 1 shows the distribution of prevalence rate of cataract in US age wise. The older age groups have a higher percentage of cataract infected persons. From fig 1, we can deduce that cataract usually

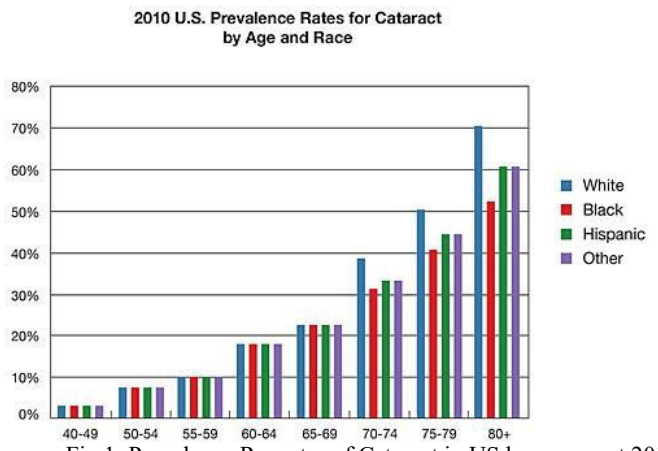


Fig 1: Prevalence Ra tes of Cataract in US by race as at 2010 (Cataracts, 2010)

manifest on people from age 40 but age of 80 and above have the highest prevalence rate.

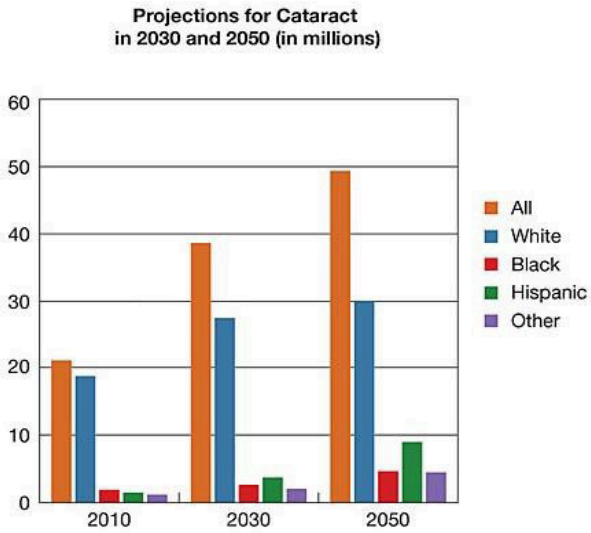


Fig 2: Projections for Cataract in 2030 and 2050 (in millions) [17]

Fig. 2 shows projected statistics of Cataract cases in the year 2030 and 2050 if the prevalence rate is not abridged. This result therefore shows how precarious cataract could be in future and therefore researchers need to foster way to tackle this challenge and therefore motivates the researcher to study this problem.

d.

III. Related Work

In recent years, there has been an increase in the studies and researches done focusing on machine learning. However, these researches have being applied in numerous fields, a good fraction of these researches aimed at improving medical diagnosis using different machine learning algorithms. Some of these researches initially started as theoretical pieces but have recently found practical applications.

[6] developed a Medical Imaging Analysis System using Artificial Neural Networks, the research showed how artificial neural networks were used for Computer-aided diagnosis systems. ANN was use in the detection and diagnosis of breast cancer and lung cancer. A computer-aided detection system identifies microcalcification clusters automatically on full field digital mammograms using six stages (pre-processing, image enhancement, segmentation of micro-calcification candidates and false-positive reduction for individual micro-calcifications, regional clustering and false positive reduction for clustered micro calcifications. A training set of 96 images where used along with 46 images for validation. The Convolution Neural Network was used to create a model to simulate the vision of vertebrate animals and use as the simplified vision machine, which performed the classification of the regions into two output types: disease and non-disease, while a linear discriminating analysis classifier was then used to differentiate clustered micro calcifications from false positives. The ROC analysis showed the system's false-positives reduction rates were 86%, 74% and 72% at sensitivities of 70%, 80% and 90% respectively when compared to classification without the Convolutional Neural Network/Linear Discriminating Analysis approach. They also applied the detection and diagnosis of bone mineral density (BMD).

In [16] A Computerized Cataract Detection and Classification was developed. The system was able to classify 156 eye images and got 50 images that contained normal eyes, 39 images with nuclear cataract, 33 images with cortical cataract and 34 with PSC. They gathered their dataset by capturing eye images with a Topcon SL-45 Scheimpflug camera. 90 images out of the 156 were used as the training set to create the modelled profile for the four categories. The remaining 66 images with unknown categories were classified into the four categories based on their distance from the modelled profiles in Euclidean space. The developed system was analysed and found to have sensitivity of 98% in detecting Cataract in images, specificity of 100% in identifying normal eyes and classification accuracy of 98%.

[7] developed a system that takes low-resolution images as input and produces a high-resolution image in the paper; Learning A Deep Convolutional Network for Image-Resolution. The system developed learns end-to-end mapping between low-resolutions and high resolutions using the deep learning convolutional neural network. The methodology was compared to some existing methods and found to outclass them, but it was also easier to implement and light to run (can be executed with a CPU). They stated that just like the sparse coding-based method, convolutional network would not be limited by internal training set but work well with external examples.

[8] carried out an ImageNet Classification with Deep Convolutional Neural Networks. In the study, a model was train using the Deep Convolutional Neural Network classifier with a dataset of 1.2 million high-resolution images. The model had 1000 different classes and when data was tested on the model, the study achieved top-1 and top-5 error rates of 37.5% and 17.0% respectively. It contained eight layers that consisted of five convolutional layers and three fully connected layers with a final 1000-way softmax. The research stated that it took six days to train the model on to GTX 580 3GB GPUs which was influenced by the dataset's size and number of layers involved in the training. The study also stated that each layer contained no more than 1% parameters required for training the model.

e.

IV. Artificial Neural Network

An Artificial Neural Network is an interconnection of nonlinear neurons. This is a machine learning classifier that was designed to model how the human brain works. It was created to act like the biological neurons (performing multiple tasks in a very short period of time).

From fig 3, we can define the following terms: "X_i" – is the input feature over n; "n" – a set of features; "W" - Weights of each corresponding feature and Finally an activation function is used

on the sum and outputs are produced [9]. This can be represented with

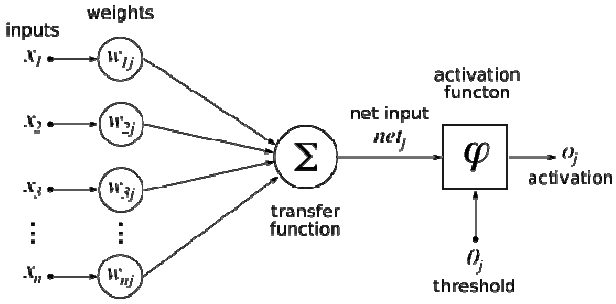


Fig 3: Showing the structure of an Artificial Neuron [15]

the equation (1):

$$y = f(x_n) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

The output y is generated from the given activation function $f(x)$ as shown in Equation (1). Where x_n is number of features over n.

$$y = \sum_{i=1}^{n}(w_i x_i + w_0) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . (2)$$

In more appropriately mathematical notation, we can represent equation (1) as equation (2) with an addition of bias weight w_0 to the sum of product of w_i and features x_i for i = 1, 2, … n .

The back-propagation technique from equation (3) was used to train the model. Whenever features x_i are passed through the classifier, it uses the gradient descent algorithm to adjust the weights w_i of features x_i so they produce the desired output as defined in the dataset used for the training. After each step, data is fed back to the classifier to retrain and training is only said to be done when the input features lead to the expected label that is when gradient's error measure is at its minimum.

f.

V. Convolutional Neural Network

The ConvNets is an ANN that has more than one layer and trains its models by adjusting its weights using optimization methods such as gradient descent algorithms. The first layer in the ConvNets is the "convolution layer" which uses the operation of convolution to search for patterns. Neurons in this network are used to perform the convolution operations i.e. filtering images to look for patterns. The neurons in each given filter share the same weight and bias thus giving them the ability to search for the same pattern in different parts of the image. The second layer is the "retified linear unit layer" which focuses on building up the patterns discovered by the convolution layer. The third layer in the ConvNets is the "pooling layer". This layer reduces the number of patterns allowing the neural network to focus only on important patterns. The final layer is called the fully "connected layer" which makes the ConvNets able to classify data into appropriate labels.

The accuracy of the ConvNets classifier is largely dependent on the size of the dataset it trains on, as it usually requires a very large set of training data. That was one of the major reasons why the initial ConvNets models did not perform very well (they did not have access to a very large dataset). Getting a very large dataset can be very stressful and difficult; the second problem has to train new models on large datasets from scratch every time they are being developed. This problem makes it almost impossible for researchers without heavy hardware specifications to implement ConvNets. It is for this reason that transfer learning was introduced so that knowledge collected in well-trained models can be transferred to the new models and used to build the new model while conserving hardware and time [10].

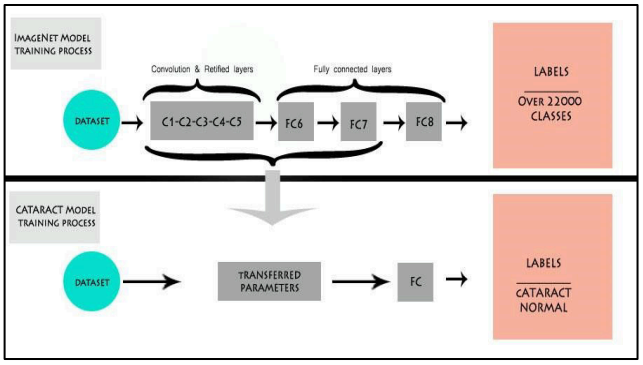


Fig 4: The system model showing how learned knowledge is being transferred from ImageNet Model to Cataract Model (Sengupta, 2003)

VI. Methodology

The system developed is a web-based application that takes images as inputs. It uses the images to train a model based on the set of images it has taken as input (it does this using the ConvNets algorithm). The model created is then used as a standard to classify eye images as malignant or benign cataract images.

The system is a three-paged web application. The first page is the landing and contains a summary and brief note on the functions of the application and the eye centre. The second page gives an educative information on cataract and the third page has the fields used to accept inputs for evaluation.

The model trained on 1.2 million images was used in the training of the new model to reduce the large dataset and heavy hardware requirements needed to develop a model that was accurate [11]. This is possible using a technique known as transfer learning in which information from another model is used along with other dataset(s) to increase the performance of the new model and to reduce the requirements needed for the computation of the new model. The "fully connected layer" was removed and the remaining part of the model (which contained information extracted by the initial layers) is treated as a fixed feature extractor for the new dataset. It computes a 2048-dimensional vector for each image that contains the activations of the hidden layer just before the classifier. Finally, a softmax function/classifier, was used as the classifier to trained to classify the new dataset [12]. The ImageNet model used was generated using the ConvNets.

The Gradient descent algorithm was used to optimize our model by finding the smallest value of a function that minimizes the error in equation (1). Its main objective is to find a value that fits a given curve so that it gives the best minimum error for the given function over point $t_o{}^p$ and $y_o{}^p$

When inputs are passed through the trainer and weights have been assigned to them, their calculations are initially not very accurate. By using the gradient descent, the weights can be adjusted until each input produces the desired output as specified in the training dataset. This process is described as the learning process. In this paper, a learning rate of 0.01 was used to descend using 4000 steps.

The solution of the gradient descent was done using the formula below:

$$E = \sum_{p} E^p \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . (1)$$

In equation (1), E computes the sum of all error as E over points P

$$E^p = \frac{1}{2}\sum(t_o^p - y_o^p)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

In equation (2), E^p is the Error measure of the slope between points $t^p{}_o$ and $y_o{}^p$.

$$\text{Then the Gradient} = \frac{\partial E}{\partial W_{ij}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where Wij is the synaptic interconnection from neuron j to neuron i.

To correct the weight W_{ij} we compute the change t_i and y_i

$$\Delta W_{ij} = \eta(t_i - y_i)x_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$[new]W_{ij} = [old]W_{ij} + \Delta W_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where η is the learning rate, and *i* is the output while *j* is the input

The softmax regression is a linear classifier implemented on the final layer of a ConvNets, which is applied mainly for classification. The networks are trained under a cross entropy which gives a nonlinear variant of multinomial logistic regression. Given that every image input in the system has only two possible label outputs, evidence needs to be gathered from the image, which determines the label it belong. The evidences gathered are then calculated to probabilities and the image then is classified into the label with the highest probability score. A weighted sum of the pixel intensities for each input image is calculated. The calculated weight could be either negative or positive. The positivity or negativity of the weight is determined by the high intensity pixels in the image that holds evidence that it beholds to the given label or evidence that it does not behold to the label. An extra bias is added to the evidence. This bias tries to suggest that some evidences are not dependent on the input. This results in the equation:

$$E_i = \sum_j w_{i,j}x_j + b_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Where i represents a given label; j represents the index of summation of pixels of input image; x is the input image, w is the weight and the evidence is resented with E.

$$Y = \text{Softmax}(E) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Using equation (7), we use the Softmax activation function to generate an output labels Y between the range of 0s and 1s.

$$\text{Softmax}(x) = \text{normalize}(\exp(x)) \ldots\ldots\ldots\ldots\ldots (8)$$

Using equation (8), we eliminate possibilities by computing the exponential of x_i (9)

$$\text{Softmax}\ldots\ldots\ldots\ldots (x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

The inputs undergo exponentiation then normalization to eliminate the possibility of having negative weights so that their sum is equal to 1 i.e. sum of probabilities equal to 1[13].

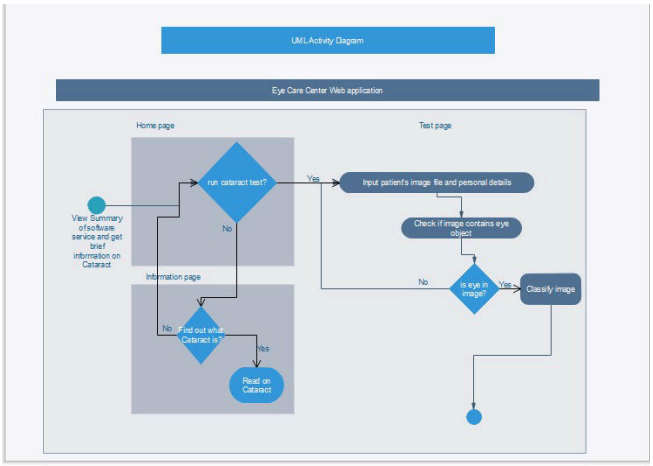


Fig Showing the Activity UML Diagram for the proposed system Fig 5 shows the flow of data in the proposed system.

VII. System Requirements

The designed system has the following system requirements;

1. Run application on a Server that has python 2.7 interpreter installed. This is because python is the server-side scripting language used for the development of the web application.
2. Tensorflow python library should be installed on the server. This is the machine learning API used in this study, which was written mainly for the CNN classifier.
3. Numpy library should be available on the server. The images used in this study had to be converted to complex arrays before operations are performed on them. Numpy is a python library that allows such operations with ease.
4. OpenCV library should be available on the server. This is a computer Vision python library that would be used in this study to implement the Viola Jones' objection detection.
5. Django framework library should be available on the server. This is a python library used to develop web applications in python.
6. Material Design Lite CSS framework is going to be the library used in writing the styles for the webpages.
7. A browser needs to be installed on the user's smart device/ Computer system in order for him to gain access to use the application.
8. A Database needs to be designed so that all tests taken are recorded on the Server (in this case SQLite).
9. The server used to host the developed system was a Linux server. This is because the Tensor flow library was written for the Linux and Mac OS and did not support the Windows OS at the time this system was developed.
10. A trained model is to be made available on the server to be used for classification.
11. The system runs on CPU with 2.7 GHz and 12 GB RAM, with dedicated graphics card of 4GB

g. VIII. Experiments and Evaluation

All the images used in the training process were retrieved from google search results with searches "human eye image" and "Cataract eye image". Training the model based on the 100 images gotten from google search result will not produce a very accurate model. So in order to save time and gain high level of accuracy, the google inception model was downloaded. This

model was created by google after training millions of images [12]. The images gotten were trained using the "retrain.py" file written by Google in the tensor flow examples on GitHub and used to create a new model. The file uses the inception model to start the training process before it works on the images. The use of the Tensor Flow CNN library eased the implementation and use of the classifier and model. A Total of 30 images were tested on the developed model and the system was evaluated using the ROC Curve and confusion matrix.

The Receiver Operating Characteristics Curve is a graph, which is used to evaluate the performance of classifiers. They initially had more application in the medical field to aid in some forms of decision making but have been adopted in machine learning and data mining researches to evaluate the performance of classifiers. The values used to plot an ROC curve are generated using a matrix known as confusion matrix. To plot an ROC curve to evaluate a binary classifier, a 2 x 2 confusion matrix is used which is the case with this paper.

Table 3: Confusion matrix table

	Actual Positives	Actual Negatives	Total
Predicted Positives	True Positives (TP)=11	False Positives (FP)=2	PP=TP+FP
Predicted Negatives	False Negatives (FN)=5	True Negatives (TN)=12	PN=FN+TN
Total	AP = TP +FN	AN = FP +TN	

$$\text{TPR} = \frac{TP}{AP} = \frac{11}{11+5} = 0.69$$

$$\text{TNR} = \frac{TN}{AN} = \frac{12}{2+12} = 0.86$$

$$\text{Precision} = \frac{TP}{PP} = \frac{11}{11+2} = 0.85$$

$$\text{F Score} = \text{Precision} * \text{TPR} = 0.85 * 0.69 = 0.59$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{11 + 12}{11 + 12 + 5 + 2} = \frac{23}{30} = 0.78$$

A Confusion matrix is a matrix, which contains information about the actual classifications and the predicted classifications performed by a classification system. It consists of four important cells:

i. True Positive (TP): This is the number of classifications that were actually positive and were correctly classified by the system. ii. False Negative (FN): This is the number of classifications that were actually positive but were wrongly classified by the system. iii. True Negative (TN): This is the number of classifications that were actually negative and were correctly classified by the system. iv. False Positive (FP): This is the number of classifications were actually negative but were wrongly classified by the system.

v. These values are analyzed and used to evaluate the system. The system used a threshold of 0.5 to classify the images. If the probability that image is in Class A (Normal) is higher than the probability that it is in Class B (Cataract), then it is classified into Class A. Class A was taken to be the positive class and Class B was taken as the negative class. vi. In order to get the values for the points to be plotted on an ROC Curve, the test set (in this case the images) are classified using different threshold values. At each threshold value used, a confusion matrix is constructed so as to get the "true positive rate" and "true negative rate" at that threshold. The ROC curve is finally plotted using the "true positive rate" values on the vertical axis and the "false positive rate" values on the horizontal axis [14].

A threshold of 0.5 was used in the experiment, which resulted in the generation of the confusion matrix below 0 => Cataractous, 1 => Normal

Table 1: Table showing TP, TN, FP, TPR, TNR and FPR values at different thresholds

SN	Threshold Value	TP	TN	FP	FN	TPR	TNR	FPR
1	0.0	16	0	14	0	1	0	1
2	0.1	15	6	8	1	0.94	0.43	0.57
3	0.2	15	8	6	1	0.94	0.57	0.43
4	0.3	14	8	6	2	0.88	0.57	0.43
5	0.4	11	10	4	5	0.69	0.71	0.29
6	0.5	11	12	2	5	0.69	0.86	0.14
7	0.6	11	12	2	5	0.69	0.86	0.14
8	0.7	10	12	2	6	0.63	0.86	0.14
9	0.8	9	14	0	7	0.56	1	0
10	0.9	8	14	0	8	0.50	1	0
11	1.0	0	14	0	16	0	1	0

Table 1 shows the truth table values computed for each given image, which is used in the generation of the ROC chart shown in fig 4. The curve shows how well our algorithm performs

The Trapezoidal formula, the Area under Curve (AUC) of the ROC curve was calculated as shown in the table fig 4

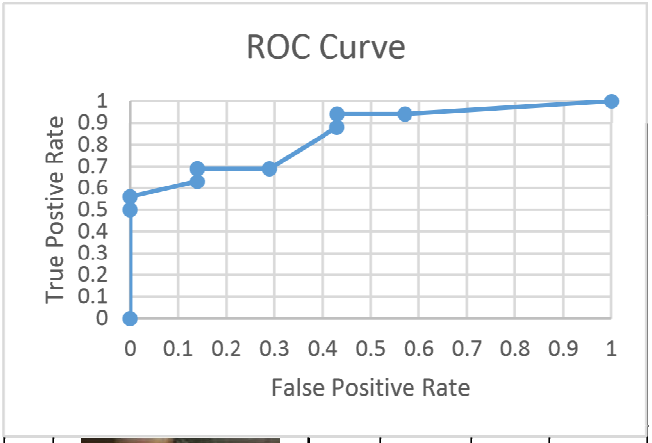
Table 2 shows the experiment result generated from tested some images on the developed system. Our system shows an interesting performance when tested with both cataract images and non-cataract images. From test image 5, our method had challenges in detecting catractous eye region due to distances of the eye from the camera. The face in image 5 really have cataract but because of the nature of the face and eye size of the patient, our system was not confident in the prediction as cataract eye. On the other way image two was predicted accurately because the affect ted eye was captured closer to the camera.

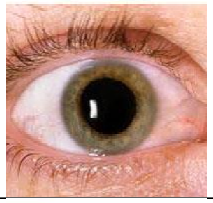
Fig 4: ROC curve generated from the values in Tables 5.3

When non-cataract images were tested on our system, in the case of images 3 and 4, the prediction was accurate as well. The two images ware actually non-cataract just as our system predicted.

Table 3 shows how the values used in the generation on the AUC for all images considered for this experiment. The proposed model generated AUC of 84%, which is better than the state-of-the-art used in detecting cataract.

Table 2: Result of some images tested with the proposed



2.		0	0	0.01	0.99
3.		1	1	0.99	0.01
4.		1	1	0.97	0.03
5.		0	0	0.40	0.60

methodology

IX. CONCLUSIONS AND FURTHER WORK

The evaluation performed on the system has beyond reasonable doubt shown how great the ConvNets Classifier is. Though the dataset was enriched with great details on the problem (images used for training were retrieved from google search results), the developed model still attained a high AUC of 84.56% and accuracy score of 78% which was influenced by the transfer learning trained on ImageNet. With a sensitivity of 69% and a specificity of 86% .This work has shown the strength of the ConvNets classifier and proves that it can be used to solve similar and even more complex problems.

This technique can be used in many other fields and is not limited to the medical field as can be found in. It showed how deep learning can be used to detect and classify cataractous and normal eye images as a CAD. It worked using transfer learning to show how time and hardware could be conserved will achieving high accuracy in the development of models.

Further work will consider fundus images so that multiclassification can be performed on the dataset.

Table 3: Table showing values used to calculate the AUC using the trapezoidal formula

i	Y-axis (TPR_i)	X-axis (FPR_i)	Area= (FPR_{i+1}-FPR_i)* (TPR_{i+1} +TPR_i)/2
1	0	0	0
2	0.5	0	0
3	0.56	0	0.0833
4	0.63	0.14	0
5	0.69	0.14	0
6	0.69	0.14	0.1035
7	0.69	0.29	0.1099
8	0.88	0.43	0
9	0.94	0.43	0.1316
10	0.94	0.57	0.4171
11	1	1	
			Total Area=0.8454
			AUC= (0.8454) *100 =84.54%

The use of the convolutional neural network can be further used to detect early cataract formation using images. In future, we intend to extend the capability of the system to be able to dictate glaucoma and other eye diseases so that CAD can reduce the rate of people blinded by eye illness.

X. REFERENCES

[1] U. D. o. Health and H. Services, "Cataract: What You Should Know," *Washington DC: NIH Publication,* 2003.

[2] "Cataracts," in *National Eye Institute* ed, 2010.

[3] C. Obinna, "Nigerians risk more blindness from cataract, if…," in *Vanguard*, ed, 2011.

[4] G. D. Magoulas and A. Prentza, "Machine Learning in medical applications," *Springer Berlin Heidelberg,* pp. 300-307, 2001.

[5] T. O. Ayodele, "Types of Machine Learning Algorithms," University of Portsmouth, United Kingdom2010.

[6] J. Jiang, P. Trundle, and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics,* vol. 34, pp. 617-631, 2010.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*, 2014, pp. 184199.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[9] X. Fern. (2011, Artificial Neural Network. *Oregon State University*. Available: http://web.engr.oregonstate.edu/~xfern/classes/cs534/notes/neuralnet-9-11.pdf

[10] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, *Learning and Transferring Mid-Level Image Representations*. MSR, New York, USA;INRIA, Paris, France, 2014.

[11] A. Krizhevsky, I. Sustsker, and G. E. Hinton, "ImageNet Classification with Deep Convolutional

Neural Networks," *Advances in neural information processing systems,* 2012.

[12] A. Karpathy. (2015, Transfer Learning. *CS231n Convolutional Neural Networks for Visual Recognition*. Available: http://cs231n.github.io/transfer-learning/

[13] TensorFlow, "MNIST For ML Beginners," in *TensorFlow*, ed, 2016.

[14] T. Fawcett, *ROC Graphs: Notes and Practical Considerations*: HP Laboratories Palo Alto, 2003.

[15] Knowino. (2010). Artificial Neural Network. Retrieved from Tel viv University: http://www.tau.ac.il/~tsirel/dump/Static/knowino.org/wiki/Artificial_neural_network.html

[16] Paul, A., Manuel, B., Michael, U., Benes, L., Valeria, F., & Kayoko, K. (2009). Computerized Cataract detection and Classification. Current Eye Research.

[17] Cataracts. (2010). Retrieved from National Eye Institute : https://www.nei.nih.gov/eyedata/cataract

XI. *APPENDIX*

This section shows the page Screenshot from the proposed system.

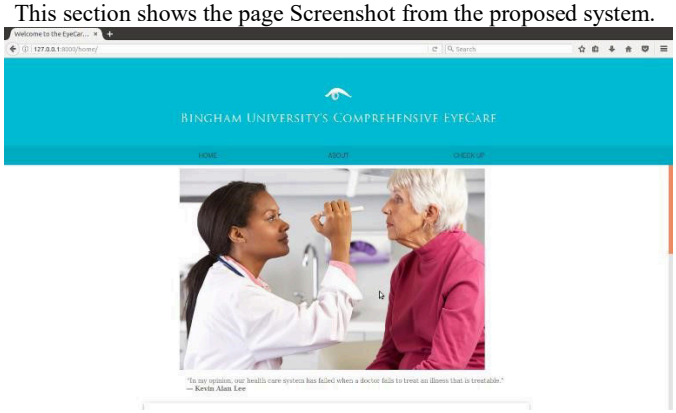


Figure 4.1: Screenshot of homepage of the Web Application

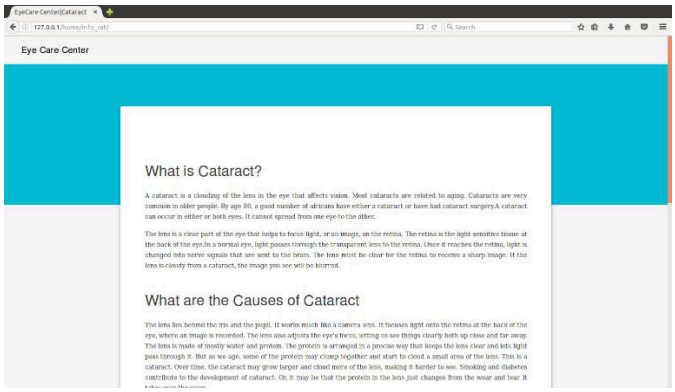


Figure 4.2: Screenshot of the about page of the web application showing information about Cataract and how to use the App.

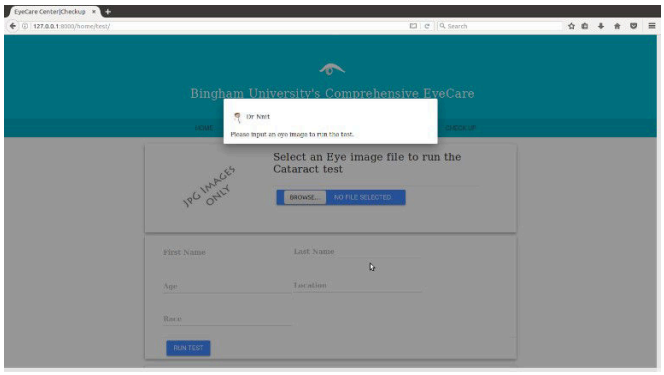
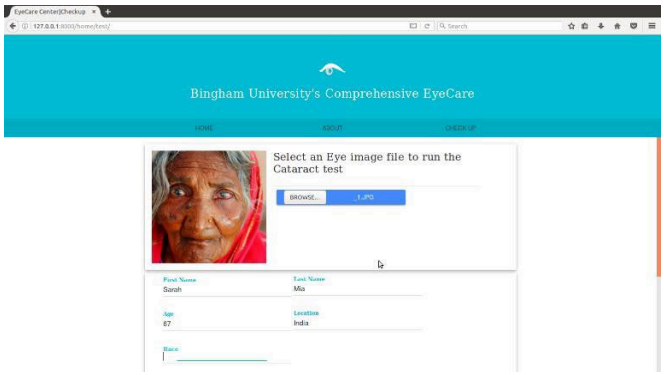


Figure 4.3: Screenshot of the Test page showing the welcom before any test is performed



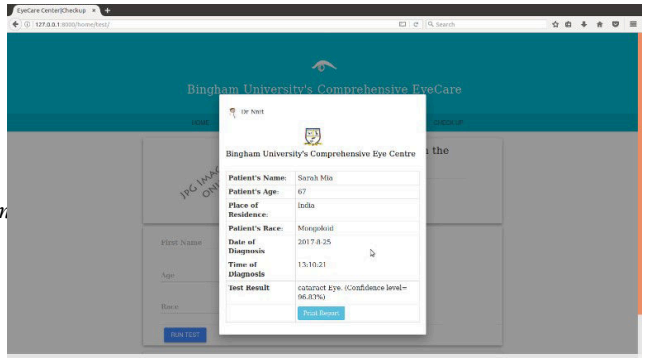


Figure 4.5: Screenshot of the Test page showing results of a diagnostic test along with thœonfidence level of the result.

Regulation of Induced Movement using Back Stepping Control

Mohammed Ahmed
Department of Electrical and Electronics Engineering
Abubakar Tafawa Balewa University
Bauchi, Nigeria
inunugoloma@yahoo.com

M. S. Huq
Department of Mechatronic Engineering
Institute of Technology
Sligo, Ireland
saif.huq@gmail.com

B. S. K. K. Ibrahim
School of Mechanical, Aerospace and Automotive Engineering
Coventry University
Coventry, United Kingdom
Ad1465@coventry.ac.uk

Abstract*—The paper presented an evaluation using the Back-Stepping control (BSC) for the regulation of Functional Electrical Stimulation (FES) aided sit-to-stand (STS) motion restoration by investigating the effect of human mass distribution across the world. The study focused on eradicating or reducing to the minimum the tasks of changing the control system parameters when applying the device for different subjects. The comparison was made with the Proportional Integral Derivative (PID) based gain scheduling based control approach. Results do indicate the robustness of the BSC control method as it portrays that despite the alteration of the masses to the extremes the control system was able to keep the tracking error approaching zero or very close to the desired with and without changes. It outperformed the PID based gain scheduling control scheme. It also operates within the permissible limits specified. Hence, it reveals the effectiveness of the BSC scheme for solving the problem of continuous readjustments making the equipment universal. It, therefore, relieve the burden of the necessity of changing the control variables continuously. The study is novel because the BSC has never been evaluated on four segments based induced STS movement in paraplegia model. The universal human mass distribution has also not been considered using the control scheme on the model

Keywords*—Control system, aided sit-to-stand, functional electrical stimulation, global human distribution, backstepping control, paraplegia

I. INTRODUCTION

Functional Electrical Stimulation (FES) is currently regarded as an essential technique utilised for movement function restoration, rehabilitation and therapy, for illnesses arising from nervous system failure. The method employs suitable levels of electrical currents. An essential characteristic of FES was the acceleration of the recovery of the neurons.

An issue still bothering the application of the FES system was not passing the clinical phase, which has some stringent requirements. The control systems tool have played an essential role in enhancing the performances of such devices. A problem worth solving according to literature selecting a more accurate and flexible control scheme capable of checkmating the burden retuning.

Therefore, the main objective of the study was to evaluate the Back Stepping Control (BSC) scheme by considering the terminal global human mass variations.

The linear control techniques are explored for the control of FES-assisted movement from sitting to standing as reported in the works of Dolan et al. [1], Poboronuic [2], Yu et al. [3], although the mathematical model of the system did not appear but for linear control methods the model has to be made linear even if it is nonlinear. Intelligent modelling technique was used by Kamnik et al. [4], and it was applied for both modelling and control by Davoodi and Andrews [5, 6]. Computer models are used and forms one of the disadvantages of the approach is the absence of a mathematical model which is vital for stability analysis and other evaluations [7] although the methods are well known for their excellent results. A sandwich of linear and nonlinear methods was presented by Previdi et al. [8]. Nonlinear approaches were proposed by Espanjani and Towhidkah [9] utilising virtual models. An advantage of employing the nonlinear control method is that it can be utilised for both linear and nonlinear systems. The nonlinear models of system are usually closer replicas to the actual systems. Moreover, utilising the nonlinear control methods can pave the way for stability studies using well-known standard techniques.

The nonlinear methods though least famous also have shortcomings that include the effect of switching in sliding mode control which could make the issue of fatigue worst, high computational time requirements in model predictive controllers and more suitability to periodic dynamics in the case of adaptive controllers. If these limitations could be solved with relative ease, they may be good candidates for the control of FES. Perhaps, after extensive studies with the nonlinear approach to ascertain the levels of stability with associated boundaries, intelligent based nonlinear techniques could also be explored. Nonlinear control has shown the potentials in improving the FES control, are more suitable considering the dynamics the dynamics of the plant, their shortcomings can be overcome or suppressed and also an avenue for extensive stability analysis exist which can be used for further enhancing the system using other or combining with other techniques.

The manuscript is made up of five sections, starting with the introduction, which gives a general overview of the work. Next is the materials and methods; explanation of the techniques and items utilised. It was followed by the results; where the results obtained are presented, and next

was the discussion; the presented are discussed. Finally, the conclusion gives a summary of the entire study.

II. Materials and Methods

In backstepping control (BSC) design complex, nonlinear systems are broke down to form subsystems. They are achieved in such a way that degrees of the subsystems were less than the primary system. Determination of Lyapunov functions and design of control laws followed accordingly. Making the design of the control law of the overall system was through the synthesis of those the subsystems in reverse and particular manner. Typically, starting with an innermost stable control system which continuously gets reinforced from the innermost subsystem to the outermost. Until finally, the desired control scheme is achieved. Hence, the name of the controller originated from the process of design mentioned [1, 2].

A. The STS Model

The STS model refers to the subject (that is the paraplegic) performing sit-to-stand manoeuvre with the aid of the FES. Relevant robotics principles as well as the muscle models presented by Ferrari and Pedotti [10, 11]. The description of the plant is given by equations (1), (2) and (3). Equation (1) is the inverse model and (2) is the forward model. The canonical form of a mathematical model is given by equation (3).

Where $D(q)\ddot{q}$ is the inertia vector, $C(q,\dot{q})\dot{q}$ is the Coriolis vector, $\dot{g}(q)$ is the gravity vector, $\ddot{q} = \ddot{\theta}$, $g(\theta,\dot{\theta}) = \frac{1}{D(q)}$, and $f(\theta,\dot{\theta}) = \frac{1}{D(q)}[-C(q,\dot{q})\dot{q} - \dot{g}(q)]$.

$$\tau = D(q)\ddot{q} + C(q,\dot{q})\dot{q} + \dot{g}(q) \quad (1)$$

$$\ddot{q} = \frac{1}{D(q)}[\tau - C(q,\dot{q})\dot{q} - \dot{g}(q)] \quad (2)$$

$$\ddot{q} = f(\theta,\dot{\theta}) + g(\theta,\dot{\theta})u \quad (3)$$

Equations (1) details are enormous. The subject is represented using four segments; the shank, thigh, trunk (from the hip to the shoulder, including the hands) and neck (neck and head). Their properties are incorporated in the various equations of the matrices. The multiple components of the muscle model are also included. The torque generated as a result of stimulation is given by (4) and (5) is the effect of dynamic friction. Others are the effect of gravity and stiffness friction. Where P_W is the pulse width of the stimulation current, δ is the time constant, G is the gain and β is the coefficient of the dynamic friction.

$$\tau_A = \frac{P_W G}{1 + s\delta} \quad (4)$$

$$\tau_D = \beta\dot{\theta} \quad (5)$$

B. The BSC Scheme

In backstepping control (BSC) design complex, nonlinear systems are broke down to form subsystems. Achieved in such a way that degrees of the subsystems were less than the central system. Determination of Lyapunov functions and design of control laws followed accordingly. Achieving the design of the control law of the overall system was through the synthesis of those the subsystems in reverse and peculiar manner. Typically, starting with an innermost stable control system which continuously gets reinforced from the innermost subsystem to the outermost. Until finally, the desired control scheme is achieved. Hence, the name of the controller originated from the process of design mentioned [12, 13].

The system is a second-order system. Therefore, the controller would be achieved using two subsystems. The first stage was as follows. Equation **Error! Reference source not found.** gives the error signals and **Error! Reference source not found.** was the system dynamics. Equation **Error! Reference source not found.** was the selected Lyapunov function for the first subsystem. Achieving the conditions in the equations **Error! Reference source not found.** and **Error! Reference source not found.** are necessary to maintain the Lyapunov criteria for guaranteed stability. Hence, the next stage was required, having achieved the first stage successfully, which continued as given by the next steps. And for the second stage, the Lyapunov function selected was as in equation **Error! Reference source not found.**. The equation **Error! Reference source not found.** carries the

$$e = \theta - \theta_d;\ \dot{e} = \dot{\theta} - \dot{\theta}_d;\ \ddot{e} = \ddot{\theta} - \ddot{\theta}_d$$
$$\ddot{\theta} = f(\theta,t) + b(\theta,t)u$$
$$V_1 = \frac{1}{2}e^2$$
$$\dot{V}_1 = e\dot{e} = e(\dot{\theta} - \dot{\theta}_d)$$
$$\dot{V}_1 \le 0;\ \forall t;\ s = \dot{\theta} - ce - \dot{\theta}_d = c_1 e + \dot{e};\ c > 0$$
$$\dot{V}_1 = es - c_1 e^2;\ If\ s = 0, then\ \dot{V}_1 \le 0, \forall t$$
$$V_2 = V_1 + \frac{1}{2}s^2$$
$$\dot{s} = \ddot{\theta} + c_1\dot{e} = \ddot{\theta} + c_1\dot{e} - \ddot{\theta}_d = f(\theta,t) + b(\theta,t)u + d(\theta,t) + c_1\dot{e} -$$
$$\dot{V}_2 = \dot{V}_1 + s\dot{s} = es - c_1 e^2 + s(f(\theta,t) + b(\theta,t)u + d(\theta,t) + c_1\dot{e} -$$
$$u = \frac{1}{b(\theta,t)}\left(-f(\theta,t) - c_1\dot{e} + \ddot{\theta}_d - c_2 s - e\right)$$
$$\dot{V}_2 = -c_1 e^2 - c_2 s^2 + sd$$
$$\dot{V}_2 \le 0;\ If\ c_1 > 0\ \&\ c_1 > 0,\ ;\ (ce^2 + cs^2) \ge sd\ \forall t$$
$$Hence, \quad e \to 0\ as\ t \to \infty$$

overall system control law and **Error! Reference source not found.** the differential of the Lyapunov function. The controller gains have to selected based on conditions given by equation **Error! Reference source not found.** to maintain the Lyapunov stability criteria [13].

A gain scheduling controller based on the PID control scheme (PID) was also implemented and compared with the BSC scheme.

III. Results

The results section describes the different results obtained in the study. It is reported based on three conditions: without any disturbance, with 53% subject mass and with 126% subject mass.

Fig. 1, Fig. 2 and Fig. 3 were the STS response with the BS and PID controllers without disturbance effect, the tracking errors during the transition, and the integrals of

the tracking errors during the transition using the controllers, respectively. Fig. 4, Fig. 4 and Fig. 6 were the control signals, the integrals of the rate of change in the stimulation/time stimulation current used by the controllers, and the integrals or cumulative amount of the stimulation/time stimulation currents during the movement respectively. Fig 7 is the phase plot of the response and Table I give the summary.

The integrals of the stimulation/time stimulation currents for the BS and the PID control schemes were 5. 2633 mA, 21.0534 mAs, and 182.1604 mA, 728.6414 mAs respectively. The PID utilised thirty-four (34) or 3360% more current than that of the BS controller. The integrals of the rate of change in the stimulation/time stimulation currents for the BS and PID controllers were 29.8138 mA/s, 119.2554 mA, and 3405.1 mA/s, 13620 mA respectively. It means the rate of change in the current using the PID was 113 times or 11300% compared to that of the BS controller. The integrals of the tracking/time tracking errors for the BS and PID controllers were 0.0580^0, 0.2319^0s, and 80.8103^0, 323.2413^0s respectively. It shows that the cumulative error of the PID controller was 1392 times more than that of the BSC. An RMSE of 0.00000390970 resulted using the BSC and 0.2697^0 with the PID controller. It indicates that it was about 68981 times compared to that of the PID control scheme. The maximum stimulation current of 24.0208 mA was used using the BSC and 316.5532 mA with the PID controller. It shows that the maximum stimulation current utilised by the PID control scheme was twelve (12) times or 1200% more than that of the BSC. The trajectory of the BSC control scheme from the phase plot in Fig. 7 show that the system is stable. A significant shortcoming shown on the PID control scheme trajectory was its oscillatory nature. Oscillations are undesirable in control systems and could lead to instability [12, 14, 15] if it grows beyond its specifications with time.

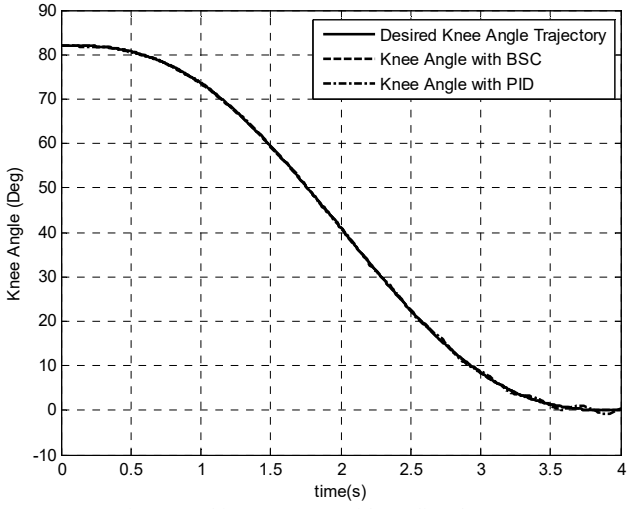


Fig. 1 Tracking Response without disturbance.

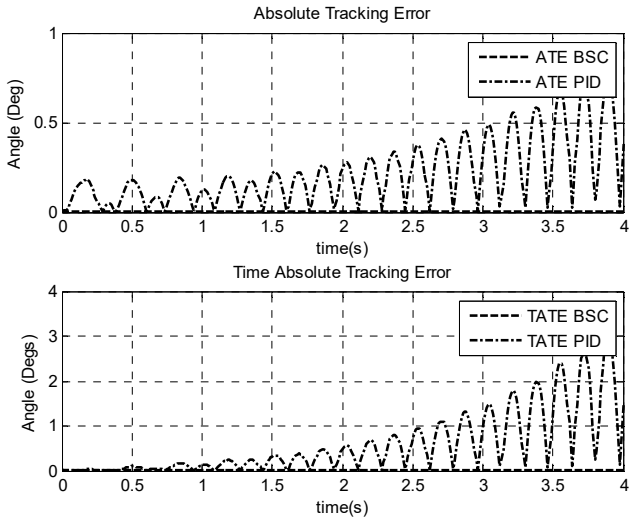


Fig. 2 Tracking Errors without disturbance.

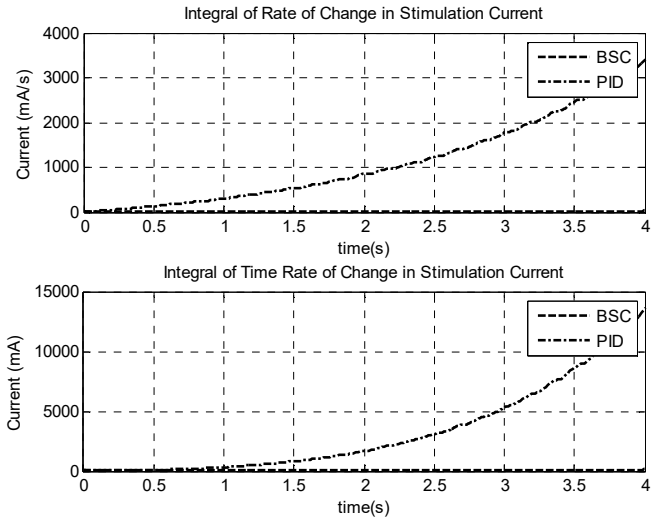


Fig. 3 Integral of tracking errors without disturbance.

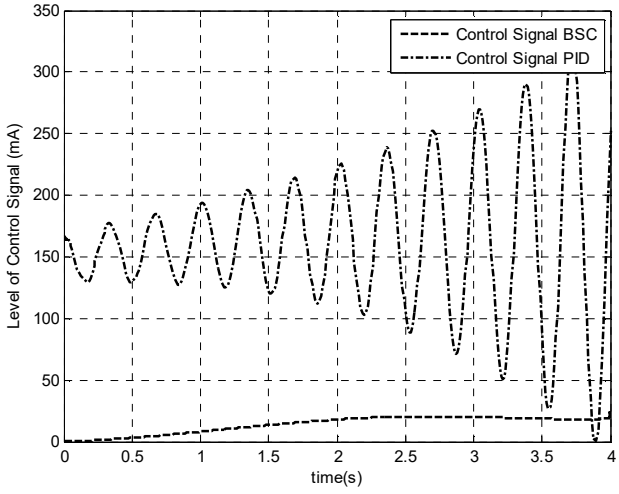


Fig. 4 Levels of Control Signals without disturbance.

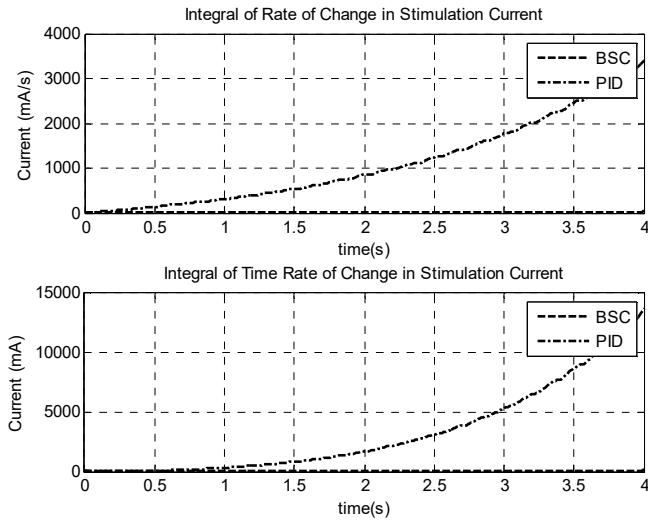


Fig. 5 Integral of Rate of Change in Stimulation Current without disturbance.

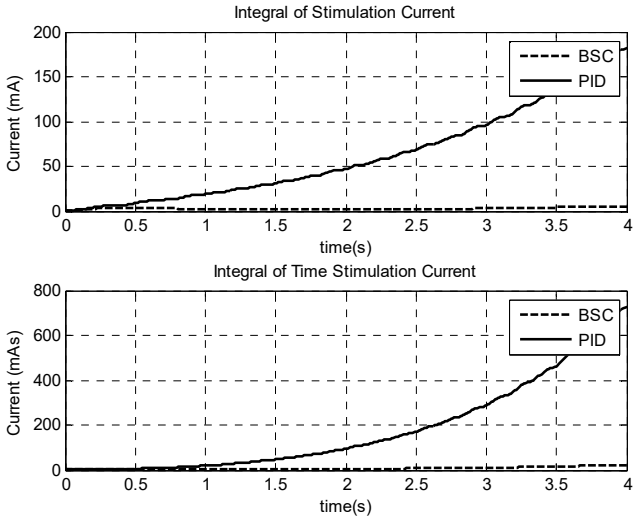


Fig. 6 Integral Stimulation Current without disturbance.

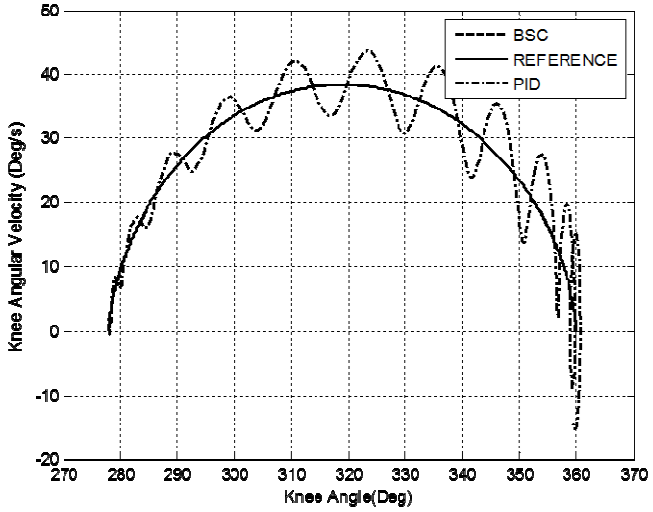


Fig. 7 Phase plot without disturbance.

RESULTS WITHOUT DISTURBANCE

Parameter	Control	
	BSC	***PID***
Integrals of Stimulation/Time Stimulation Current (mA/mAs)	5.2633/ 21.0534	182.1604/ 728.6414
Integrals of Rate of Change in Current/Time Current (mA^{-s}/mA)	29.8138/ 119.2554	3405.1000/ 13620.0000
RMSE (Deg)	3.9097u	0.2697
Integrals of Error/Time Error (Deg/Degs)	0.0580/ 0.2319	80.8103/ 323.2413
Maximum Control Signal (mA)	24.0208	316.5532

Fig. 8, Fig. 9 and Fig. 10 are the tracking response of the knee angle during STS transition with the expected minimum amount of subject mass (53% of the subject's mass), tracking error graphs, and the integrals of the tracking errors. Fig. 11, Fig. 12 and Fig. 13 are were the control signals, the integral (cumulative) of the rate of change in the stimulation/time stimulation current used by the controllers, and the integrals or cumulative amount of the stimulation currents during the movement respectively. Fig. 14 is the phase response with 53% subject mass and Table II is the summary of the results.

The integrals of the stimulation/time stimulation currents for the BS and the PID control schemes were 2.7887 mA, 11.1547 mAs, and 54.5686 mA, 218.2742 mAs respectively. The PID utilised nineteen (19) times or 1850% more current than that of the BS controller. The integral/integral time of the rate of change in the stimulation currents for the BS and PID controllers were 13.6848 mA/s, 54.7392 mA, and 1353.3 mA/s, 5413.3 mA respectively. It means the rate of change in the current using the PID was 98 times or 9800% compared to that of the BS controller. The integral/integral time of the tracking errors for the BS and PID controllers were 0.0580^0, 0.2319^0s, and 37.8825^0, 151.5301^0s respectively. It shows that the cumulative error of the PID controller was 630 times or 63000% more than that of the BSC. An RMSE of 0.0000039096^0 resulted using the BSC and 0.0762^0 with the PID controller. It indicates that it was about 19489 times compared to that of the PID control scheme. The maximum stimulation current of 10.8019 mA was used using the BSC and 72.0630 mA with the PID controller. It shows that the maximum stimulation current utilised by the PID control scheme was six (6) times or 567% more than that of the BSC. The phase portraits in Fig. 14 indicate that the trajectory with the BS control scheme would produce a stable system with the BS control system and unstable with the PID counterpart.

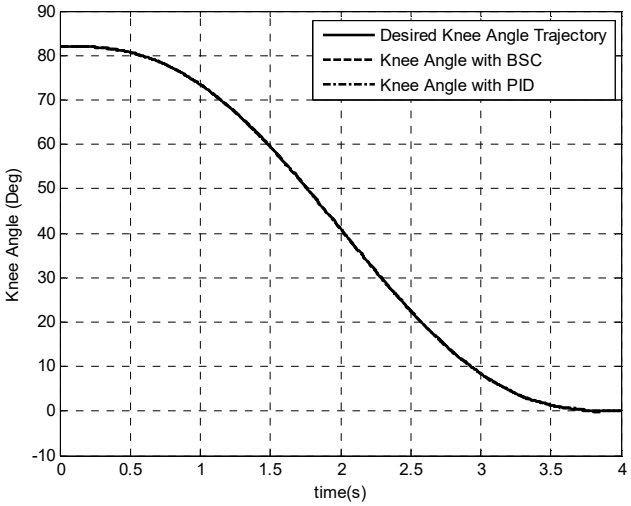


Fig. 8 Tracking Response with 53% mass.

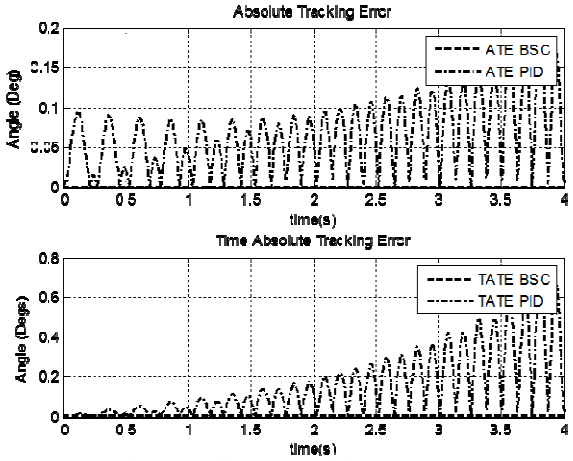


Fig. 9 Tracking Errors with 53% mass.

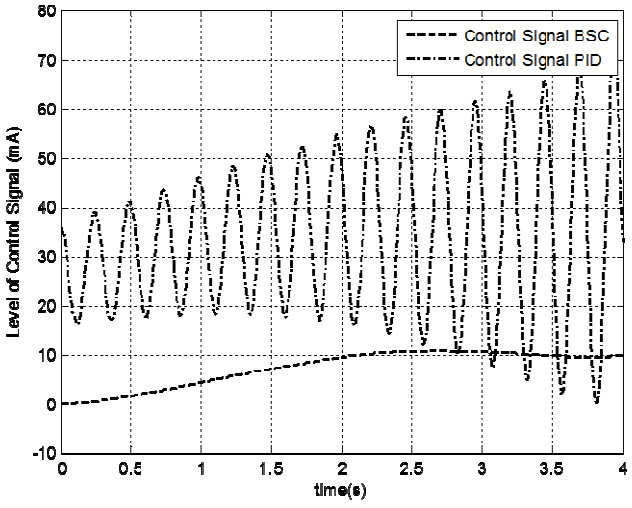


Fig. 10 Levels of Control Signals with 53% mass.

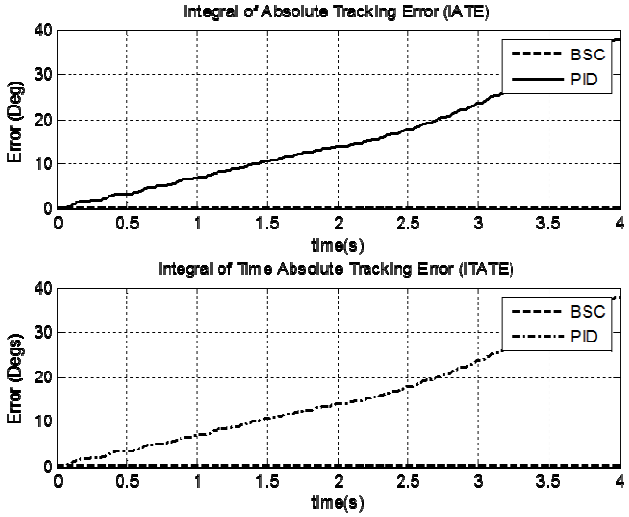


Fig. 11 Integrals of the tracking errors with 53% mass.

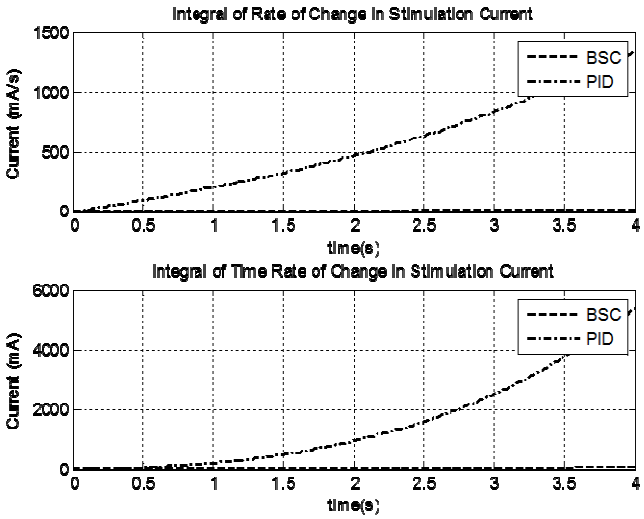


Fig. 12 Integral of the rate of change in current with 53% mass.

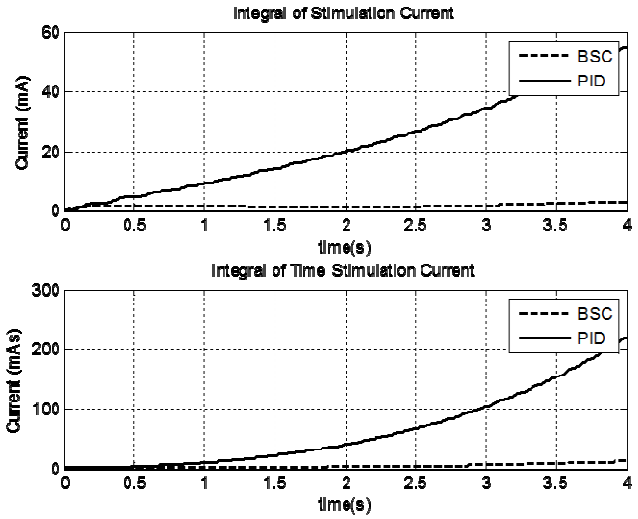


Fig. 13 Integral of stimulation current with 53% mass.

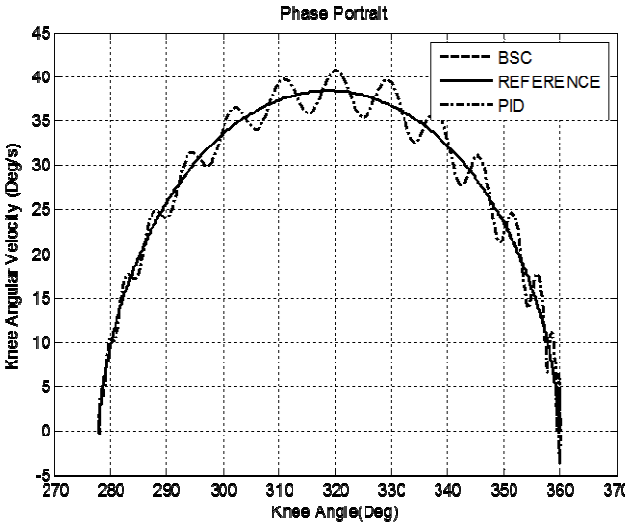


Fig. 14 Phase plot with 53% mass.

RESULTS WITH MASS REDUCED TO 53%

Parameter	Control	
	BSC	***PID***
Integrals of Stimulation/Time Stimulation Current (mA/mAs)	2.7887/ 11.1547	54.5686/ 218.2742
Integrals of Rate of Change in Current/Time Current (mA^{-s}/mA)	13.6848/ 54.7392	1353.3/ 5413.3
RMSE (Deg)	3.9096u	0.0762
Integrals of Error/Time Error (Deg/Degs)	0.0580/ 0.2319	37.8825/ 151.5301
Maximum Control Signal (mA)	10.8019	72.0630

Fig. 15. Fig. 16 and Fig. 17 are the tracking response of the knee angle during STS transition with the expected maximum amount of subject mass (126% of the subject's mass), the tracking errors, and the integrals of the tracking errors. Fig. 18, Fig. 19 and Fig. 20 are the control signals, the integrals of the rate of change in stimulation currents, and the stimulation currents during the movement. Fig. 21 is the phase plot with 126% subject mass and Table III is the tabulation of the results.

The integral/integral time of the stimulation currents for the BSC and the PID control schemes were 6.6314 mA, 26.5256 mAs, and 269.6443 mA, 1078.6 mAs respectively. The PID uses approximately forty (40) times or 3984% more current than the BS control scheme. The integral of the rate of change in the stimulation/time stimulation currents for the BS and PID controllers were 37.6315 mA/s, 150.5260 mA, and 4315.7 mA/s, 17263.0 mA respectively. It means the rate of change in the current using the PID is 114 times or 11377% compared to that of the BSC. The integral of the tracking/time tracking errors for the BS and PID controllers were 0.0580^0, 0.2319^0s, and 100.9881^0, 403.9524^0s, respectively. It shows that the cumulative error of the PID controller was 1740 times or 174000% higher than of that of the BSC. An RMSE of 0.0000039096^0 resulted using the BSC and 0.4051^0 with the PID controller. It indicates that it was about 103615 times compared to that of the BS control scheme. The maximum stimulation current of 29.3535 mA was used using the BSC and 517.2153 mA with the PID controller. It shows that the maximum stimulation current utilised by the PID control scheme was 17 times or 1659% more than that of the BSC scheme. The phase graphs in Fig. 21 shows that the BS trajectory is stable while unstable with the PID controller.

III. DISCUSSIONS

The response parameters of the FES aided STS movement without considering the effects of mass variations shows the superiority of the BSC over the PID control scheme. Results show that the cumulative current, the cumulative rate of change in current, RMSE, tracking error, cumulative tracking error, RMSE and the level of the control signal, were better than that of the PID control method. The stimulation current was higher by 3360%, the current stimulation rate of change was higher by 11300%, Integral tracking error greater by 139200, RMSE more significant by 6898100% and the maximum amount of stimulation current larger by 1200% for the PID compared to the BS control method. Additionally, the phase plane trajectory was better in the case of the BS control. The oscillatory nature of the trajectory was an indicator of the unsuitability of the PID control scheme for the plant under study (paraplegics).

Consideration given was the universal human masses distribution in evaluating the system performance. Since its application was expected to be generalised; that is, the implementation of the control system to other subjects without the need for re-tuning. The selection of 45kg as minimum and 107kg as a maximum was according to literature. The stimulation current was higher by 1850%, the current stimulation rate of change was higher by 9800%, Integral tracking error greater by 63000%. RMSE was bigger by 1948900% and the maximum amount of stimulation current more significant by 567 % for the PID compared to the BS control method for a 53% change in the subject's mass. And the stimulation current was higher by 3984% in the case of 126% mass, the current stimulation rate of change was greater by 11377%, Integral tracking error greater by 17400%. RMSE bigger by 10361500% and the maximum amount of stimulation current larger by 1659 % for the PID compared to the BS control method for a 126% change in the subject's mass. Despite the changes in mass, results indicate that the cumulative current, the cumulative rate of change in current, RMSE, tracking error, cumulative tracking error, RMSE and the level of the control signal, were better for the BS than that of the PID control technique. Results showed that similar stability properties emerged as in the case of the original subject's mass. And also confirmed the robustness of the BS control scheme over the PID in the mass variation range.

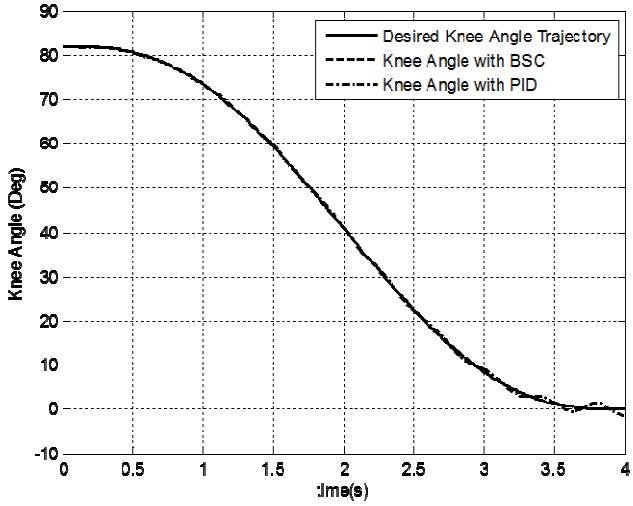


Fig. 15 Tracking response with 126% mass.

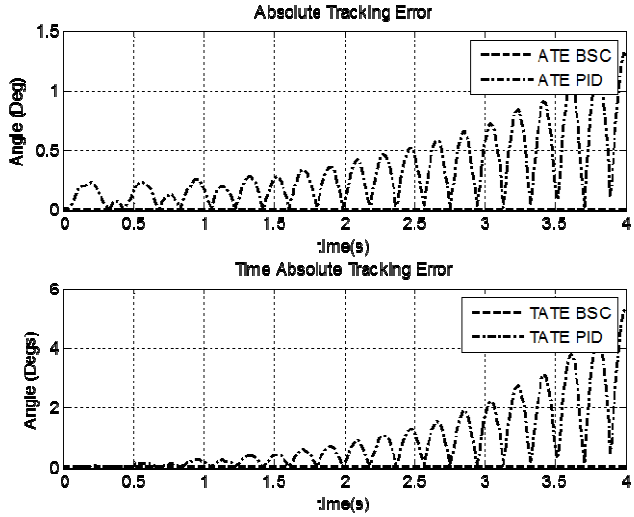


Fig. 16 Tracking error response with 126% mass.

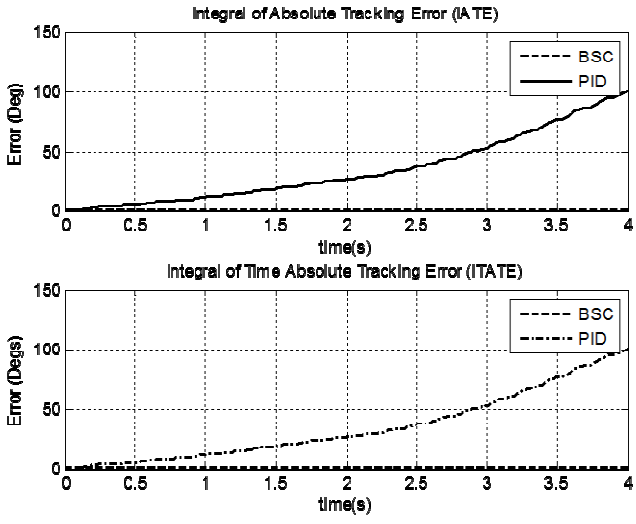


Fig. 17 Integrals of Tracking Errors with 126% Mass.

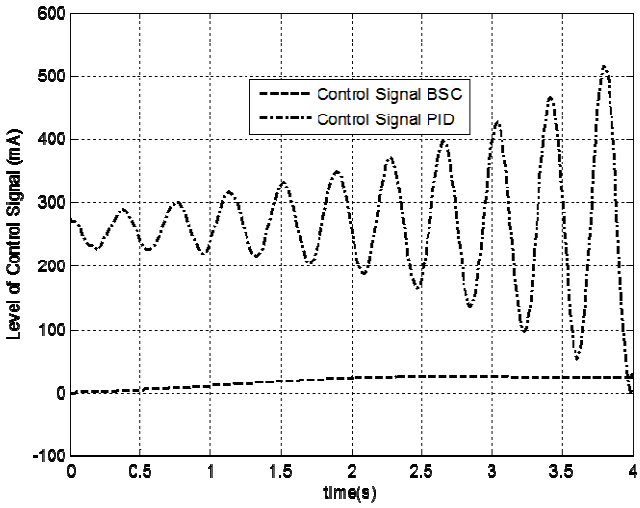


Fig. 18 Control Signal with 126% Mass.

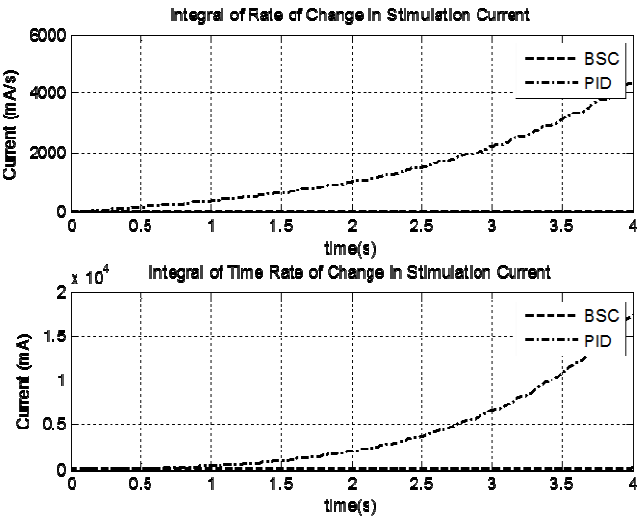


Fig. 19 Integrals of Rate of Change in Stimulation Current with 126% Mass.

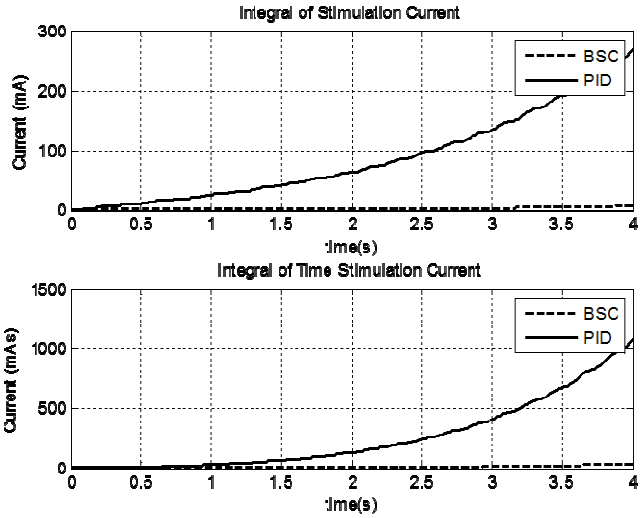


Fig. 20 Integral of stimulation current with 126% mass.

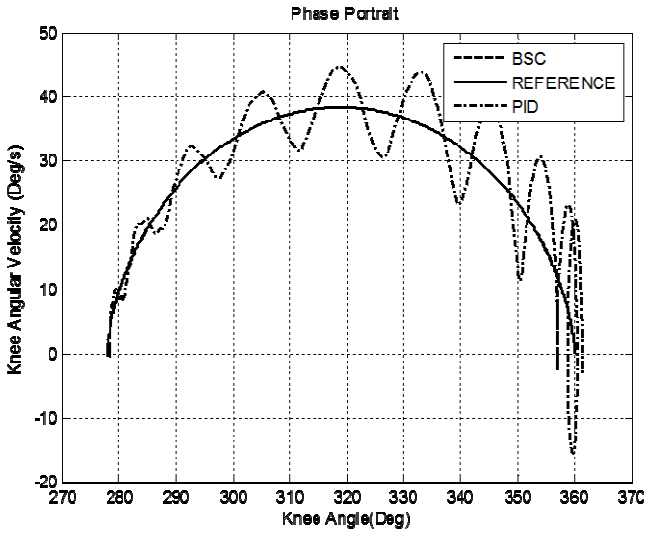


Fig. 21 Phase plot with 126% mass.

RESULTS WITH MASS INCREASED TO 126%

Parameter	Control	
	BSC	***PID***
Integrals of Stimulation/Time Stimulation Current (mA/mAs)	6.6314/ 26.5256	269.6443/ 1078.6000
Integrals of Rate of Change in Current/Time Current (mA^{-s}/mA)	37.6315/ 150.5260	4315.7/ 17263
RMSE (Deg)	3.9096u	0.4051
Integrals of Error/Time Error (Deg/Degs)	0.0580/ 0.2319	100.9881/ 403.9524
Maximum Control Signal (mA)	29.3535	517.2153

IV. CONCLUSION

According to literature, the FES aided STS movement does require improvement to pass clinical acceptance. In that regards, evaluation of the BS control scheme becomes important in the quest for a better solution. The control scheme was a nonlinear control scheme and has robustness capability. The BSC maintained superiority when compared with the PID. The BS control scheme showed a high level of ability to suppress disturbance caused by the extremes of the mass distribution globally. It was also stable. The PID control scheme showed high levels oscillations in the control signals and phase plots, which worsen in the presence of disturbances and aggravate with times, which was a treat to stability. Therefore, the proposed scheme was expected to aid tremendously in achieving clinical validity in FES application. The novelty of the study could be because the BSC was the unavailability of such for the FES induced STS manoeuvre utilisation. Additionally also, was the evaluations of the expected extremes of human global mass distribution.

ACKNOWLEDGEMENT

Special thanks of gratitude to the Centre for Graduate Studies, Universiti Tun Hussein Onn Malaysia, Faculty of Electrical and Electronic Engineering Universiti Tun Hussein Onn Malaysia, the Research Management Office Universiti Tun Hussein Onn Malaysia, Abubakar Tafawa Balewa University Bauchi, Nigeria and the Tertiary Education Trust Fund Abuja, Nigeria for their support.

REFERENCES

[1] M. J. Dolan, B. J. Andrews, and P. Veltink, "Switching curve controller for FES-assisted standing up and sitting down," *IEEE Transactions on Rehabilitation Engineering,* vol. 6, pp. 167-171, 1998.

[2] M. Poboroniuc, S. , "New experimental results on feedback control of FES-based standing in paraplegia," *AL.I.CUZA University Scientific Annals of Biophysics, Medical Physics and Environment Physics, tom.I, ,* vol. vol. III, pp. 83-89, 2007.

[3] N.-Y. Yu, J.-J. J. Chen, and M. S. Ju, "Closed-loop control of quadriceps/hamstring activation for FES-induced standing-up movement of paraplegics," *Journal of Musculoskeletal Research,* vol. 5, pp. 173-184, 2001.

[4] R. Kamnik, J. Q. Shi, R. Murray-Smith, and T. Bajd, "Nonlinear modeling of FES-supported standing-up in paraplegia for selection of feedback sensors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 13, pp. 40-52, 2005.

[5] R. Davoodi and B. J. Andrews, "Computer simulation of FES standing up in paraplegia: A self-adaptive fuzzy controller with reinforcement learning," *IEEE Transactions on Rehabilitation Engineering,* vol. 6, pp. 151-161, 1998.

[6] R. Davoodi and B. J. Andrews, "Optimal control of FES-assisted standing up in paraplegia using genetic algorithms," *Medical engineering & physics,* vol. 21, pp. 609-617, 1999.

[7] M. Huq and M. Tokhi, "Genetic algorithms based approach for designing spring brake orthosis–Part II: Control of FES induced movement," *Applied Bionics and Biomechanics,* vol. 9, pp. 317-331, 2012.

[8] F. Previdi, M. Ferrarin, S. M. Savaresi, and S. Bittanti, "Closed-loop control of FES supported standing up and sitting down using Virtual Reference Feedback Tuning," *Control Engineering Practice,* vol. 13, pp. 1173-1182, 2005.

[9] R. M. Esfanjani and F. Towhidkhah, "Application of nonlinear model predictive controller for FES-assisted standing up in paraplegia," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 6210-6213.

[10] M. Ferrarin and A. Pedotti, "The relationship between electrical stimulus and joint torque: A dynamic model," *IEEE transactions on rehabilitation engineering,* vol. 8, pp. 342-352, 2000.

[11] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: modelling, planning and control*: Springer Science & Business Media, 2010.

[12] H. K. Khalil, "Nonlinear systems, 3rd," *New Jewsey, Prentice Hall,* vol. 9, 2002.

[13] J. Liu and X. Wang, *Advanced sliding mode control for mechanical systems: design, analysis and MATLAB simulation*: Springer Science & Business Media, 2012.

[14] N. S. Nise, *CONTROL SYSTEMS ENGINEERING, (With CD)*: John Wiley & Sons, 2015.

[15] J.-J. E. Slotine and W. Li, *Applied nonlinear control* vol. 199: prentice-Hall Englewood Cliffs, NJ, 1991.

Steamblockly: A Block-Oriented Programming Environment to Enhance and Engage Interest in Software Engineering

Jerry D. Emmanuel
Institute of Computing and ICT
Ahmadu Bello University
Zaria, Nigeria
jerryemmanuel8@gmail.com

Donfack A. F. Kana
Institute of Computing and ICT
Ahmadu Bello University
Zaria, Nigeria
donfackkana@gmail.com

Ummi Abdulsalam
Software Department
Steamledge Limited
Kano, Nigeria
ummiabdulsalam@steamledge.com

Zainab A. Ladan
Steamledge Limited
Kano, Nigeria
azladan@steamledge.com

Ahmad Idris
Steamledge Limited
Kano, Nigeria
ahmadidris@steamledge.com

Latinwo Olugbenga
Steamledge Limited
Kano, Nigeria
latinwoolugbenga@steamledge.com

***Abstract*—This paper presents a block-oriented visual programming environment called Steamblockly for HTML (SFH) which is built on Google Blockly. With the aid of this block-oriented environment, programming beginners can easily write and learn HTML programs by dragging and dropping connectable visual blocks while they need not be familiar with the ever-evolving tags of HTML. The environment has 3 major components which are; Block Category, Workspace and Code Generated Area. Output of connected blocks are generated with the click of a button. As a result of its friendly user interface, the environment easily engages and effectively increases the interest of users on learning programming. This paper also presents applicable areas such as elementary schools and tertiary institutions where SFH can be used.**

Keywords—steamblockly, google blockly, block programming, visual programming environment, html

I. INTRODUCTION

Computers and their corresponding tools have become an integral part of every human being ranging from the use of smart devices, mobile phones, laptops, desktops and even autonomous vehicles. Computers have made our lives easier such that complicated and repetitive tasks are programmed to be processed quickly and efficiently with the aid of softwares or programmable hardwares [1]. Many commercial and free softwares or applications are available for doing different kinds of tasks on a computer and there are times when there might be need to developed custom standalone or web applications to meet specific needs. Mostly, this is achievable with the use of programming which can be in any language such as Hypertext Markup Language (HTML), Python, Java and C. However, this paper focuses on HTML.

Even though programming in HTML has many advantages especially with the advent of some drag-and-drop platforms such as Dreamweaver. However, learning to write several lines of syntactically correct code can be a huge task and can take a lot of time to master especially for grade level students, students in tertiary institutions and beginners, generally. This is most at times because of the complex and unfriendly UI. This is where Steamblockly for HTML (SFH) plays an important role. Developed at Steamledge Limited [2], an education and technology company in Nigeria, SFH is a Google Blockly-based Visual Programming Environment (VPE) where programming beginners can easily write and learn HTML programs by dragging and dropping connectable visual blocks while they need not be familiar with its ever-evolving tags.

a. Visual Programming Environment

Similar to an Integrated Development Environment (IDE), a Visual Programming Environment (VPE) is a programming environment that allows a user to create language-specific programs primarily through graphical manipulation [3]. Some common interaction models in VPEs are:

- Dragging blocks around a screen (e.g. Scratch [4])
- Using flow diagrams, state diagrams, and other component wiring (e.g. Pure Data [5])
- Using icons or non-text representation (e.g. Kodu [6])

Many VPEs still use text, or combine text with visual representations. Every VPE has a grammar and a vocabulary. Together they define the set of concepts that can be easily expressed with the language. The grammar is the visual metaphor used by the language: blocks, wires, etc. The vocabulary is the set of icons, blocks, or other components that allow you to express ideas [7].

b. Google Blockly

Developed by Google and first released in May 2012, Blockly is an open source library that makes it less tedious to add block-oriented visual programming to applications whether web apps or mobile apps. Blockly provides a block-based Graphical User Interface (GUI) editor and a code base for generating selected block corresponding code in text-based languages [7]. By default, the library has generators for JavaScript, Lua, PHP, Dart and Python and it's able to interface with custom-made generators for other programming languages.

On its own, Blockly is not a programming language. It provides a grammar and a representation for programming that developers can use in their apps. Code is represented by

blocks, which may be dragged around the screen. The blocks have connection points where they can be attached to other blocks and chained together as a code snippet [7]. Fig. 1 shows how blocks of code are connected to one another and the corresponding representation in JavaScript programming language. The connected blocks in Fig. 1 sets a count variable to 1. For every time the count variable is less than or equal to 3, a string "Hello World!" is alerted, after which the count variable is incremented by 1.

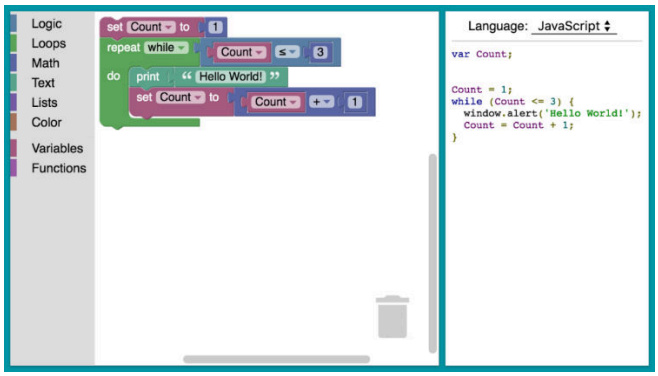


Fig. 1: Google Blockly Interface

Blockly has been used for learning and coding hardware, creating educational robot and programming animated characters on a screen. Innovatively, Blockly has been used for security purposes such as user authentication using Captcha. As displayed in Fig. 2, Instead of images, blocks are displayed to the user and the user rearranges certain blocks in a meaningful manner.

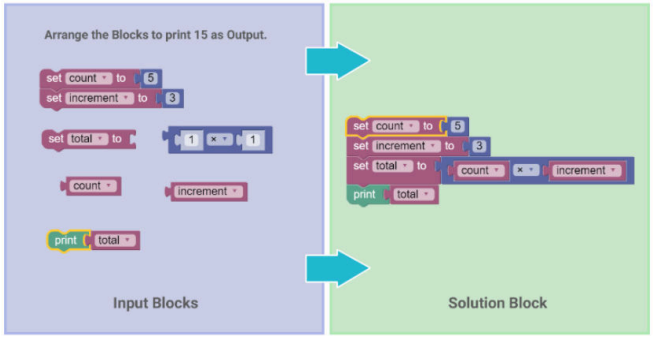


Fig. 2: Using Blockly as Captcha for Enhanced Security

Also, as displayed in Fig. 3 traditional password for user authentication has been improved using Blockly where user is expected to arrange certain blocks to form his/her password.

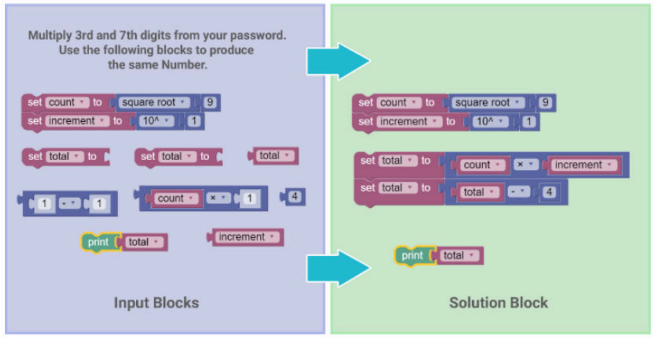


Fig. 3: Using Blockly as Captcha and Password for Secured, Fast and Easy Authentication

In addition, Blockly has been used by industries to automate minor but frequent system updates. With this the entire system does not need to be reprogrammed by a software engineer. Even if there's a need to reprogram the system, no special skill is required to do that. Fig. 4 considers a robotic paint system where the actual painting procedure changes frequently.

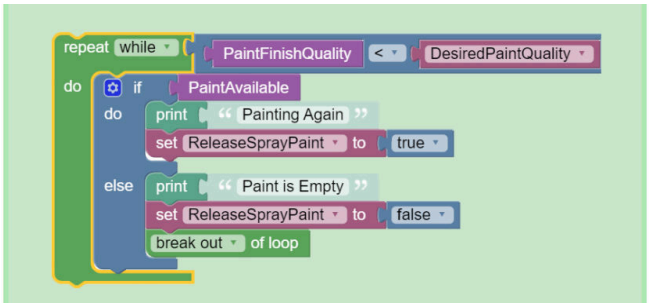


Fig. 4: Automated Paint System using Blockly

II. Related Blockly-oriented Environment

The following subsections presents related Blockly-based environments developed to perform functions such as software and hardware programming.

a. Blockly-C

Blockly-C, a Blockly-base visual programming environment translates blocks into C codes instead of JavaScript, Python, or PHP. For students, Blockly-C can increase their learning achievement and interest on C programming because it reduces the opportunity of spelling and syntax errors. Blockly-C framework is made up of three units which include C blocks, XML-to-C translator and gcc compiler. The three units are integrated together with the aid of C# WinForms which was also used in developing its User Interface (UI). As shown in Fig. 5, Blockly-C program can be compiled and executed by pressing the button labeled A, while button denoted by B loads XML of block from cloud storage or local disks either for modification or execution. Clicking the button labeled C stores the XML of blocks in edition and converts the XML into C codes [8].

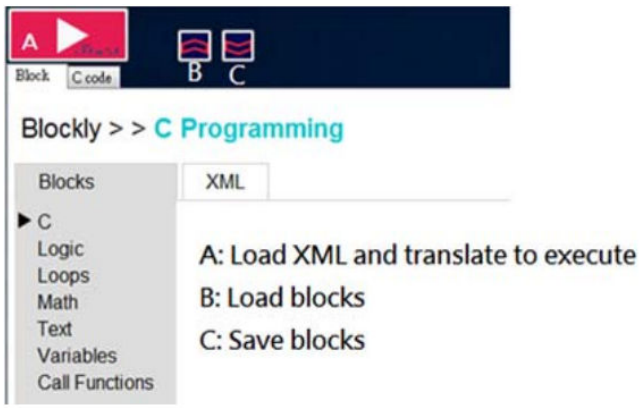


Fig. 5: User interface of Blockly-C

b. *EBot Blockly*

Ebot (Educational Robot) Blockly is a great app developed using Blockly. It provides virtual educational kits containing three platforms electronics, mechanical and software. The idea is to provide the users with basic knowledge of how each component works and how they can be combined to make a larger project. It acts as a prototype for testing new ideas without buying any real components.

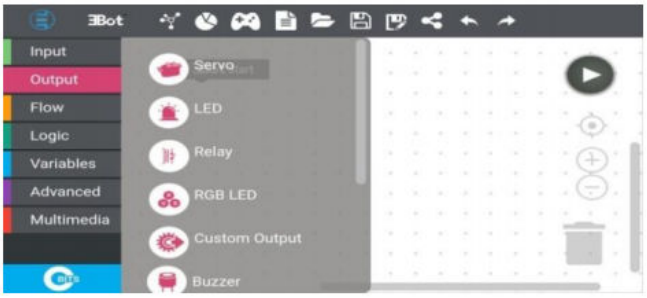


Fig. 6: Ebot Blockly - Tools and Interface

Not all users from electronics or mechanical field are aware of programming and hence it provides a graphical interface to build logic. This design can be then converted into C++ code which can be executed [9].

c. *CPF Arduino Blockly*

CPF Arduino Blockly is a great tool to learn and code hardware. It supports Arduino Leonardo via Acer Cloud to communicate with available hardware. It provides assembly instructions and remote control features to control all your devices and also helps to log reports.

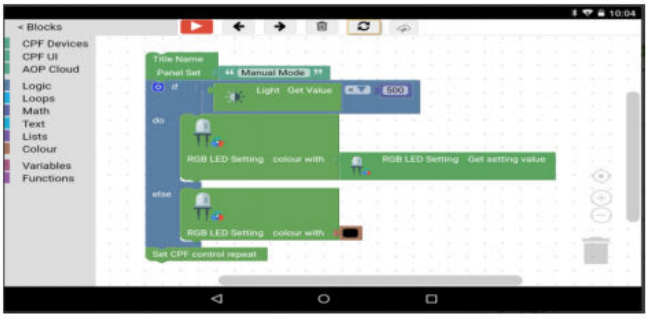


Fig. 7: Building logic using CPF Arduino Blockly

The app uses prewritten functions which are designed to eliminate the time required to operate hardware at the base level. This increases coding speed and efficiency. Since it uses Arduino as an interface between Google blockly and real hardware, it can support all hardware supported by the Arduino. All of this hardware is included under CPF devices section of the app.

III. STEAMBLOCKLY

In the following subsections, Steamblockly framework and UI are discussed extensively:

A. Steamblockly Framework

Steamblockly VPE is comprised of 3 major components as shown in Fig. 8. These include Steamblockly or SFH Blocks, HTML Generator and Generated Code UI. Steamblocky runs on web browser such as Google Chrome, Firefox, Explorer etc. In this paper, Steamblockly Block or SFH Block are synonymous, hence the phrase might be used interchangeably.

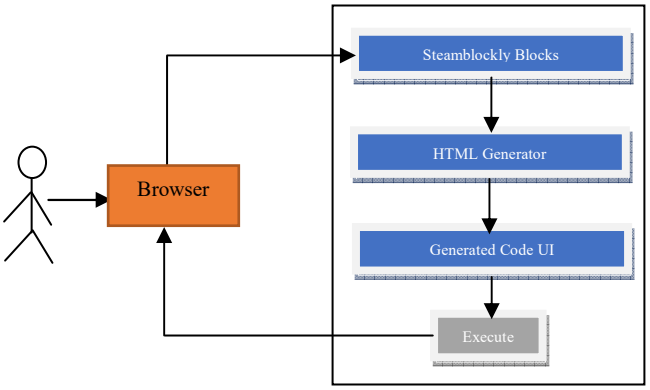


Fig. 8: Steamblockly Framework

Google Blockly has different types of blocks, and it can translate blocks into different programming languages. Users can define their blocks using two different methods. The first method involves the use of Block Factory and the second method uses Blockly API. Predefined Blockly blocks have no support for HTML, hence Steamblockly blocks were defined using the first method (Block Factory) which generate Extensible Markup Language (XML) representation for Block Categories and SFH Blocks. Some of the categories are as shown in Fig. 9.

```
<category name="Basic HTML" colour="230">...
</category>
<sep></sep>
<category name="Formatting" colour="210">...
</category>
<sep></sep>
<category name="Forms and Input" colour="290">...
</category>
<sep></sep>
<category name="Frames" colour="20">
  <block type="frames_iframe"></block>
</category>
<sep></sep>
<category name="Images" colour="260">...
</category>
<sep></sep>
<category name="Audio/Video" colour="330">...
</category>
<sep></sep>
```

Fig. 9: Steamblockly Block Categories

Block Factory is also used to define block labels, connection type and block color. All of which are represented in XML format. Whilst each block represent an HTML tag, the syntactically correct HTML code for each block is defined by the Generator. For instance, if a defined block appears with label "***Paragraph***", the HTML generator

represents the block as ** ** and displays it on Generated Code UI when dragged and dropped on workspace. The resulting web output of selected blocks is also processed by the browser.

B. Steamblockly User Interface

SFH VPE is made up of three major components, these include the Block Category, the Workspace and the Code Generated Area. Following subsections discusses these components and their functionalities.

i. Block Category

Block oriented environment is composed of Grammar and Vocabulary. The grammar is the visual description used in the environment such as blocks, wires etc while vocabulary is the set of icons, blocks, or other components that allow you to express ideas. In this paper the grammar is described using blocks while the vocabulary adopts the natural language approach. Natural language blocks use standard written sentences for most of their blocks. Blocks that use readable sentences can be used to express more complex concepts than icons, while still feeling familiar and intuitive for users. Sentences are also easier to read and more understandable than code for many users. If well designed, users of any age can find these experiences fun and challenging.

Steamblockly blocks are categorize based on their functionalities and each category is given a name and a specific color. Fig. 10 shows categories of blocks and some blocks in the Basic HTML block category.

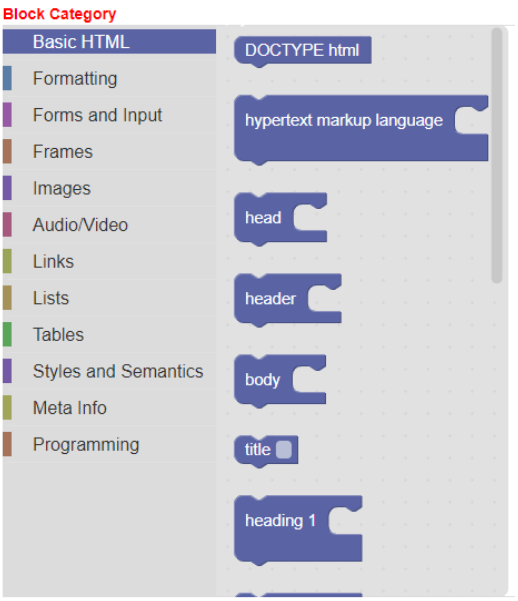


Fig. 10: Steamblockly Categories and some blocks in Basic HTML Category.

i. Workspace

When blocks are dragged from Block category they are dropped on the workspace. Workspace provides a platform for connecting blocks to each other. The platform also has a zoom-in and zoom-out feature for increasing or decreasing blocks' font size. Fig. 11 shows a set of connected blocks for a "Hello World" project. The generated code is discussed in subsequent section.

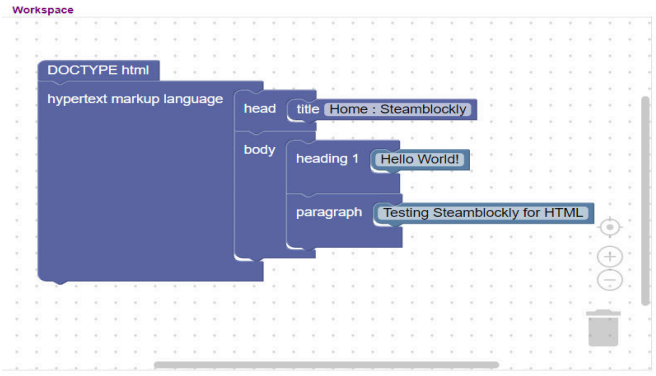


Fig. 11: Steamblockly Workspace

ii. Code Generated Area

This section displays the corresponding HTML codes for the connected blocks on the workspace. Once a block is dragged to the workspace. The HTML codes for the block is automatically generated on the Code Generated Area. The codes are appropriately structured eliminating possible syntax errors and are properly indented which makes learning interesting and easier for learners. Fig. 12 shows generated codes for connected blocks on workspace. To see the output, the user only clicks on "Run" button and the result of the webpage is as shown in Fig. 13. The entire Steamblockly interface is as shown in Fig. 14.

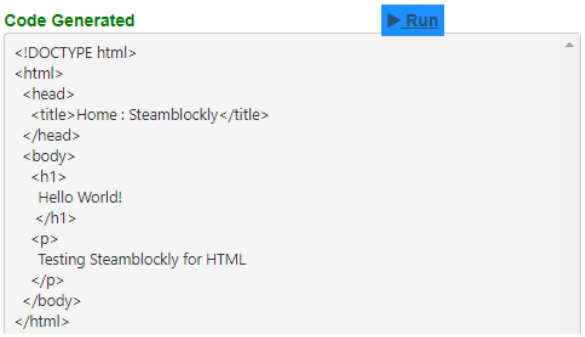


Fig. 12: Steamblockly Code Generated Area

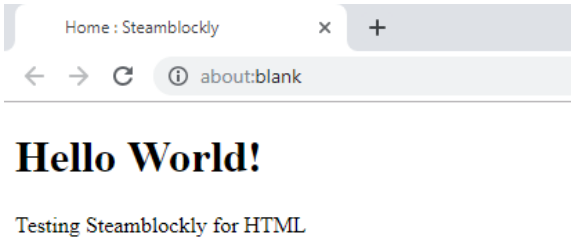


Fig. 13: Webpage Output of Steamblockly Program in Google Chrome Browser

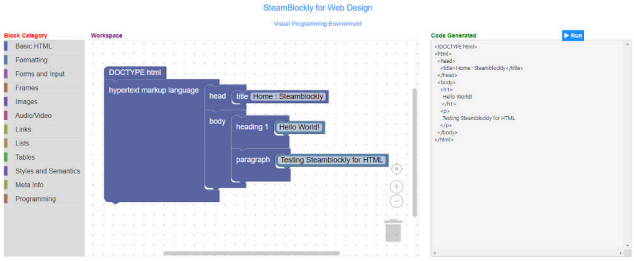


Fig. 14: Steamblockly UI with sample program

IV. AREAS WHERE STEAMBLOCKLY CAN BE USED

a. Steamblockly as Teaching aid for Elementary School Students

One of the great selling features of block-based programming is the lack of syntax errors [12]. Which means, grade level and secondary school students can develop web applications or web sites without worrying about probable errors while coding. Writing tags by the students using other IDEs is prone to errors such as syntax errors. Another thing is when these errors become quite much and frequent, grade level students could lose interest in programming easily. But with the aid of Steamblockly, students do not need to bother about the correctness of the syntax as it is handle automatically by the system. Hence, Steamblockly could serve as teaching aid for elementary students.

b. Steamblockly as Tertiary Institutions Programming Environment

Students in tertiary institutions prefer more applied fields of information and communication technologies in comparison with the informatics and programming [11]. University Computer Science students or students in related field who have never written code before simply let go of programming because of perceived complexity of writing several lines of code. To solve this problem, Steamblockly presents a user friendly UI for users especially level one students. No syntax errors and no problems with the structures of operators. All this gives confidence to students and allows them to focus on algorithms and scenarios. This can effectively and efficiently engage their interest in programming. In fact, programming instructors or teachers would find this very much helpful when teaching.

c. Steamblockly as Personal Web Development Learning Tool

One of the great strengths of Blockly is its simplicity [10]. Steamblockly being built on Blockly extends this simplicity. In most cases, block-based programming environments are Internet-based, which Steamblockly is not an exemption this allow access from anywhere, anytime and the ability to share created products with friends and followers. These features make it possible to use block-based programming for students with different level of knowledge and experience. Individuals who do not have access to Internet can also download the offline version of the software and use.

V. CONCLUSION AND FUTURE WORK

In this paper, a Blockly-based visual programming environment called Steamblockly was presented. Primarily, Google Blockly translates blocks into JavaScript, Dart, Python, or PHP. However, no provision was made for HTML. Hence, this paper presented an enhancement of this platform for HTML. This colorful and easy-to-use environment helps grade level students in learning the basics of HTML for web development. Also, the platform promises to engage and increase the conversion rate of programming beginners in tertiary institutions. Steamblockly like other block-oriented environment uses the drag-and-drop approach. Blocks are dragged from Block Category unto Workspace while corresponding code representation of the blocks are automatically generated in Code Generated Area. Resulting web page developed are seen simply by clicking a run button. Because of the evolving tags of HTML, this work focuses more of the most basic HTML as future work we intend to extend this work by adding more HTML tags especially HTML5 tags. Also, Steamblockly can be developed as visual programming environment for other programming languages such as Swift, R and Java.

REFERENCES

U. Athawale, K. Yadav, and V. Yadav, “Review of Google Blockly and its Innovative Use,” International Journal of Computer Sciences and Engineering, Vol. 6, Issue 6, 2018, pp. 1262 – 1266.

Steamledge Limited. Retrieved May 4, 2019, from https://www.steamledge.com

Visual Programming Language. Retrieved May 5, 2019 from https://en.wikipedia.org/wiki/Visual programming language

Scratch - Imagine, Program, Share. Retrieved May 5, 2019, from https://scratch.mit.edu/

Pure Data. Retrieved May 5, 2019, from https://puredata.info/23

Kodu. (2010, April 5). Retrieved May 5, 2019, from https://www.microsoft.com/en-us/research/project/kodu/

E. Pasternak, R. Fenichel, and A. N. Marshall, “Tips for Creating a Block Language with Blockly,” in 2017 IEEE Blocks and Beyond Workshop, 2017, pp. 21 – 24.

T. Y. Liang, H. T. Peng, and H. F. Li, “A block-oriented C Programming Environment,” in 2016 International Conference on Applied System Innovation (ICASI), 2016, pp. 1 - 4.

Innovative Methods, User-Friendly Tools, Coding, and Design Approaches in People - Oriented Programming Book, Accessed on: 6.05.2018.

J. Protzenko, “Pushing Blocks all the way to C++,” in IEEE Blocks and Beyond Workshop (Blocks and Beyond), 2015, pp. 1 – 5.

T. Glushkova, “Application of Block Programming and Game-Based Learning to Enhance Interest in Computer Science,” Journal of Innovations and Sustainability, Vol. 2, No. 1, 2016, pp. 21 – 32

N. Fraser, “Ten Things We’ve Learned from Blockly,” in 2015 IEEE Blocks and Beyond Workshop, 2015, pp. 49 – 50.

AN IMPROVED DIGITAL IMAGERY COMPRESSION APPROACH USING HUFFMAN AND CHINESE REMAINDER THEOREM SCHEMES

Mohammed Babatunde Ibrahim
Department of Computer Science
Kwara State University, Malete, Nigeria
Imbamok@gmail.com

Kazeem Alagbe Gbolagade
Department of Computer Science
Kwara State University, Malete, Nigeria
Kazeem.gbolagade@kwasu.edu.ng

Abstract-*The advancements in networks and information communication technology enable image transmissions across channels and links to different parties globally. However, the frequency and size of image transmitted over networks continue to limit applications area including remote sensing, databases, information conveyance, and so on. This paper prepares digital image for effective storage and transmission as may be required. First, Huffman coding is used to generate reduced size and low bits digital images. Next, Chinese Remainder Theorem is utilized to balance the pixel makeup of the resulting compressed still image in order to remove redundancy. The proposed image compression scheme experimentation included four different digital images of standard sizes such as TIF, GIF, BMP, and JPG. On the bases of Peak Noise to Signal Ratio (PNSR), Mean and Compression Rate (CR) performance parameters, the outcomes indicates that original grayscale images (TIF and GIF) outweighed colour images (BMP and JPG). The proposed compression scheme performs better than other investigative schemes on the basis of data quality recovered (PSNR) by 60.13% to 59.62% respectively. More so, the proposed scheme's performance surpasses that of motion picture (or video) RNS-based encoding scheme by 58.63% to 41.37% to respectively.*

Keywords: Compression, Huffman, Chinese Reminder's Theorem (CRT), PNSR, Mean and CR.

I. INTRODUCTION

Over the past two decades, there has been an increasing and widespread use of digital data in the largest proportion in several areas of human endeavours [1]. In 2008, International Data Corporation (IDC) projected 2.25×1021 bits of digital information was available from various transactions [2]. This volume is expected to surpass 6×1023 bits in another two decades. Precisely, computing technologies and newest information technologies have enabled the personal users to create and transmit data especially through computers, smart phones and tablets on every minute basis across the globe [3]. These have been largely connected to the growth of Internet technologies which supported massive data generation, processing, transmission and distribution [1].

Data compression is an encoding technique which reduces data size substantially for important applications such as storage systems and communication networks in highly secured manner [4,5,6]. Also, data compression methods decrease the size of information to be transmitted or stored through the process of redundancy elimination in information without loss during reconstruction of the original data. In practice, there are several file formats that can be effectively compressed including text, image, video, and audio. At large, compression implementations are broadly categorized into lossy or lossless, dictionary or non-dictionary schemes [7]. One major feature of the lossy compression schemes is that data losses occur during reversing of original input data. In contrast, lossless compression schemes retain the precise composition of the original file without data losses [8].

Digital images have been discovered a veritable means of conveying secret information which gave rise of issues such as security. Compression schemes have been applied on images for the purpose of downsizing the original file in order to attain enhanced sizes for effective communication and storages throughout across networks and devices [7]. The entire process of compression relies on rate of availability of every character encoding to construct binary depictions of original digital image files. A number of compression algorithms are available including Huffman, Lempel-Ziv-Welch (LZW), Redundant Residue Number System (RRNS) and Residue Number System (RNS). This paper proposes the implementation of Huffman and CRT compression

schemes on diverse images files formats to effectively produce secure and compressed images.

II. LITERATURE REVIEW

A. The Concept of Huffman Coding

In the case of lossless image compression, Huffman coding is one of the most deployed schemes. It has been noted as effective for image compression in some degrees [9]. The operation of Huffman encoding commences by determining the probability of every image symbol. The probabilities of these symbols are organized into a descending format giving rise to tree leaf nodes. In the event of individual coding of the symbols, the Huffman code is constructed by integrating the small probable symbols. Thereafter, the operation is continued until two probabilities of two composite symbols remained. Consequently, Huffman codes are realized from labels of the code tree generated earlier. The shortest length binary code for a two-symbol source is usually between the symbols 0 and 1. In fact, the Huffman codes depicting the symbols can be generated by considering the branch digits in sequence starting at the root node to the corresponding terminal leaf or node. Coding redundancy is minimized by means Huffman coding as illustrated in Fig. 1 [9].

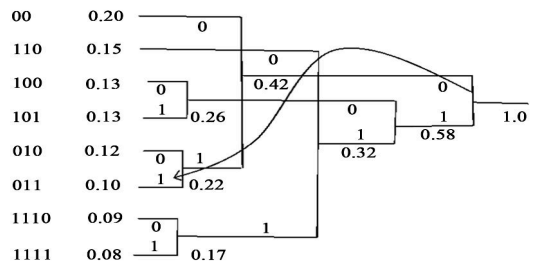


Fig. 1. Huffman coding operational structure. [10]

Huffman coding process relies on these assumptions. (a) most reoccurring symbols are assigned smaller code words against the symbol with reoccurring least; and (b) the two symbols reoccurring least are the equal length.

The regular code length is released by the determining the average of the multiplication of probability of the symbol and corresponding amount of bits encoded. The study in [11,12] offered more insights. Whereas; the efficiency of Huffman code is determined by dividing the entropy against the average length. The entire of process of Huffman builds the optimal code for a collection of probabilities and matching symbols on the premise that the symbols are simultaneously encoded [9].

B. The Concept of Residue Number System

Residue Number System (RNS) is composed of a collection of moduli but, they are independently distributed. An integer is used to the modulus and arithmetic processes of every residue separately. The benefits of choosing the RNS as against the conversional system span from fault tolerance, carry-free operation, parallelism to modularity. These attributes makes RNS to be top choice in Digital Signal Processing (DSP) applications including convolution, digital filtering, image processing, and fast Fourier transform [13].

Let $\{m_1, m_2, m_3, ..., m_n\}$ be a set of positive integers all greater than 1. m_i is known as a modulus, and the n-tuple set $\{m_1, m_2, m_3, ..., m_n\}$ is called a moduli set. Suppose that an integer number Y for every modulus in $\{m_1, m_2, m_3, ..., m_n\}$, then, $y_i = Y \bmod m_i$, (which will be depicted as $|Y|_{mi}$). Thus the number Y in this system is represented as Y = (y_1, y_2, y_3, …, y_n), $0 \le y_i \le m_i$. To minimize redundancy, the moduli set must be pairwise or relatively prime. Thus, gcd (m_i, m_j) = 1 for $i \neq j$, where gcd means the largest common divisor of m_i and m_j .

Let $M = \prod_{i=1}^{n} m_i$, then the RNS depiction is exclusive for given integer Y $Y \in [0, M-1]$. M is the dynamic range [13,14].

RNS architectures are normally made up of three major components including, a binary-to-residue converter, residue arithmetic units, and a residue-to-binary converter as shown in Fig. 2. The residue-to-binary converter is the most composite component in RNS structure. One modification has been proposed to the RNS design especially, speedy conversion process in order to minimize the incidence of the RNS fast arithmetic, though, a slow conversion procedures. The majority of available reverse converters are offshoot of Chinese Reminders Theorem (CRT) or the Mixed Radix Conversion (MRC). The reason for the popularity of CRT is due parallelized computation against MRC, which is a sequential procedure with a large amount of arithmetic operations. The key setback for CRT is that, it needs modulo-M

operation on the assumption that M is a large number. By nature, the procedure is time inefficient and expensive with focus on area end energy dissipation. The New CRT eliminates the drawbacks of either the conventional CRT or MRC schemes giving rise to notable improvement in the conversion schemes or algorithms [15]. The usability of the New CRT replies on the selected set of moduli [16]. Due to the merits of CRT, it will be adopted in this research paper.

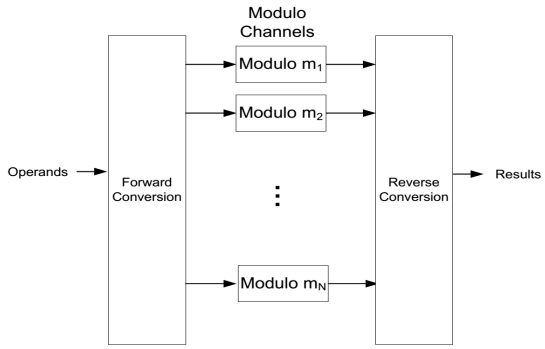


Fig. 2. General structure of an RNS processor. [17]

C. Chinese Reminder Theorem (CRT)

The hypothesis of numeric theory, which is expressed as, in case one knows the fragments of the division of an entire number n by a couple of numbers, one can choose curiously whatever is left of the division of n by the consequence of these numbers, under the condition that the divisors are pairwise co-prime [13,18] is known as the Chinese Reminder Therorem (CRT). The CRT will decide a number n that at the point when divided by some given numbers (divisors) leave given fragments [17]. Assume $n_1,...,n_k$ be entire numbers more prominent than 1, which are consistently called moduli or divisors. Assume N implies the consequence of n_i. The CRT hypothesis declares that if n_i are pairwise co-prime, and if $a_i,...,a_k$ are numbers with the result that $0 \le a_i < n_i$ for each *i*, at that point there is one and just a single whole number x, to such an extent that $0 \le x < N$ and the rest of the Euclidean division of x by n_i is a_i for each *i*.

Given a moduli set $\{m_1, m_2, m_3, ..., m_n\}$ with $\gcd(m_i, m_j)$ =1 for $i \neq j$ and dynamic range $M = \prod_{i=1}^{n} m_i$, with respect to CRT an integer Y denoted with the RNS illustration is $(y_1, y_2, y_3, ..., y_n)$ can be altered from its residue form as $Y = \left| \sum_{i=1}^{n} M_i \left| M_i^{-1} y_i \right|_{mi} \right|_M$,where $M_i = \frac{M}{m_i}$ and M_i^{-1} is the multiplicative inverse of M_i with respect to m_i.

The application of the known moduli sets was used along with RNS when integrated with LZW for securing an efficient data encoding and decoding. MATLAB was used for simulation of few selected documents to ascertain the efficiency of the proposed LZW-RNS by [7] showed an enhanced performance above 15% (15.6250kb) with decrease in data size, a rise of 40% (37kb) in the original file size as well as 2.8645s in computational time than the original scheme.

An improvement in the conventional Huffman's data encoding algorithm was proposed by applying RNS [19]. The study presented a conventional Huffman's encoding which was used to enhance data compression in which the rate of recurrence of every character is used to generate binary codes. Similarly, [20] improved the security of image in digital form using Arnold Transforms and RNS for encoding of image. Digital images observations have various applications which require adequate security across transmission media such as networks.

Another study by [21] identified the capability of the 3G networks to revolutionize data and information sharing among peoples in an extraordinary style as compared to the networks of 2G and 2.5G. Consequently, the 3G was developed using KASUMI block cipher that is incapable of performing error detection and correction. Therefore, RNS was introduced as a technique based on modulus projection method which reduced considerably the computation overhead for RRNS codes and decoding. Furthermore, this challenge was dealt with using hybrid scheme requiring integer recovery process according to [22].

The Redundant Residue Number System (RRNS) was exploited in wireless networks and mobile systems for data sharing established on peer-to-peer protocol. The recognition and rectification of error attributes were applied with multi-level RNS to advance an innovative method which performed considerably superior [23].

In [24], a hybrid of LZW and Huffman image compression was developed, where the results

showed that CR of 47.61% and PSNR of 92.76% of the original image. The focus of this paper is on the implementation of RNS and Huffman compression schemes on digital images for the purpose of image size reduction, increased reliability and security as well as faster execution time or usage performance.

III. RESEARCH METHODOLOGY

The compression algorithms experimentations were carried out on MATLAB in two stages for different images and dimension as shown in Fig. 3.

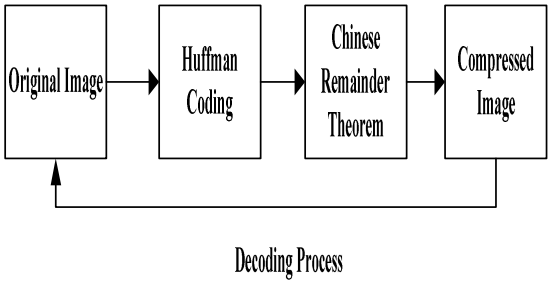


Fig. 3. The structure of the proposed image compression scheme.

In Fig. 3, the process of encoding the original images begins with Huffman coding; and CRT is applied to obtain the compressed images. The processes of retrieving the original image data after complete compression operation start at reversing from the last stage of encoding the original image. In the first stage, the entire procedure of the image compression using Huffman coding is described in Table 1.

TABLE I. HUFFMAN CODING IMAGE COMPRESSION PROCEDURE

Input: Grayscale/RGB images Output: Partially compressed image	
1	**INPUT** dimension of grayscale and RGB image files: 512x512
2	**CONVERT RGB image to grayscale image**
3	**RESIZE** images to new dimension to remove irrelevant sections: 401x391
4	**CALCULATE** probabilities of the resized image and sort probabilities from lowest to highest as '1' and '0' respectively.
5	**STORE** the resultant character dictionary as nodes of tree and create priority queue.
6	**ENCODE** map for the character sequence and associated map.
7	**COMPRESS** output file using encoding map as header for decoding purpose.
8	**GENERATE the image file as compressed file**.
9	**OUTPUT** the 8 bits sequence stored in the map.

The second stage of the image compression makes use of the Huffman coding partially-compressed image file with Chinese Reminder Theorem (CRT) as contained in Table 2.

TABLE II. CHINESE REMAINDER THEOREM IMAGE COMPRESSION PROCEDURE

Input: Huffman partial-compressed image Output: Fully compressed image	
1	**INPUT** dimension of compressed file: 401x391
2	**GENERATE** moduli set and compute the product *M = m1*m2*m3… mk*
3	**COMPUTE** the product of the remaining moduli for each modulus *m1 = M/m1*
4	**REDUCE** the product modulo of given modulus to obtain reduced numbers s1, s2, s3, … , sk
5	**FIND** the reciprocal modulo of the reduced numbers as *t1, t2, t3,…, tk*. From *s1x* = 1(mod *m1*).
6	**GET** the sum *of r1s1t1 + r2s2t2 + … + rksktk.*
7	**REDUCE** the sum modulo ***M*** from step 1 to attain the CRT ***N***
8	**OUTPUT** the fully compressed image based on Huffman and CRT schemes.

The original image decompression involves the process of decoding the Huffman coding and enhancements applied by the CRT as illustrated in Fig. 4.

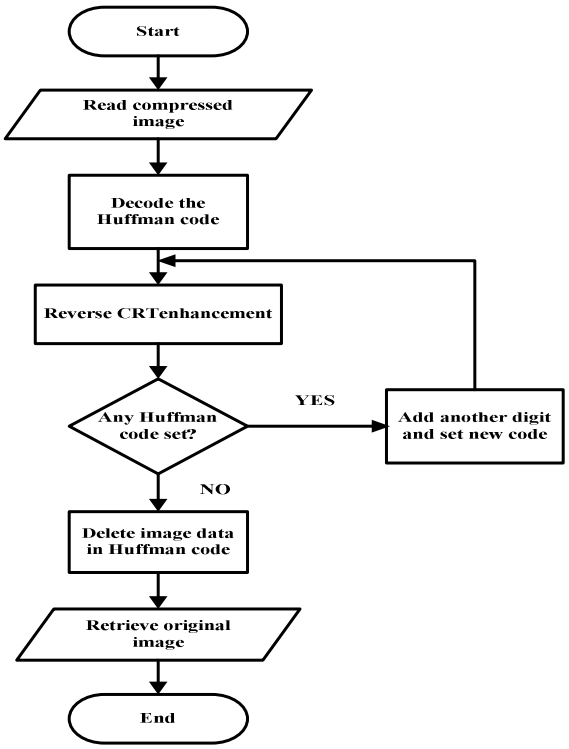


Fig. 4. The original image decoding process

The evaluation parameters used to determine the strength and effectiveness of two compression algorithms understudy include:

Quality: This can be attained with two parameters Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). These calculate the rate of deviation of original image and decompressed image. The cumulative square error between decompressed image and original image is the mean square error (MSE), while peak signal to noise ratio denotes the peak error of quality. The values of MSE and PNSR must lower and higher respectively whenever decompressed image quality outperforms the original image as expressed in Equation 1.

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}(C(x,y) - S(x,y))^2 \quad -- \quad 1$$

where,
(x, y) are image's two coordinates,
(M, N) are image's two dimensions,
$C(x, y)$ is decompressed image generation function, and
$S(x, y)$ is the original image generation function.

PSNR is used to determine the quality of image. This reveals that decompressed image with higher value of PSNR possesses high quality. PSNR is measured in decibels (dBs) or 8 bits per pixel between two images given by Equation 2:

$$PSNR = 10\log_{10}\left[\frac{MAX^2}{MSE}\right] \quad ---- \quad 2$$

where,
MAX^2 is the maximum value of pixel in original image, and
MSE is Mean Square Error risk function of cumulative square error between original image and compressed image.
Compression ratio: This refers to compression power which is realised from the size of data illustration formed by means of algorithms used in compressing. It is measured as the ratio between of the real size of image and compressed size of image given by Equation 3.

$$CR = \frac{US}{CS} \quad ---- \quad 3$$

where,
CR is compression ratio of image
US is the size of uncompressed image
CS is the size of compressed image
Therefore, the amount of space savings which is the reduction in size compared to the size of the uncompressed image given by Equation 4.

$$Space\ savings = \left(1 - \frac{US}{CS}\right) \times 100\% \quad -- \quad 4$$

whenever, the uncompressed data rate is determined, the compression ratio can be realized from the compressed data rate. The amount of redundancy in original data can be calculated in percentage, which is space savings rate.

IV. RESULTS

In this paper, the implementation of the image compression approach that combined the independent concepts of Huffman coding and Chinese Remainder Theorem are presented in two major phases for four kinds of image types including GIF, TIF, JPG and BMP. In the first phase, the colour images were converted to grayscale image to produce defined standard colourspace and reduce the human-perceived achromatic intensity characteristics similar to the original grayscale images [25]. The resulting images serve as original image input data for the Huffman coding; and thereafter Chinese Remainder Theorem procedure is applied to produce completely compressed images. For the demonstration, the GIF uncompressed image format was used as input data as shown in Fig. 5.

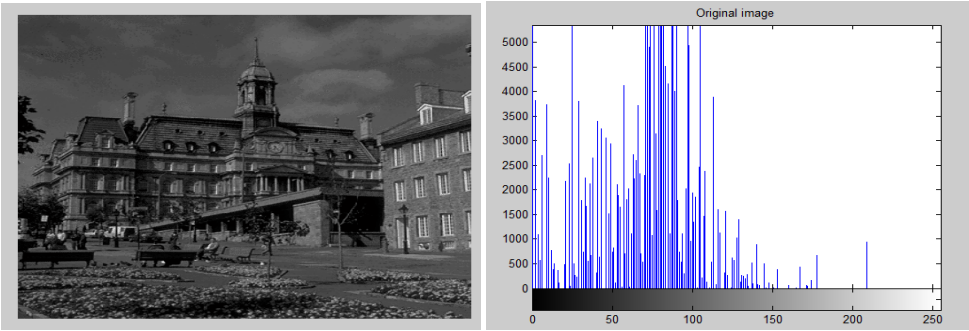


Fig. 5. Uncompressed or original image and histogram

In Fig. 5, the image type composition is a standard grayscale type displayed in darkish black and white. This image contains huge redundancy and irrelevant contributing to large size obtained, which is expected to be removed during compression procedure. The next stage involves the application of both concept of Huffman coding and Chinese Remainder Theorem for the complete image size reduction to release compressed image of the original image as shown in Fig. 6.

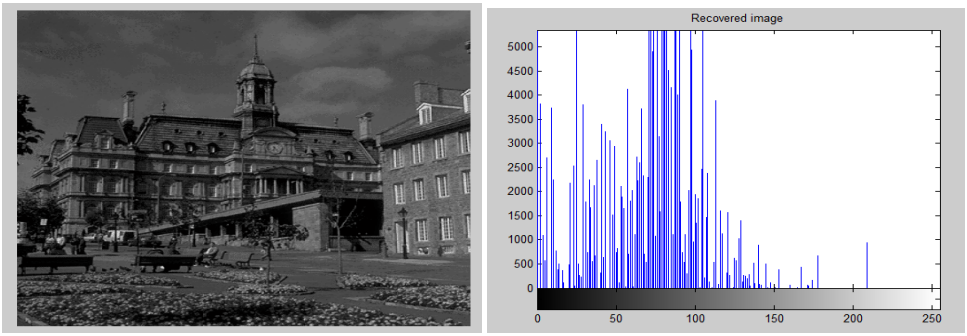


Fig. 6. Compressed image and histogram.

In Fig. 6, the difference between the original or uncompressed image and the compressed image is minimal except for slight increased luminance of the original image. These stages of compression were applied on the remaining image samples understudy. The outcomes of compressing the various images types, original sizes, compressed sizes after Huffman coding and CRT procedures for the selected parameters are presented in Table 3.

TABLE 3: OUTCOMES OF COMPRESSION PROCEDURE OF HUFFMAN AND CRT SCHEMES

Image type	Original size (KB)	Compressed size (KB)	CR (%)	PSNR (%)	MSE (%)	Space saved (%)
GIF	284	196	1.45	60.13	30.80	44.90
TIF	121	87	1.39	59.91	27.09	39.08
JPG	202	156	1.29	38.70	28.82	29.49
BMP	385	292	1.32	35.89	29.01	31.85

In Table 3, the quality of data realized from compressed images is best for GIF and TIF. This is the same for the amount of space saved after successful compression of image of the type GIF and TIF. Reason being that grayscale images can be directly manipulated by the proposed compression scheme without further enhancements operations. Conversely, colour images have been found to perform poorly when compared to those of grayscale. The original images (that is, JPG and BMP) require a conversion operation into grayscale level before compression scheme is applied on the resulting images. In effect, there are no significant differences between compressed and uncompressed images representations as indicated by histogram plots in Figures 6 and 7. This is attributed to the relative distribution of bits in both images (uncompressed and compressed images files); though, the compressed images have closeness of bits after removal of redundancy that gave them higher luminance. The overall performance of the proposed image compression scheme revealed proportionate reduction in the sizes of original images as well as image data retention for the recovered images. However, the proposed scheme is more efficient when compared to the hybrid of Huffman and LZW compression scheme [24] in terms of quality of recovered data (PSNR) by 60.13% to 59.62%. Consequently, the compression operation offered by the proposed scheme for space savings, CR, PSNR and Mean after recovery of original data was better than the previous schemes. Similarly, the value of PSNR for motion picture (video files) RNS-based encoding scheme in [26] falls short of the results obtained in this paper by 41.37% to 58.63% respectively.

V. CONCLUSION

The study has revealed that image compression can be substantially attained by means of large redundancy removal. The compression scheme is composed of compression and decompression method based on Huffman coding and CRT, which provided more effective compression outcomes for the sampled still images. Aside time efficiency, large size images can benefit from the proposed image compression scheme. The process of decompressing the still images showed significant similarities between the recovered images and their respective initial images; it can be said that the proposed still image compression scheme is a lossless compression approach. There is need to deploy other compression algorithms with CRT schemes on other image files formats (such as WebP, AVIF and HIEF) for better outcomes of image size reduction and data retention.

REFERENCES

[1] H. Qui, “An Efficient Data Protection Architecture Based on Fragmentation and Encryption”. Unpublished Ph.D. Thesis, Institute of Sciences and Technologies, Paris Institute of Technology, Paris, France, 1-136, 2017.

[2] F. Berman, Got Data? A Guide to Data Preservation in the Information Age. Communications of the ACM, vol. 51, no. 12, 50–56, 2008.

[3] V. Mayer-Schönberger and K. Cukier, “Big data: A revolution that will transform how we live, work, and think”. Houghton Mifflin Harcourt, 2013.

[4] V. Cappellini, “Data Compression and Error Control Techniques with Applications”, Third Edition, Academic Press, London, ISBN: 0-8194-2427-7, pp: 9-37, 1989.

[5] A. Ammar, A. Al-Kabbany, M. Youssef and A. Emam, “A Secure Image Coding Scheme using Residue Number System”, Proceedings of the 18th National Radio Science Conference, (NRCS`01), Egypt, pp: 339-405, 2001.

[6] N. Saud, N. Rameez, A. Raja Ali and S. Faisal, “Optimized RTL Design and Implementation of LZW Algorithm for High Bandwidth Application”, pp 279-285, 2011.

[7] A. Alhassan, K. A. Gbolagade and E. K. Bankas, “A Novel and Efficient LZW-RNS Scheme for Enhanced Information Compression and Security”. International Journal of Advanced Research in Computer Engineering & Technology, vol. 4, no. 11, 4015-4019, 2015.

[8] B. Vijayalakshmi and N. Sasirekha, “Lossless Text Compression for UNICODE TAMIL Documents”, ICTACT Journal on Soft Computing, vol. 8, no. 2, 1635-1640, 2018.

[9] J. H. Pujar and L. M. Kadlaskar, “A New Lossless Method of Image Compression and Decompression using Huffman Coding Techniques”, Journal of Theoretical and Applied Information Technology, vol. 15, 18-23, 2010.

[10] V. G. Dubey and J. Singh, “3D Medical Image Compression Using Huffman Encoding Technique”, ECE Department, RIEIT Railmajra, SBS Nagar, Punjab, International Journal of Scientific and Research Publications, vol. 2, no. 9, 2012.

[11] K. Gupta, R. L. Verma and Md. Sanawer Alam, “Lossless Medical Image Compression Using Predictive Coding and Integer Wavele Transform based on Minimum Entropy Criteriat”. International Journal of Application or Innovation in Engineering & Management (IJAIEM), vol. 2, 98-106, 2013.

[12] K. Mishra, R. L. Verma, S. Alam and H. Vikram, “Hybrid Image Compression Technique using Huffman Coding Algorithm”. International Journal of Research in Electronics & Communication Technology, vol. 1, 37-45, 2013.

[13] K. A. Gbolagade and S. D. Cotofana, “Residue-to- Decimal Converters for Moduli Set with Common Factors”, 52nd IEEE International Midwest Symposium on Circuits and Systems, pp 264-627, 2009.

[14] M. Lu, “Arithmetic and Logic in Computer Systems”, John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

[15] Y. Wang, “Residue-to-binary converters based on new chinese remainder theorems”, IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing, vol. 47, no.3, pp. 197-205, 2000.

[16] K. A. Gbolagade, “Effective Reverse Conversion in Residue Number System Processors”. PhD Thesis. The Netherlands, 2010.

[17] A. Omondi and B. Premkumar, “Residue Number Systems-Theory and Implementation”, Imperial College Press Advances in Computer Science and Engineering: Texts vol. 2, 2007.

[18] B. Arutselvan.and R. Maheswari, “CRT Based RSA Algorithm For Improving Reliability And Energy Efficiency With Kalman Filter In Wireless Sensor Networks”, International Journal of Engineering Trends and Technology (IJETT), vol. 4, no. 5, pp 1924-1929, 2013.

[19] A. Alhassan, I. Saeed and P. A. Agbedemnab, “The Huffman’s Method of Secured Data Encoding and Error Correction using Residue Number System (RNS)”, Communication on Applied Electronics (CAE) Journal, Foundation of Computer Science (FCS), vol. 2, no. 9, pp. 14-18, New York, USA, 2015.

[20] S. Alhassan and K. A. Gbolagade, “Enhancement of the Security of a Digital Image using the Moduli set {2n-1, 2n, 2n+1}”, International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), vol. 2, no. 7, pp 2223-2229, 2013.

[21] H. Mahyar, “Reliable and High Speed KASUMI Block Cipher by Residue Number System Code”, World Applied Sciences Journal, vol. 17, no. 9, pp 1149-1158, 2012.

[22] K. A. Amusa and E. O. Nwoye, “Novel Algorithm for Decoding Redundant Residue Number System (RRNS) Codes”, IJRRAS, vol. 12, no. 1, pp 158-163, 2012.

[23] A. Barati, M. Dehgan, A. Movaghar and H. Barati, “Improving Fault Tolerance in Ad-Hoc Networks by using Residue Number Systems”, Journal of Applied Sciences, vol. 8, pp 3272-3278, 2008.

[24} F. A. Ajala, A. A. Adigun and A. O. Oke, “Development of Hybrid Compression Algorithm for Medical Images using Lempel-Ziv-Welch and Huffman Encoding”. *International Journal of Recent Technology and Engineering*, vol. 7, no. 4, pp. 1-5, 2018.

[25] C. Saravanan, “Color Image to Grayscale Image Conversion”, 2nd International Conference on Computer Engineering and Applications, vol. 2, pp 196-199, IEEE, 2010.

[26] C. N. Vayalil, M. Paul and Y. Kong, “A Residue Number System Hardware Design of Fast-Search Variable-Motion-Estimation Accelerator for HEVC/H.265”, IEEE Transactions on Circuits and Systems for Video Technology, 29(2), 572-581, 2019.

Neuro-fuzzy ensemble techniques for the prediction of turbidity in water treatment plant

S.I. Abba
Department of PPD and M
Yusuf Maitama Sule University
Kano, Nigeria

Rabiu Aliyu Abdulkadir
Department of Electrical and Electronic Engineering
Kano University of Science and Technology
Wudil, Kano, Nigeria

M.S. Gaya
Department of Electrical and Electronic Engineering
Kano University of Science and Technology
Wudil, Kano, Nigeria

M.A. Saleh
Department of Electrical and Electronic Engineering
Kano University of Science and Technology
Wudil, Kano, Nigeria

Parveneh Esmaili
Department of Electrical and Electronic Engineering
Near East University
TRNC

Mustapha Bala Jibril
Department of Electrical and Electronic Engineering
Near East University
TRNC

***Abstract*—providing a satisfactory and reliable prediction tool for Turbidity in water treatment plant is quite an essential task for various environmental and public health perspective. In this paper, a neuro-fuzzy approach is developed using two different optimizations of fuzzy inference system (FIS) (i.e. hybrid and backpropagation) to predict the treated Turbidity at Tamburawa water treatment plant (TWTP). Subsequently, a neuro-fuzzy ensemble technique was applied to improve the performance of the two optimizations. For this purpose, the daily recorded data of turbidity ($Turb_R$) (µs/cm), conductivity ($Cond_T$) (mS/cm), total dissolve solid (TDS_R) (mg/L), chloride (mg/L) and suspended solid ($SS_{R)}$ (mg/L) and Hardness ($Hardness_T$) (mg/L) from TWTP were obtained. The predictive models were evaluated based on two numerical indicators (determination coefficient and root mean square error). The obtained results indicated that neuro-fuzzy hybrid increased the performance accuracy of neuro-fuzzy backpropagation optimization up 16% and 15% in both the training and testing phase respectively. For neuro-fuzzy ensemble results, the performance proved that hybrid ensemble increased the prediction efficiency of backpropagation ensemble up to 18% in the testing phase. Hence, for the prediction of Turbidity in TWTP both the hybrid FIS optimization and ensemble hybrid FIS optimization showed excellent accuracy while for its recommended to employed ensemble techniques in case of backpropagation FIS optimization. The ensemble methodology proved to be implemented as a real-time prediction model that can provide a brilliant approach for environmental sustainability.**

Keywords— water treatment plant, ensemble techniques, neuro-fuzzy, turbidity, Tamburawa-Kano

VI. INTRODUCTION

The major function of municipal water treatment plant (MWTP) is to supply healthy drinking water, void from impurities for public utilization[1]. MWTP is a set of processes employed to eliminate contaminant such as solids, bacteria, and micro-organisms from raw water meant for public consumption. An efficient water treatment plant (WTP) is necessary in order to satisfy drinking water standards set by law. Accurate operation and control of WTP are highly challenging because of many factors including biological, physical, chemical and biochemical parameters present in the process[2]. If the standard limits of the parameters in WTP are not well maintained, there is going to be a severe health and environmental issues. Therefore, appropriate control techniques must be used to keep these parameters at their desired limits. One way to achieve suitable control is to model WTP performance by employing robust modeling tools based on salient parameters of the process[3].

Nevertheless, modeling of a complex process such as WTP is nontrivial, this is because of the dilute mixture of several proportions in addition to non-linear processes (physical, chemical, biological and biochemical) exhibited by WTP [4]. Therefore, simulating WTP with mathematical and linear models result in unreliable performance [5]. Over the years, several linear models have been introduced in the literature to represent the performance of WTP. However, these models are generally less accurate, expensive and time-consuming (Solgi et al., 2017; Suen and Eheart, 2003). With recent developments in artificial intelligence (AI) based modelling, black-box models like Support Vector Machine (SVM), Artificial Neural Network (ANN) and Adaptive

neuro-fuzzy inference system (ANFIS) have provided an avenue for modelling non-linear systems with high level of accuracy [6],[7],[8].

Based on the conducted literature survey, it is found that many studies have been carried out with ANN models to study the performance WTP. Although there are promising results from these models their accuracy is affected by parametric uncertainties. Hybrid models such as adaptive neuro-fuzzy inference systems (ANFIS) which combined the learning capability of ANN and robustness of fuzzy systems have provided better performance in comparison with individual models. The study of other AI models like SVM for modeling of WTP is rare in the literature. Although there is no consensus on the best AI model for modeling WTP, it is believed that using ensemble strategy would produce higher performance. The ensemble methods have been widely applied to model several engineering processes [9],[10],[11],[12]. In this regard, [10], [11] demonstrated the use of ensemble approach to predict Biological Oxygen Demand (BOD) and several WQ parameters. Similarly, [2] proposed AI-based ensemble model for the prediction of WTP performance. The prediction result is better than the individual AI-based models. In this paper, a neuro-fuzzy ensemble technique is proposed to predict the treated Turbidity at Tamburawa water treatment plant (TWTP). The approach is developed by using two different optimizations of fuzzy inference system (FIS) (i.e. hybrid and backpropagation).

VII. Materials and Methods

A. Tamburawa water treatment plant (TWTP)

Kano State is located in the northern part of Nigeria. The state has a total landmass of 20,131 km^2. It is bounded by Katsina on the west, by Kaduna State on the south-west and by Jigawa and Bauchi on east and southeast respectively. The state has 44 local government areas and divided into three different zones, named as Kano Central, Kano South, and Kano North. The state has the highest population in Nigeria, with a population of more than 9,383,682 and a population density of 470 per/sq km [13]. The Kano metropolitan alone has a population of 2,163,225. The current rate of population growth in the state is 2.9% per annum [14]. The Tamburawa water treatment (TWTP) plant with a capacity of producing 150ML portable water per day to covers communities in Kano city and the surroundings. Kano River is treated in the plant exceeding the minimum standard WQ limits of world health organization using conventional treatment processes [13]. Fig.1 shows the map of the study area.

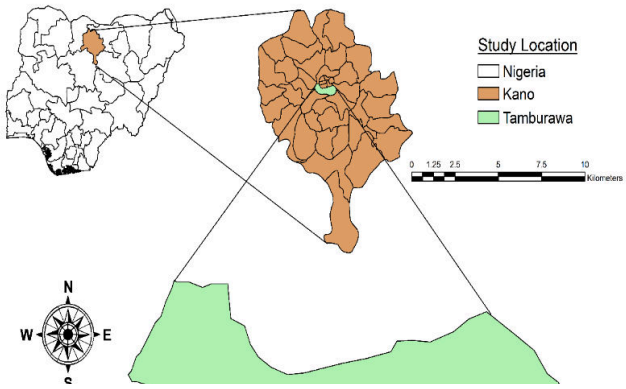


Fig.1, Tamburawa water treatment plant study area

B. Model development and data analysis

A neuro-fuzzy (ANFIS) model was proposed for modeling the turbidity in TWTP using three different model combination. For this purpose, two different optimization approach (hybrid and backpropagation) of fuzzy inference system (FIS) were employed and subsequently, the neuro-fuzzy ensemble was carried out to improve the final prediction accuracy. For input and outputs relationship, the analysis of data is important in any modeling, as such pre-processing and post-processing of the were carried out before the training of the model[15]. The measured daily data for the year 2015 with 196 instances contained raw turbidity (TurbR) (µs/cm), conductivity (CondT) (mS/cm), total dissolve solid (TDSR) (mg/L), chloride (mg/L) and suspended solid (SSR) (mg/L) and Hardness (HardnessT) (mg/L) as the input of the model and turbidity (TurbT) (µs/cm) as corresponding the output. Table 1 shows the descriptive statistical analysis used for this study. After normalization as in Eq. 1, the data were portioned into 75% and 25% for training and testing phase. For the development of the model, Fig. 2, show the flowchart of the used model in this paper.

$$y = 0.5 + \left(0.5 \times \left(\frac{x - \bar{x}}{x_{max} - x_{min}}\right)\right) \quad (1)$$

Where $\bar{X}$, X_{max}, X_{min} and σ indicates the mean, maximum, minimum and standard deviation and respectively.

For better modeling, it is important to select the appropriate input combination, therefore a set of three different models (M1, M2 and M3) were derived on the basis of significant input variables as:

$$\left.\begin{aligned} \boldsymbol{M1} &= Turb_R + SS_R \\ \boldsymbol{M2} &= Turb_R + SS_R + Cl_R + Hadness_R \\ \boldsymbol{M3} &= Turb_R + Cond_R + TDS_R + Hadness_R + Cl_R + SS_R \end{aligned}\right\} \quad (2)$$

Table 1. Descriptive data statistics of the considered data

Parameters	$\bar{x}$	X_{max}	X_{min}	σ	Correlation with $Turb_T$
$Turb_R$ (mg/L)	200.7667	1796.0000	51.0000	220.2917	0.4448
$Cond_R$ (mS/cm)	116.5067	257.0000	53.0000	43.7466	-0.0584
TDS_R (mg/L)	57.4705	106.400	17.9000	20.7953	-0.1074
$Hardness_R$ (mg/L)	36.3028	53.8700	24.6100	6.33222	0.1697
Cl_R (mg/L)	13.4560	33.5600	8.8800	4.1255	0.1959
SS_R (mg/L)	154.200	1248.0000	34.0000	161.4377	0.4170
$Turb_T$ (mg/L)	0.8181	3.4100	0.2000	0.5390	1.000

Where y is the normalized data. x is the measured data, $\bar{x}$ is the mean of the measured data, x_{max} is the maximum value of the measured data and xmin is the minimum value.

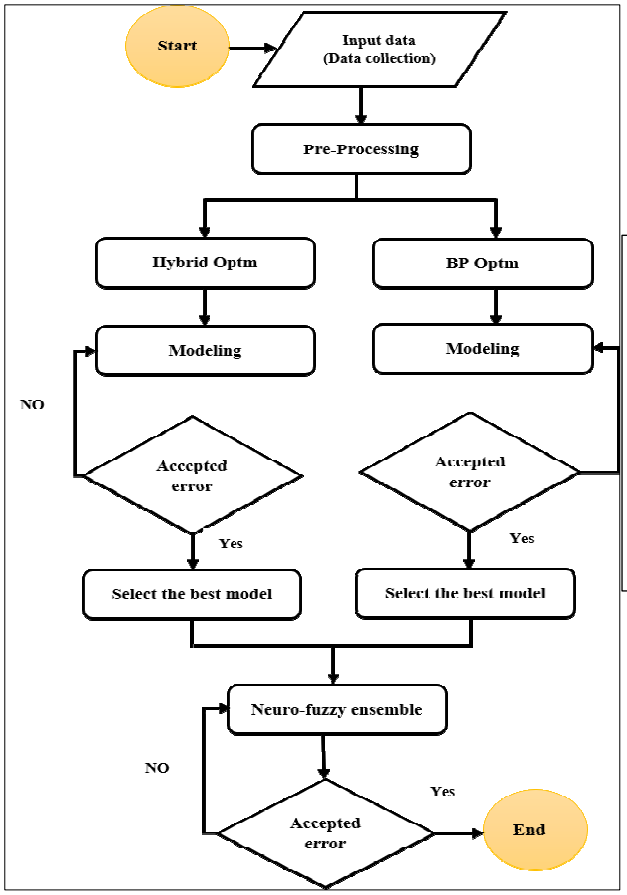


Fig. 2. Flow chart of the proposed models employed in this paper

C. *Adaptive neuro-fuzzy inference system (ANFIS)*

ANFIS is a hybrid integration of learning power of artificial neural network and a fuzzy logic-based technique. ANFIS employed least squares methods and backpropagation gradient descent which are being constructed by hybrid learning algorithm and iteratively tuning and adjusting a fuzzy membership function parameter [16]. The objective behind training an ANFIS system is to govern resultant parameter and optimal premise by training the FIS system model with ANFIS in order to modify the membership function parameter to cope with the training database on error chosen criterion, when ANFIS contains training data and checking data, the least square data model error is selected which is having parameter related to the FIS model[17]. Owing to the hybrid nature of ANFIS serve as an important tool in the modeling process such as environmental prediction [18]. The most utilized membership functions in ANFIS model are triangular (Trimf), trapezoidal (tramf), gaussian (gaussm), bell-shaped (gbellmf) and sigmoidal (sigmf) [19]. For more explanation of PCA refers to the studies of [20],[21].

Assuming 'x' 'y' is the input and 'f'' is the output of a fuzzy inference system the first-order Sugeno type as the following rules (Eqs. 3 and 4).

Rule (1): if $\mu(x)$ is A_1 and $\mu(y)$ is B_1; then $f_1 = p_1x + q_1y + r_1$ (3)

Rule (2): if $\mu(x)$ is A_2 and $\mu(y)$ is B_2; then $f_2 = p_2x + q_2y + r_2$ (4)

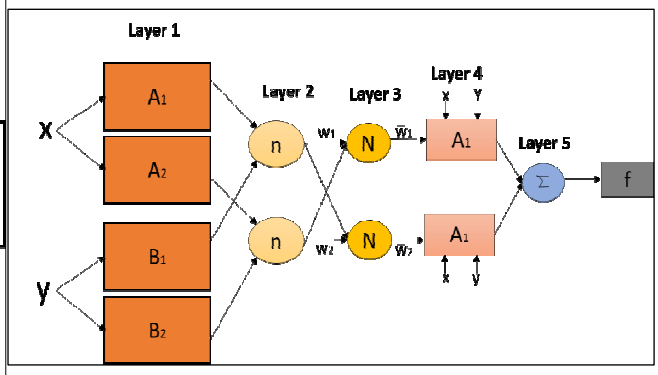


Fig. 3. The general structure of ANFIS [21].

Neuro-fuzzy Ensemble techniques

Ensemble learning is a machine learning which combines the process of multiple predictors in order to enhance the model performance. The branch of machine learning dealing with multiple homogenous or heterogeneous models is collectively termed as ensemble learning [22]. Ensemble techniques had been widely applied in several fields such as prediction, classification, time series, and regression problems. The primary goal for this technique follows the concept of enhancing the performance of the single model, by combining the results of the various individual models[2].

Recently, non-linear ANN ensemble has been employed by some researchers and all of them reported the merit of the technique [9]. In this study, non-linear neuro-fuzzy ensemble techniques were used based on the two different optimization FIS (hybrid and backpropagation) to improve the performance of single models. However, it is notable that, other AI models may also be integrated in an ensemble manner [23].

D. Model performance criteria

For comparative evaluation of the performance accuracy of the prediction modeling, two criteria were used as a multi-criteria approach for assessing the models, in which model performance was evaluated using one statistical error and goodness-of-fit measures, including the root mean- squared error (RMSE) and coefficient of determination (R2) [24].

$$R^2 = 1 - \frac{\sum_{j=1}^{N}[(Y)_{obs,j} - (Y)_{com,j}]^2}{\sum_{j=1}^{N}[(Y)_{obs,j} - \overline{(Y)}_{obs,j}]^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left((Y)_{obs,j} - (Y)_{com,j}\right)^2} \quad (6)$$

Where N, Y_{obs}, Y_{com} and $\overline{Y}$ indicates the number of samples, observed data, computed data and mean of observed, respectively, for all the considered parameters [25], [26].

VIII. Result and Discussion

In this paper, the descriptive statistical analysis and correlation matrix to explore the type, degree, and extent of the relation between the raw and treated water is applied. It is clear from Table 1 that, the raw water values of some parameter are higher than the expected designed value, such as SS_R and TDS_R which are mainly caused by Turbidity. Fig. 4, shows the time series plot of the raw and treated Turbidity, the plot demonstrated the degree and efficiency of TWTP based on the scale range of raw and treated water.

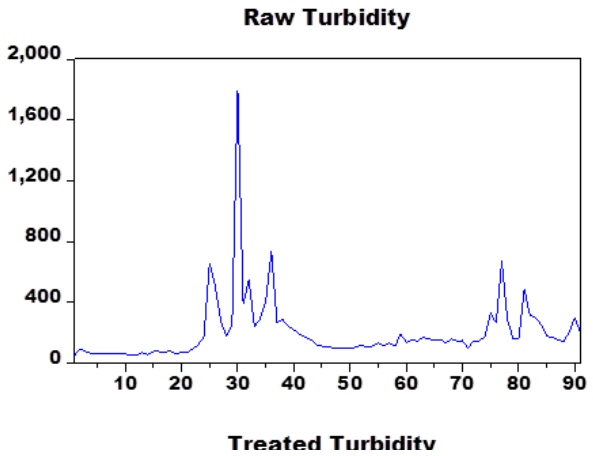


Fig.4 Time series of raw and treated turbidity at TWTP

For both training and testing phase, Sugeno-type FIS with sub-clustering was used due to it is appropriateness for ANFIS modeling. The membership function (MFs) parameters were trained by a different set of model combination (input and output variables) through backpropagation and hybrid optimization algorithms. The error tolerance and epoch number are iterated using Trial-and-error procedure in order to find the best model and ANFIS architecture. Table 2 depicts the obtained results of ANFIS using backpropagation and hybrid optimization algorithms.

Table 2. The result of the optimization algorithms

		Training		Testing	
Optimization	**Model**	**R^2**	**RMSE**	**R^2**	**RMSE**
Hybrid	M1	0.5339	0.1561	0.5319	0.1456
	M2	0.7394	0.1474	0.7317	0.1036
	M3	0.8926	0.0620	0.8896	0.0034
Back-propagation	M1	0.4018	0.1691	0.4016	0.1034
	M2	0.6531	0.1569	0.6368	0.1389
	M3	0.7268	0.1487	0.7374	0.1023

It can be seen from the Table 2 that among the three model combinations, M3 is the best performing model for both hybrid and backpropagation algorithms with the minimum value of RMSE as 0.0034 and 0.1023 in the testing phase for hybrid and backpropagation respectively. despite the better performance of M3, M2 also can be served the predictive performance and proved to be merit, the accuracy of M2 could be attributed to the high correlation of SS_R which is the major factor affecting the turbidity in TWTP. However, the direct comparison for the two algorithms models shows that the hybrid outperformed the back-

propagation algorithms in the prediction of Turbidity, this is due to the combined nature of hybrid algorithms. The closet assessment of the results indicates that hybrid algorithms increased the performance accuracy of backpropagation up 16% and 15% in both the training and testing phase respectively. The superiority of the hybrid algorithm was demonstrated in all the three-model combination (M1, M2, and M3) this can justify by the radar plots in Fig. 5 of R^2 in both training and testing. The radar plot is employed to compare both the training and testing values based on the different scores. However, from the multivariate analysis of the radar chart, the high score indicates better performance which is same in the case of identifying the range of R^2. Similarly, Fig. 6 shows the time series of the observed vs computed models, from the Fig.6 we can vividly observe the agreements accuracy of the fitted values on hybrid algorithms than that of backpropagation in predicting the performance of Turbidity.

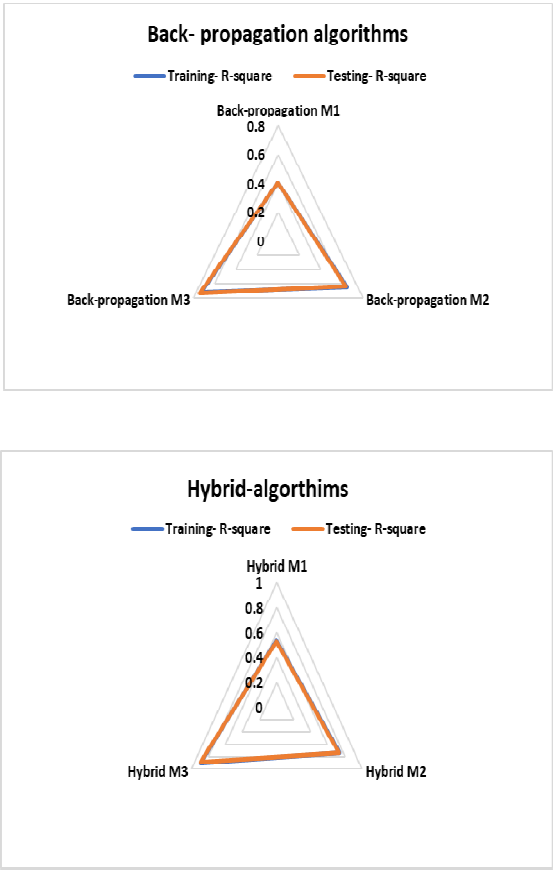


Fig.5 Radar chart for Turbidity in both training and testing phase of two optimization

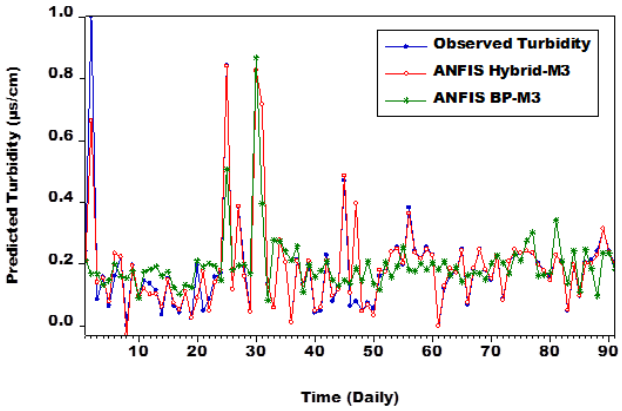


Fig. 6 Time series of the observed vs computed value for the best model

It has been reported by [23], [26] that, there is no single algorithm that provides the best prediction accuracy, as such the difference in performance of the models at different points could be due to distinctions in their mode of operations, their responses to inputs, and their ability to capture different data phenomena. Therefore, combining the results of these models in form of an ensemble framework could increase the generalization capability of the modeling by properly capturing the target and hence to improve the prediction performance. Table 3 shows the comparison results of two neuro-fuzzy ensemble algorithms, the obtained results indicated that the neuro hybrid ensemble model is superior to backpropagation ensemble model. This due to the remarkable merging with regards to R^2 and RMSE. Further investigation of the results depicted that, using for predicting the Turbidity at TWTP using neuro-fuzzy model both hybrid optimization algorithms and ensemble hybrid are satisfactory and can be recommended. In the other hand for backpropagation optimization, ensemble backpropagation emerged to be more satisfactory than using normal backpropagation optimization.

Table 3. Comparison results of two neuro-fuzzy ensemble algorithms

	Training		Testing	
Ensemble Techniques	**R^2**	**RMSE**	**R^2**	**RMSE**
Hybrid ensemble	0.9483	0.0613	0.9879	0.0033
BP-ensemble	0.8589	0.0260	0.8053	0.0797

In the same way, Fig.7 depicts the bar plot of the performance indicator for the two ensemble techniques, it was reported that the smaller the values of RMSE the more accurate the predicting results while in terms of R^2 the reverse is the case. In addition, Table 3 proved that the hybrid ensemble increased the prediction efficiency of backpropagation ensemble up to 18% in the testing phase.

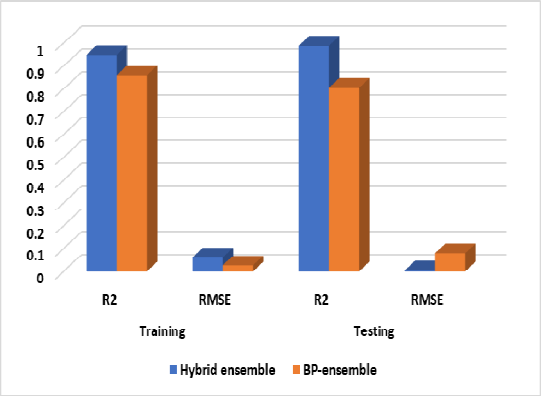


Fig.7 The bar plot of the performance indicator for the two ensemble techniques

IX. CONCLUSION

This research described the implementation of neurofuzzy ensemble using two hybrid and back-propagation training optimization to predict the turbidity in water treatment plant, based on daily measured data obtained from Tamburawa water treatment plant. The obtained results indicated that neurofuzzy ensemble is encouraging with high performance accuracy than the used of single optimization fuzzy inference system. The assessment of the results also indicates that hybrid algorithms increased the performance accuracy of back- propagation up 16% and 15% in both the training and testing phase respectively. For neuro-fuzzy ensemble results, the performance proved that hybrid ensemble increased the prediction efficiency of back- propagation ensemble up to 18% in the testing phase. The result also recommended the implementation of neuro-fuzzy model in an ensemble manner in comparing with other artificial intelligence so as to enhance the prediction of the single model. It is suggested that other AI may be applied with the combination of ensemble techniques so as to come up with a neuro-fuzzy ensemble model in order to produce higher accuracy and more reliable prediction.

References

[1] T. Gómez, G. Gémar, M. Molinos-Senante, R. Sala-Garrido, and R. Caballero, "Assessing the efficiency of wastewater treatment plants: A double-bootstrap approach," *J. Clean. Prod.*, vol. 164, pp. 315–324, 2017.

[2] V. Nourani, G. Elkiran, and S. I. Abba, "Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach," *Water Sci. Technol.*, 2018.

[3] M. M. Hamed, M. G. Khalafallah, and E. A. Hassanien, "Prediction of wastewater treatment plant performance using artificial neural networks," *Environ. Model. Softw.*, vol. 19, no. 10, pp. 919–928, 2004.

[4] B. E. Emrah Dogan, Asude Ates, Ece Ceren Yilmaz, *Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand*, vol. I, no. December. Wiley InterScience, 2015.

[5] M. S. Gaya, N. Abdul Wahab, Y. M. Sam, and S. I. Samsudin, "ANFIS Modelling of Carbon and Nitrogen Removal in Domestic Wastewater Treatment Plant," *J. Teknol.*, vol. 67, no. 5, 2014.

[6] O. Kisi and Z. Mundher, "Catena The potential of hybrid evolutionary fuzzy intelligence model for suspended sediment concentration prediction," *Catena*, vol. 174, no. May 2018, pp. 11–23, 2019.

[7] M. Hameed, S. Shartooh, S. Zaher, and M. Yaseen, "Application of artificial intelligence (AI) techniques in water quality index prediction : a case study in tropical region , Malaysia," *Neural Comput. Appl.*, vol. 28, no. s1, pp. 893–905, 2017.

[8] G. Elkiran, V. Nourani, S. I. Abba, and J. Abdullahi, "Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river," *Glob. J. Environ. Sci. Manag.*, vol. 4, no. 4, pp. 439–450, 2018.

[9] E. Sharghi, V. Nourani, and N. Behfar, "Earthfill dam seepage analysis using ensemble artificial intelligence based modeling," *J. Hydroinformatics*, p. jh2018151, 2018.

[10] S. E. Kim and I. W. Seo, "ScienceDirect Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers," *J. Hydro-Environment Res.*, vol. 9, no. 3, pp. 325–339, 2015.

[11] B. B. Gulyani and A. Fathima, "Introducing Ensemble Methods to Predict the Performance of Waste Water Treatment Plants (WWTP)," vol. 8, no. 7, 2017.

[12] K. T. Lee, J. Ho, H. Kao, G. Lin, and T. Yang, "Using Ensemble Precipitation Forecasts and a Rainfall-Runoff Model for Hourly Reservoir Inflow Forecasting during Typhoon Periods," *J. Hydro-environment Res.*, 2018.

[13] M. S. Gaya *et al.*, "Estimation of turbidity in water treatment plant using hammerstein-wiener and neural network technique," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 5, no. 3, pp. 666–672, 2017.

[14] N. population commission (NPC)., "Gender and Sustainable Development2. Nigeria population census 1991, Analysis.," 2016.

[15] S. I. Abba and G. Elkiran, "Effluent prediction of chemical oxygen demand from the astewater treatment plant using artificial neural

network application," in *Procedia Computer Science*, 2017, vol. 120.
[16] S. I. Abba, S. J. Hadi, and J. Abdullahi, "River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques," in *Procedia Computer Science*, 2017, vol. 120.
[17] A. Solgi, A. Pourhaghi, R. Bahmani, and H. Zarei, "Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting biochemical oxygen demand (BOD)," *Ecohydrol. Hydrobiol.*, vol. 17, no. 2, pp. 164–175, 2017.
[18] J. R. Jang, "ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System," vol. 23, no. 3, 1993.
[19] A. Solgi, H. Zarei, V. Nourani, and R. Bahmani, "A new approach to flow simulation using hybrid models," *Appl. Water Sci.*, vol. 7, no. online, p. 16, 2017.
[20] S. P. Methods, "SS symmetry Performance Comparison of ANFIS Models by Input," 2018.
[21] V. N. G. Elkiran, S. I. Abba, and J. Abdullahi, "Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river," vol. 4, no. 4, 2018.
[22] N. M. Baba, M. Makhtar, S. Abdullah, and M. K. Awang, "CURRENT ISSUES IN ENSEMBLE METHODS AND ITS APPLICATIONS," vol. 81, no. 2, pp. 266–276, 2015.
[23] V. Nourani, G. Elkiran, J. Abdullahi, and A. Tahsin, "Multi-region Modeling of Daily Global Solar Radiation with Artificial Intelligence Ensemble," *Nat. Resour. Res.*, 2019.
[24] M. Hameed, S. S. Sharqi, Z. M. Yaseen, H. A. Afan, A. Hussain, and A. Elshafie, "Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia," *Neural Comput. Appl.*, vol. 28, pp. 893–905, 2017.
[25] D. R. Legates and G. J. McCabe Jr., "Evaluating the Use of 'Goodness of Fit' Measures in Hydrologic and Hydroclimatic Model Validation," *Water Resour. Res.*, vol. 35, no. 1, pp. 233–241, 2005.
[26] G. Elkiran, V. Nourani, and S. I. Abba, "Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach," *J. Hydrol.*, p. 123962, Jul. 2019.

Learning from Imbalanced Data

Danung, Simtong Langchel
Computer Science Department
Americcan University of Nigeria
Yola, Nigeria
simtong.danung@aun.edu.ng

Musa Yusuf
Computer Science Department
Bingham University
Karu, Nigeria
yusufmusa@binghamuni.edu.ng

Ogedebe Peter
Computer Science Department
Baze University
Abuja, Nigeria
peter.ogedebe@bazeuniversity.edu.ng

Abstract*— In the field of Machine learning, the classification of data with imbalanced class distributions has posed a significant drawback in the performance attainable by most classification systems, which assume relatively balanced class distributions. This problem is especially crucial in many real world application domains, such as medical diagnosis, fraud detection, network intrusion, etc., which are of great importance in machine learning and data mining. This research paper explores meta-techniques, which are applicable to most classifier learning algorithms, with the aim to advance the classification of imbalanced data. Multiplayer Perceptron and Naïve Bayes Trees Algorithm has being employed to comparatively study the accuracy and determine the most effective and efficient classifier for both learning, time reduction and accuracy improvement. This study begins by discussing the problems that arise when learning with imbalanced dataset, reviewing various methods and techniques that address the class imbalanced problems, both at the data level and the algorithmic level. It also goes ahead to discuss a number of evaluation metrics that have been developed to assess classifier performance on imbalanced dataset, the two algorithms presented were used for the learning process and a model was built for both algorithms, which was tested over ten datasets. It also covers experimental analysis to determine the most effective model. It ends with a discussion on the experimental results of classification of real-world imbalanced data. Compared with the Naïve Bayes Trees algorithms, the Multi-Layer Perceptron Neural network algorithm is superior in achieving better measurements regarding the learning objectives from imbalanced data.

Keywords—: Imbalance data, Machine Learning, Naive Bayes Trees, Multiplayer Perceptron, Classification

I. INTRODUCTION

An imbalanced dataset is a dataset whose classification categories are not equally represented, that is the classes are not divided into equal class. The class imbalance problem typically occurs in a classification problem, where there are many more instances of some classes than others. In such cases, standard classifiers tend to be overwhelmed by the large classes and ignore the small ones. In practical applications of imbalance data, the ratio of the small to the large classes can be drastic, such as 1 to 100, 1 to 1,000, or 1 to 10,000 (and sometimes even more). These problem is prevalent in many applications such as fraud/intrusion detection datasets, face recognition [1] text classification [2, 3], medical diagnosis/monitoring [4], Activity Recognition [5], Cancer Malignancy grading [6], . It is worth noting that in certain domains as mentioned earlier the class imbalance is intrinsic to the problem. For example, within a given setting, there are typically very few cases of fraud as compared to the large number of honest use of the offered facilities. However, class imbalances sometimes occur in domains that do not have an intrinsic imbalance. This will happen when the data collection process is limited (e.g., due to economic or privacy reasons), thus creating "artificial" imbalances. Conversely, in certain cases, the data abounds and it is for the scientist to decide which examples to select and in what quantity. In addition, there can also be an imbalance in costs of making different errors, which could vary per case.

A number of solutions to the class-imbalance problem were previously proposed; at the data level, algorithmic level and Hybrid level [7, 8]. At the data level, these solutions include many different forms of re-sampling such as random oversampling with replacement, random under sampling, directed oversampling (in which no new examples are created, but the choice of samples to replace is informed rather than random), directed under sampling (where, again, the choice of examples to eliminate is informed), oversampling with informed generation of new samples, and combinations of the above techniques. At the algorithmic level, solutions include adjusting the costs of the various classes so as to counter the class imbalance, adjusting the probabilistic estimate at the tree leaf (when working with decision trees), adjusting the decision threshold, and recognition-based (i.e., learning from one class) rather than discrimination-based (two class) learning. Many of these solutions are discussed in the papers presented in the workshops or are referred to in the active bibliography on the topic.

A recent work of Zhang et., al.[9] Tackled class imbalanced datasets problem, on the basis of the margin theory, they defined the majority class margin and the minority class margin hence forming a balanced set of sample through over-sampling of the class margin of the minority sample, to test for the learning theorem on the sample set, they implemented AdaBoost which was used to form MROBoost algorithm. Here to tackle misclassification samples of the negative minority class margin the initial weight of the samples in the AdaBoost algorithm is given a level of skewness to enable correct classification. They tested based on the performance measures using UCI dataset, comparing AdaBoost, RDSBoost, ROSBoost; and it was seen that the MROBoost performed better than the others did.

Li et., al. [10] applied a cost effective approach known as Extreme Learning Machine to learn and tackle imbalanced dataset problems. They achieve that by their proposed method known as Data Distribution Based Weighted Extreme Learning Machine (D-WELM). Their methods performed well with all the original advantages of the Extreme learning machine of speed and considers both class distribution and sample size.

[11] conducted a comparative analysis to examine the evaluation suitability between Receiver Operating Characteristics Curve (ROC) and Matthew Correlation Coefficient (MCC) based on degree of consistency and degree of discriminancy when used in evaluating imbalanced datasets. They perform experimentation on 54 different imbalanced datasets with imbalanced ratios ranging from 1% to 10%. The study established and recommended AUC to be a better measure than MCC.

In section 1 of this paper, we present the general introduction of the study of learning imbalanced data and its applications. Sections 2 describes the Methodology. The features of the imbalanced data used in the study are discussed in section 3 while section 4 is fully devoted the experimental approach used in this paper. However, Section 5 and 6 deals with the Experimentation and summary and conclusion. Finally, recommendations are discussed in section 6 presents

II. METHODOLOGY

a. *Algorithmic Level Approaches*

In this approach several methods have been proposed with good performance on unbalanced data, at this level, solutions include adjusting the cost of the various classes so to counter the imbalanced class, adjusting the probabilistic estimate at the tree leaf (when working with decision trees), adjusting the decision threshold, and recognition-based (i.e., learning from one class) rather than discrimination-based (two class) learning. These approaches include modified SVMs [12], Decision trees[13], k nearest neighbor KNN [14], neural networks, [15], and rough set based algorithms, probabilistic decision tree and learning methods. But in this study we are going to base on just three of the methods which are discussed below;

b. *Artificial Neural Networks*

Neural Network or Artificial Neural Network is an information processing technique inspired by the way biological brain system works. A neural network contains a number of interconnected processing nodes (or neurons) working in parallel to solve a particular problem.

Neural networks are powerful in deriving meanings from complex or imprecise data, which can be used to understand or recognize things that are too complex to be noticed by other methodologies. A neural network simulates human brains by learning expertise from examples, and stored knowledge in interneuron connection strengths known as synaptic weights. In our experiment, we applied multilayer perceptron (MLP)

S/no	Datasets	#Ex.	#Atts.	Class(+,-)	Validation	I R	Group
1	Abalone9-18	731	8	(positive, negative)	5 folds	16.4	2
2	Vehicle0	846	18	(positive, negative)	5 folds	3.25	1
3	New-thyroid1	215	5	(positive, negative)	5 folds	5.14	1
4	Vowel0	988	13	(positive, negative)	5 folds	9.98	2
5	Wisconsin	683	9	(2,4)	5 folds	1.86	1
6	Newthyroid2	215	5	(positive, negative)	5 folds	5.14	1
7	Ecoli4	336	7	(positive, negative)	5 folds	15.8	2
8	Glass1	214	9	(positive, negative)	5 folds	1.82	1
9	Ecoli1	336	7	(positive, negative)	5 folds	3.36	1
10	Ecoli2	336	7	(positive, negative)	5 folds	5.46	1

PERFOEMANCE METRIC

which is the most commonly used neural network architecture.

c. *Multilayer Perceptron (MLP)*

MLP is a supervised network which requires a labeled training data for learning. Back propagation is used to adjust the weights a small amount at a time in a way that reduces the error. The ultimate goal of the training process is to reach an optimal solution based on our performance measurement.

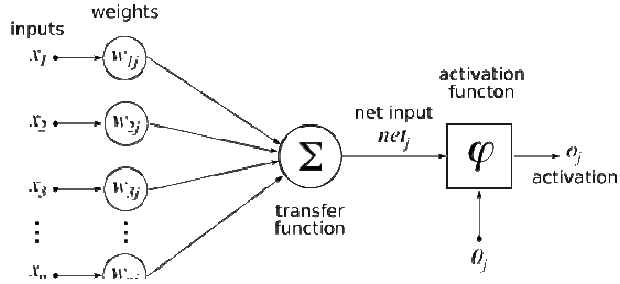


Fig 3: Showing the structure of an Artificial Neuron

The multilayer perceptron (MLP) or Multilayer feedforward network Building on the algorithm of the simple Perceptron, the MLP model not only gives a perceptron structure for representing more than two classes, it also defines a learning rule for this kind of network [16].

The MLP is divided into three layers: the input layer, the hidden layer and the output layer, where each layer in this order gives the input to the next. The extra layers give the structure needed to recognize non-linearly separable classes.

d. *Naïve Bayes Trees*

Naïve Bayes Trees Classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features[17].

The Bayes' theorem is stated mathematically in equation 1 shown below;

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where A and B are events.

P (A) and P (B) are the probabilities of A and B without regard to each other.

P (A | B), a conditional probability, is the probability of A given that B is true.

P (B | A), is the probability of B given that A is true. The aforementioned Naïve Bayes theorem has been one of the most efficient methods in used.[13].

b. DATA SET

Table I shows the details of datasets used in the experiment. The datasets were obtained from Knowledge

Extraction based on Evolutionary Learning (KEEL) [18, 19]. For each dataset, the number of examples (#Ex.), number of attributes (#Atts.) class name for each class (minority(-) and majority(+)), the validation, the Imbalance Ratio (IR) and the ratio group where Group 1 are imbalance ratio between 1.5 and 9 and Group 2 imbalance ratio higher than 9.

III. Experiments and Result

In this section, we apply the proposed learning approach to ten benchmark datasets taken from the keel database repository, the benchmarks used are the Abalone19, Abalone9-18, Vehicle0, New-thyroid1, Page-blocks0, Vowel0, Segment0, Yeast6, Wisconsin, Newthyroid2, Ecoli4, Iris0, Glass1, Ecoli1, and Ecoli2. These datasets are summarized in table I.

Our aim is to study the learning and classification capability of the two algorithms which are the Naïve Bayes Tree algorithm and the Multilayer Perceptron approach from the neural networks and evaluate those models on the imbalanced datasets presented in table IV whish shows that MLP outperformed NBT in terms of Recall and Precision.

4.1 Performance Measure

Classifier performance measures uses evaluation metrics which play an important role in machine learning, they are used to guide and evaluate the learning algorithms. A variety of common metrics as stated above are defined based on the confusion matrix (also called a contingency table). A two by two contingency table is shown in table II and table III. Table II presents the description of each of the performance metric.

Description of Perfoemance Metric

SN	Performance Metric	Interpretation of performance metric
1	$ACC = \frac{TP+TN}{TP+FN+FP+FN}$	Accuracy is one of the most commonly used empirical measure even though it does not distinguish between the numbers of correct labels of different classes, which in the framework of imbalanced problems therefore may lead to erroneous conclusions.
2	$Precision = \frac{TP}{(TP)+(FP)}$	The precision was used to evaluate the classifiers' percentage of positive predictions made by the classifier of which is correct
3	$Recall = \frac{TP}{(TP)+(FN)}$	The recall was used to evaluate the classifiers 'percentage of true positive patterns in the data that are correctly detected by the classifier
4	$F-measure = \frac{2*PrecisionxRecall}{Precision+Recall}$	F-measure defined as the Harmonic mean of recall and precision (Fawcett, 2006). A high level of F-measure signifies a both high value for precision (purity metric) and recall
5	$sensitivity = \frac{TP}{(TP)+(FN)}$ $Sensitivity = Recall$	We used Sensitivity also known as Recall to evaluate the percentage of true positive patterns that are correctly detected by the classifier; hence, it can be referred to as the True positive rate.

.Table II depicts the confusion metric attributes used in the evaluation of the classifier's accuracy, we also used the actual sum given by;

$$ActualClassSum = (TP + FN) + (FP + TN) \ldots\ldots (2)$$

Equation 2 estimates the sum of the positive and negative in respect to the Actual class which is the column wise sum representation of table II

$$PridictionClassSum = (TP + FP) + (FN + TN) \ldots\ldots (3)$$

Equation 3 estimates the sum of the positive and negative in respect to the Prediction class which is derived from row wise sum representation of table II

$$\mathbf{Total} = ActualClassSum = PridictionClassSum \ldots. (4)$$

The total shows the summation of both the column wise sum and the row wise sum that is sum of both the actual class and the prediction class which both should be equal as shown on table II

Confusion Metrix

		Actual Class		
		Positive	**Negative**	Total
Predicted Class	**Positive**	True Positive (TP)	False Positives (FP)	TP + FP
	Negative	False Negatives (FN)	True Negatives (TN)	FN + TN
	Total	TP + FN	FP + TN	**Total**

4.2 Performance Monitoring Metrics

When dealing with class imbalance, Sensitivity, Specificity and Geometric mean metrics have usually been adopted to monitor the classification performance on each class separately. These metrics measures are being used when performance of both classes is concerned and is expected to be simultaneously high. The way in which these measures are adopted in this paper is presented in the table IV.

IV. Summary and Conclusion

In this paper research work, we discussed the problems that arise when learning with imbalanced data sets; we reviewed various methods and techniques that address the class imbalanced problems, both at the data level and the algorithmic level. We also discussed a number of evaluation metrics used to assess classifier performance on imbalanced datasets. We presented two models derived from NBT and MLP whose performance was tested on ten datasets which for the learning process and a model was built for both algorithms which was tested over fifteen datasets. Experimental results show that the Multilayer Perceptron approach can effectively improve the classification accuracy and learning process on imbalanced datasets.

Hence, from the findings, it is recommended that for learning from imbalanced dataset the MLP should be considered since it has exhibited a better classification performance compared to NBT while maintaining higher accuracy, and minimal error rate should be used.

V. SUGGESTION FOR FUTURE RESEARCH

Imbalanced data in machine learning has a several success stories in application to learning process. In spite of this, machine learning communities and researches need to address a number of issues to improve our understanding of the imbalanced classification. Below is a list of problems and research directions that are worth focusing in this regard.

There is a need to carry out detailed comparative study on multiclass imbalanced datasets.

There needs to create a suite of benchmark data (with ground truth) which include both static and dynamic data with qualitative and quantitative attributes, because current benchmarks are limited to static data.

II. REFERENCES

[1] S. Rodda and U. S. R. Erothi, "Class imbalance problem in the network intrusion detection systems," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 2685-2688.

[2] L. Mathews and S. Hari, "Learning from imbalanced data," in *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*, ed: IGI Global, 2019, pp. 403-414.

[3] T. Munkhdalai, O.-E. Namsrai, and K. H. Ryu, "Self-training in significance space of support vectors for imbalanced biomedical event data," *BMC bioinformatics,* vol. 16, p. S6, 2015.

[4] F. Ladayya, S. W. Purnami, and Irhamah, "Fuzzy support vector machine for microarray imbalanced data classification," in *AIP Conference Proceedings*, 2017, p. 030022.

[5] X. Gao, Z. Chen, S. Tang, Y. Zhang, and J. Li, "Adaptive weighted imbalance learning with application to abnormal activity recognition," *Neurocomputing,* vol. 173, pp. 1927-1935, 2016.

[6] B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Applied Soft Computing,* vol. 38, pp. 714-726, 2016.

[7] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence,* vol. 5, pp. 221-232, 2016.

[8] G. G. Sundarkumar and V. Ravi, "A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance," *Engineering Applications of Artificial Intelligence,* vol. 37, pp. 368-377, 2015.

[9] Z. Zhang, Z. Chen, W. Dai, and Y. Cheng, "An Over-sampling Method Based on Margin Theory," in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 2019, pp. 506-510.

[10] M. Li, Q. Sun, and X. Liu, "Data Distribution Based Weighted Extreme Learning Machine," in *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, 2019, pp. 1-6.

[11] C. Halimu, A. Kasem, and S. Newaz, "Empirical Comparison of Area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for Evaluating Machine Learning Algorithms on Imbalanced Datasets for Binary Classification," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, 2019, pp. 1-6.

[12] C. Zhang, Y. Zhou, J. Guo, G. Wang, and X. Wang, "Research on classification method of high-dimensional class-imbalanced datasets based on SVM," *International Journal of Machine Learning and Cybernetics,* vol. 10, pp. 1765-1778, 2019.

[13] H. Yang, S. Fong, R. Wong, and G. Sun, "Optimizing classification decision trees by using weighted naïve bayes predictors to reduce the imbalanced class problem in wireless sensor network," *International Journal of Distributed Sensor Networks,* vol. 9, p. 460641, 2013.

[14] G. Haixiang, L.Yijing, L. Yanan, L. Xiao, and L. Jinling, "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification," *Engineering Applications of Artificial Intelligence,* vol. 49, pp. 176-193, 2016.

[15] U. Bhowan, M. Zhang, and M. Johnston, "Genetic programming for classification with unbalanced data," in *European Conference on Genetic Programming*, 2010, pp. 1-13.

[16] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE transactions on neural networks and learning systems,* vol. 24, pp. 888-899, 2013.

[17] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia,* vol. 18, p. 60, 2006.

[18] J. Alcalá-Fdez, L. Sanchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell*, et al.*, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing,* vol. 13, pp. 307-318, 2009.

[1] S. Rodda and U. S. R. Erothi, "Class imbalance problem in the network intrusion detection

systems," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 2685-2688.

[2] L. Mathews and S. Hari, "Learning from imbalanced data," in *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics*, ed: IGI Global, 2019, pp. 403-414.

[3] T. Munkhdalai, O.-E. Namsrai, and K. H. Ryu, "Self-training in significance space of support vectors for imbalanced biomedical event data," *BMC bioinformatics,* vol. 16, p. S6, 2015.

[4] F. Ladayya, S. W. Purnami, and Irhamah, "Fuzzy support vector machine for microarray imbalanced data classification," in *AIP Conference Proceedings*, 2017, p. 030022.

[5] X. Gao, Z. Chen, S. Tang, Y. Zhang, and J. Li, "Adaptive weighted imbalance learning with application to abnormal activity recognition," *Neurocomputing,* vol. 173, pp. 1927-1935, 2016.

[6] B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Applied Soft Computing,* vol. 38, pp. 714-726, 2016.

[7] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence,* vol. 5, pp. 221-232, 2016.

[8] G. G. Sundarkumar and V. Ravi, "A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance," *Engineering Applications of Artificial Intelligence,* vol. 37, pp. 368-377, 2015.

[9] Z. Zhang, Z. Chen, W. Dai, and Y. Cheng, "An Over-sampling Method Based on Margin Theory," in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 2019, pp. 506-510.

[10] M. Li, Q. Sun, and X. Liu, "Data Distribution Based Weighted Extreme Learning Machine," in *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, 2019, pp. 1-6.

[11] C. Halimu, A. Kasem, and S. Newaz, "Empirical Comparison of Area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for Evaluating Machine Learning Algorithms on Imbalanced Datasets for Binary Classification," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, 2019, pp. 1-6.

[12] C. Zhang, Y. Zhou, J. Guo, G. Wang, and X. Wang, "Research on classification method of high-dimensional class-imbalanced datasets based on SVM," *International Journal of Machine Learning and Cybernetics,* vol. 10, pp. 1765-1778, 2019.

[13] H. Yang, S. Fong, R. Wong, and G. Sun, "Optimizing classification decision trees by using weighted naïve bayes predictors to reduce the imbalanced class problem in wireless sensor network," *International Journal of Distributed Sensor Networks,* vol. 9, p. 460641, 2013.

[14] G. Haixiang, L.Yijing, L. Yanan, L. Xiao, and L. Jinling, "BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification," *Engineering Applications of Artificial Intelligence,* vol. 49, pp. 176-193, 2016.

[15] U. Bhowan, M. Zhang, and M. Johnston, "Genetic programming for classification with unbalanced data," in *European Conference on Genetic Programming*, 2010, pp. 1-13.

[16] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE transactions on neural networks and learning systems,* vol. 24, pp. 888-899, 2013.

[17] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia,* vol. 18, p. 60, 2006.

[18] J. Alcalá-Fdez, L. Sanchez, S. Garcia, M. J. del Jesus, S. Ventura, J. M. Garrell*, et al.*, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing,* vol. 13, pp. 307-318, 2009.

[19] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez*, et al.*, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic & Soft Computing,* vol. 17, 2011.

[19] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez*, et al.*, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic & Soft Computing,* vol. 17, 2011.

TABLE IV PERFOEMANCE METRIC

		Metric Measures															
		Correctly Classified Instances (%) value		InCorrectly Classified Instances (%) value		TPRate		TN Rate		Precision		Recall		F-Measure		ROC Area	
Algorithm		NBT	MLP	NBT	MLP	NBT	MLP	NBT	MLP	NBT	MLP	NBT	MLP	NBT	MLP	NBT	MLP
No	**Dataset**																
1	Abalone9-18	94.66	**94.80**	5.33	5.19	0.94	0.948	0.674	0.54	0.93	**0.94**	0.95	**0.95**	0.94	0.95	0.72	0.90
2	Vehicle0	93.26	**97.16**	6.73	2.83	0.93	0.972	0.12	0.061	0.93	**0.97**	0.93	**0.97**	0.93	0.97	0.96	0.99
3	New-thyroid1	96.74	**98.13**	3.25	1.86	0.96	0.981	0.12	0.05	0.96	**0.98**	0.97	**0.98**	0.97	0.98	0.99	0.99
4	Vowel0	97.87	**99.69**	2.12	0.30	0.97	0.997	0.07	0.02	0.98	**0.99**	0.97	**0.99**	0.98	0.99	0.97	0.99
5	Wisconsin	**96.85**	94.84	3.14	5.15	0.96	0.948	0.03	0.055	**0.96**	0.94	**0.96**	0.95	0.97	0.95	0.82	0.98
6	Newthyroid2	96.74	**98.13**	3.25	1.86	0.96	0.981	0.12	0.05	0.96	**0.98**	0.96	**0.98**	0.97	0.98	0.99	0.99
7	Ecoli4	96.72	**98.51**	3.27	1.48	0.96	0.98	0.33	0.14	0.96	**0.98**	0.96	**0.98**	0.96	0.98	0.93	0.97
8	Glass1	**77.10**	73.36	22.89	26.63	0.77	0.73	0.29	0.32	**0.76**	0.73	**0.77**	0.74	0.76	0.73	0.80	0.74
9	Ecoli1	89.58	**91.96**	10.41	8.03	0.89	0.92	0.27	0.19	0.89	**0.91**	0.89	**0.92**	0.89	0.92	0.93	0.96
10	Ecoli2	93.15	**95.83**	6.84	4.16	0.93	0.95	0.21	0.12	0.93	**0.95**	0.93	**0.95**	0.93	0.96	0.89	0.95

Fraud Detection in Credit Card and Application of VAT Clustering Algorithm: A Review

A. A. Abdulrazaq, M. B. Abdulrazaq, I. J. Umoh , E. A. Adedokun

***Abstract*—The evolution of secured and reliable internet facilities has greatly influenced the advancement in the global e-commerce, making online transaction more efficient and even more promise for the future decades. The uses of credit card become more popular for online purchases due to the ease and availability of internets. Financial fraud pattern also changes and increases rapidly with the development of modern technology which conversely increases the level of fraudulent transactions in credit card resulting in huge losses. However, several credit card fraud detection techniques have been developed to address the problem. In this paper, different fraud detection method have been reviewed and classified based on the approaches used. Also, the major limitation and reasons why most methods are not very efficient are also presented. Furthermore, this paper proposed the application of VAT algorithm as a clustering algorithm to improve on the limitations of credit card detection methods.**

***Index Terms*— Credit Card, e-commerce, Fraud Detection Techniques, VAT Clustering Algorithm**

1. INTRODUCTION

The new age information technology and advancement in communication systems had enormously aids the rapid development and popularization of mobile devices such as smart phones, palm top and other portable communication gadgets. Online shopping and banking become a popular medium of daily purchases and business transactions [1]. Since credit card enables card holders to borrow funds for payment of goods and services, the numbers of transactions by credit cards are now increasing geometrically [2].

However cases of e-commerce criminal dealings and fraudulent activities in credit card is also increasing and spreading all over the globe resulting to massive financial losses [3]. According to [2], more than $31 trillion were generated worldwide by online payment systems in 2015 which is 7.3% increases compared to 2014. Worldwide losses from credit card fraud increased to $21 billion in 2015, and trend indicate that by 2020, it will possibly hit $31billion. Although fraud can be perpetrated through several means, including mails, chats, phones and internet but online media especially the internet is the most popular one. Because of the global availability of the web and possibility of users easily hiding their identities over internet transactions, there is a rapid growth of unauthorized actions and misuse of credit card over this medium [4]. In addition, since the bandwidth of internetworking channels have been improved, fraudsters tend to form fraud networks for information exchange among themselves and also collaboration all over the world. Hence, online frauds committed over the internet become the most popular ones due to their nature [5].

Due to the continuous huge losses as a result of fraudulent acts and misuse of credit card, the development of preventive measures becomes inevitable. Credit card fraud detection is an important means to prevent fraud activities. Literatures have revealed several approaches that have been used for credit card fraud detection [2], [6], [7], each methods focuses on increasing the detection rate while minimizing the false alarm rate. In this paper, different method of credit card detection have been reviewed including decision tree, support vector machine (SVM), Hidden Markov Model (HMM), genetic algorithm (GA), random forest classifier, artificial neural network (ANN), artificial immune system (AIS), Bayesian network (BN), Fuzzy logic system, inductive logic programming (ILP), case-base reasoning, peer grouping (clustering method) and probability density estimation (PDS).

2. CREDIT CARD FRAUD

Fraud can be defined as criminal or wrongful deception with the intention of obtaining goods and services without payment or accessing unauthorized funds from another individual account without leading to any form of direct legal implications [5], [8]. In order to curtail fraud and losses due unauthorized use of credit card and other fraudulent activities, two main mechanisms are involving; fraud preventive measure and fraud detective system.

In prevention methods, precaution measures are taken to avoid the fraud events. But when there is failure in fraud prevention, then the problem of detection is considered. Fraud detection involves identifying suspect, illegitimate and wrongful behavior of credit card profile [9]. The main aim of the designing a system for fraud detection is to examine all transaction for the possibility of either fraudulent or unauthorized regardless of the prevention measures and also to determine fraudulent transaction immediately the fraudster begins to perpetrate a fraudulent transaction [5], [12].

The credit card fraud can also be online or offline fraud depending whether the fraudster gain access to the credit card or its data details. The offline fraud involves the stealing of the credit card. The credit card is physically stolen by an unauthorized person and then used it to order for goods and other transaction. In online fraud, it only requires the information about the card and it is sufficient to make purchase and transactions through mobile phones or internet [8].

III CONCEPT OF CREDIT CARD FRAUD DETECTION

As earlier mentioned, the goal of any credit card fraud detection system is to prevent fraud activities or

transactions. Generally, the detection is done by examining every new transaction pattern and comparing with the known or existing card holder behavioral pattern for classification as either legitimate or fraudulent. If pattern deviate from the card profile, it is declared suspected or fraudulent transaction. Fraud detection techniques are usually categorized into two main classes which include the anomaly detection and classifier-based detection.

3. Anomaly Detection Techniques

Anomaly detection technique also known as the unsupervised fraud detection technique employs the uses of credit card transaction history. In unsupervised methods, no prior data sets which indicate the transaction state to be fraud or legitimate. It simply calculates the distance between the data points in space to ascertain the transaction consistencies. By determining the distance between the incoming transaction and the cardholder's profile, it can effectively identify any incoming transaction with deviant with the cardholder's profile [10].

4. Classifier-Based Detection Techniques

Classifier-based detection is also known as the supervised fraud detection method. This involves training of data sets where previous known legitimate and fraud cases are used to develop a model which will then produce a suspicion score for the new transactions. The classifier is trained using supervised learning methods base on certain predefined features which is been extracted from the card holder behavioral pattern. The supervised learning is to extract fraud features from fraud transactions in order to identify every normal transaction from the fraud ones [2].

IV. TECHNIQUES OF FRAUD DETECTION IN CREDIT CARD

a. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a class of supervised learning model. It is a binary classifier which is associated with learning algorithms capable of analyzing and recognizing patterns for classification tasks. The basic concept of SVM is toward finding an optimal hyper-plane that can separate instances of two certain classes in a linear order. This hyper-plane is usually assumed to be located in the gap between some marginal instances called support vectors. In the case of data that are linearly inseparable, SVM is incorporated with kernel functions for efficient classification as an extension to linearly inseparable data [11]. A kernel function like the radial basis function (RBF) is used to show the projection of the data points in a space with high dimension using dot product of the two points. It is a transformation that disperses dataset through mapping a new space from the input space that the instances are more possibly to be linearly separable. Hence, SVM can easily learn input spaces with complex relationship. For any classification tasks, considering a labeled set of training instances that is associated with a class, a new space (incoming) instance can be determined using SVM training algorithm to find a hyper-plane for the new incoming instances into one among the two classes. The feature space which corresponds to hyper-plane side is used for the prediction of each new data point class. [11]. In credit card fraud detection, SVM serves as an efficient classifier in distinguishing fraudulent transaction from those of the legitimate. SVM in [5], [12], [13] have been successfully applied for fraud detection in credit card.

b. Neural Network

Artificial neural network (ANN) is a network consisting of set of nodes that are interconnected to mimic the functioning operation of the human brain. Every node is associated with a weight connected to many other nodes in adjacent layers. Every individual nodes receives an input from other connected nodes and employ the weights mutually with a simple function to generate an output values. ANNs have different configuration which includes supervised, unsupervised and hybrid methods [14]. In supervised learning approach, among the most common configuration is the back propagation (BP) network which usually minimizes its objective function using multi stage optimization method in a dynamic way for a generalized delta rule. BP methods are mostly used in fraud analysis method. Considering fraud detection model for credit card, samples of both legitimate and non-legitimate card holder behavioral pattern that are correlated with the features that are been extracted are used to create models for the classifier.

The BPN algorithm take longer time and requires intensive training and turning of its parameters including the learning rate and amount of hidden neurons in order to achieve the best result. In the field of fraud detection, back propagation approach which is a supervised neural networks are well known effective tool that have applied in numerous studies [15], [16], [17].

c. Hidden Markov Model (HMM)

A Hidden Markov Model is an unsupervised fraud detection method which is stochastic in nature. It is also double embedded in nature and can be applied to systems that are more complex in comparison with the popular traditional Markov system. In fraud detection for credit card, the HMM is usually trained to model the normal behavior pattern card holder transaction profiles. Considering the HMM model, any incoming transaction can be identified as been fraudulent with low probability of acceptance by model. The profile of every user transaction contains some certain information about last 10 transactions which may include time, location, category and amount of for each transaction [11]. Incorporation of HMM into online fraud detection models is also possible in receiving and verifying transaction details to determine whether it is fraudulent or not. HMM produces high false positive rate and are also capable of processing transactions of larger size. It has been successfully applied in literatures including [18], [19], [20].

d. *Random Forest Algorithm (Bagging Ensemble Classifier)*

Random Forest detection techniques also known as bagging classifier supervised method which model and construction by the combination of several independent-based classifier. This technique has an exceptional data classification with a very less overfitting problem. By applying the multiple trees concept based on the majority vote for the individual predictions, the variance and error rate of the classifier will be reduced and it result is much more better when compared with single-base classifier [2], [21].

Each classifier represents the profile of single card holder's behavioral pattern and the classifier set is usually assigned to individual cardholder in group. Thus, from a similar group, each element (member) has many specific profiles. Hence, considering group's profiles in place of individual profiles, the behavioral patterns of the cardholder's is thereby enriched, better prediction of any deviant behavior from the group profile can be effective identified. In [2], the random forest algorithm has been successful applied in their work with remarkable results.

e. *Artificial Immune system (AIS)*

An artificial immune system (AIS) is an adaptive – based system which is inspired through the operation of the biological immune system. AIS models can handle difficult problems including data clustering, intrusion detection, and classification and search problems. The natural AIS models contain complex networks of specialized organs, tissues, cells and chemical molecules. It has the capability of recognizing, destroying and remembering unlimited large amount of pathogens to include viruses, multi-cellular parasites, bacteria, as well as fungi. The ability of AIS model to distinguish between self and non-self give a protection to the organism. This distinctive capability makes the AIS efficient in handling complex problems involving data clustering and classification [22].

For problems related to fraud detection in credit card, the AIS system is only trained using the legitimate behavioral pattern of the cardholder. However, it is capable of identify any deviation from the card holder profile and such considered been suspicious. It can also be easily incorporated as a data aggregation and clustering algorithm with other system for better fraud detection [9].

f. *Bayesian Network.*

Bayesian network model uses a graphical representation of conditional dependencies among random variables. These networks are important when determining the probability of unknown given that of the known probabilities which is uncertain. Bayesian networks system can also be modeled to known uncertain incoming data in a situation where some certain data is already known. The main objective of using Bayes rules is usually to predict the class label that is associated to a given vector of attributes (features) [11], [23].

For the purpose of detecting fraud, two Bayesian networks can be constructed to represent the behavior pattern of the card holder user. One network is designed to model behavioral pattern under the assumption that the user is illegitimate (I) and the other network is model under the condition the user is a legitimate (L), data from legitimate user is used to develop the 'user net'. Considering the incoming emerging date, legitimate net will be adapted for a specific user. New transactions can then be classified by inserting to the two networks and finally, indicate the type of behavioral pattern (legitimate/illegitimate) based on the corresponding probabilities. Employing Bayes rule, it predict the result of every new transaction to be fraud or not [7]. In [24], Bayesian network was successfully applied for fraud detection in credit card.

g. *Fuzzy Logic based system*

Fuzzy logic systems are based on the rules of fuzzy. It is concern with data that are uncertain, both the input and output variables are pre-defined fuzzy sets which try to express number values in terms of linguistic variables (e.g. big, small, short and large). Models that are based on fuzzy logic includes the fuzzy neural model and fuzzy Darwinian model. In Fuzzy Neural model, it involves learning from certain data which information is not certain but are common and applicable in real world situation. Fuzzy neural model can be employed in detection of fraud in individual credit cards by accelerating the rule of induction. On the other hand, fuzzy darwinian system evolves of the Evolutionary-Fuzzy model which evolves fuzzy rules through genetic programming. While applying the rules, the model can identify and classify fraudulent transaction from the normal ones. This system is a combination of genetic programming (GP) and fuzzy expert model, this make the model to be very fast but highly expensive. Literature reveals that fuzzy logic based system have been successfully applied in fraud detection in credit card including [25], [26],[27].

h. *Inductive logic programming*

This is a supervised fraud detection method which employs a set of samples (positive and negative) used in describing a concept for predicting first order logic used in classifying new instances. While using the ILP in data classification, expression for components with complex relationship, achieving attributes classification becomes very easily [11]. For fraud detection in credit card, ILP can be applied as a classifier in identifying fraudulent transactions. [28] have successfully incorporates ILP with simple homophily-based classifier on a relational database to model a labeled data for fraud detection in credit card model.

i. *Case-based Reasoning (CBR)*

The basic concept of CBR approach involves solving past problems by adapting certain solutions in order to solve new problems. In CBR, a database is developed where cases of past experiences are described and stored to be retrieved for future encounter of similar cases. These cases can be applied for identification and classification solutions. A CBR model usually search for a matching case from the stored cases when encountered with a new problem. In this approach, when new case are presented to the model, it consider both the training

and test data to determine a subset of cases that are most similar in order to predict a result for the newly presented or encountered case. For fraud detection in credit card, CBR can be used as a classifier in checking new pattern of credit holders, comparing and matching case to identify similar pattern. CBR can also incorporate with other algorithm as in [29] where nearest neighbor matching is applied with CBR for fraud detection in credit card. Other hybrid application of CBR can be found in [30].

j. Peer grouping (clustering method)

Peer grouping fraud detection techniques is a class of clustering method. It is an unsupervised approach that examine the behavioral pattern of an account or case with respect to monitoring of the time. The peer group method is capable of identifying certain patterns that have distinct properties that are completely different from other pattern at one moment in time whereas which have similar properties in the previous. Those patterns are usually flagged as suspicious to alert a fraud analyst to investigate such cases. The hypothesis of the peer group analysis is that every behavior of all accounts are been monitored to detect any significant deviate from previous behavior of any account, such account will be notified and then investigated for genuiness [31].

This approach is been used to detect fraud in credit card by examining the card holders transaction profiles. Any transactions with behavior different from their peer group are referred to as fraudulent transactions [3]. Figure 1 show the taxonomy of credit card fraud detection method.

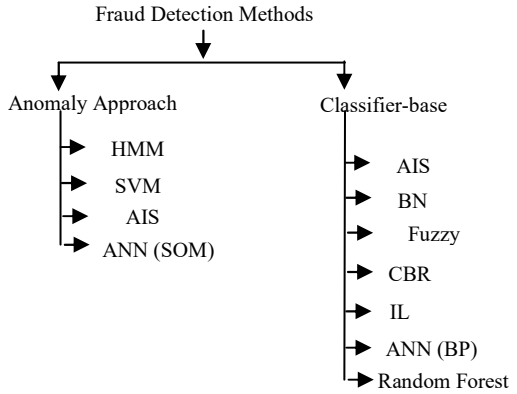


Figure 1: Taxonomy of credit card fraud detection methods

Table 1 present the comparison of various credit card fraud detection method considering the pros and cons of different methods and table 2 present the comparison analysis table for implementation of past work

Table 1.0: Pros and cons of various fraud detection methods

Method	Pros	Cons
SVM (Support Vector Machine)	Ability to minimized effectively classification error	Problem of data overfitting during training
Neural network	It have high detection rate a d effective in handling data with complex relationship.	It requires excessive training time especially with large data set
HMM (Hidden Markov Model)	It has a very fast detection rate	Difficulty in handling large data set.
Random forest	It is very fast in detection and can be easily implemented	
AIS (Artificial Immune System)	Can be easily integrated with other models and very effective in handling noisy data	It requires excessive training of data
Bayesian Network	It has a very good curacy and fast detection rate	It require training of data excessively
Decision tree	It is highly flexible and easy to implement	It requires continuous checking of model condition.
ILP (Inductive Logic Programming)	Effective in handling data of different type including data with complex relationship	Low accuracy especially while handling noisy data.
Case-base Reasoning	It has the ability to handle large dataset and also flexible to update	Problem of low accuracy when handling noisy data
Peer Grouping	Capable of handling large dataset and flexible in updating	It has the problem of slow detection rate
Fuzzy logic	It has a very good accuracy and fast detection rate	It is expensive to implement

-

Table 2.0: Comparison analysis table for implementation of past work

Source	Methods	Detection rate	Computational complexity	Data structure	Metrics for evaluation
[5]	SVM	Low	Higher	Real data (National bank's credit card data warehouse	False alarm rate using 7 (seven) different classifier
[2]	Decision Tree Bagging Ensemble Decision Tree	High Faster Fast	Low Low Lower	Real data (Data set from UCSD-FICO Competition	Fraud catching rate (FCR), false alarm rate (FAR), Balanced classification rate (BCR), Mathews correlation coefficient (MCC)
[31]	Neural network	Fast	high	Real data from W3C's standard, Geo-location API Erukaapi (GEOLOCATOR) Service	Accuracy rate %
[9]	AIS Modified AIS	Fast Faster		Real data set (Transaction records registered by a POS)	Detection rate %, FP rate % and FP ratio
[24]	Bayesian with two (2) clusters Bayesian with two (3) clusters	Faster Fast	High Higher	Synthetic data generated using Markov Modulated Poisson Process (MMPP), Genuine Gaussian Distribution Module (GGDM), and Fraud Gaussian Distribution Module (FGDM).	True positive rate (FPR) and False positive rate (FPR).

V. LIMITATION AND CHALLENGES OF CREDIT CARD DETECTION METHODS

The greatest challenge faced while detecting fraud in credit card model is the nature of data. Understanding and analyzing the data set for transaction in credit card is extremely difficult and has prune serious limitation to fraud detection system. Literatures [2, 3, 7, 10,11, 19, 32] have summaries the issues resulting from datasets problems to include the following:-

- *Misclassification of transaction due to large dataset*: Credit card transaction involve large amount of datasets which need to be processed within a short time in order to gain acceptance as legitimate or rejection as fraudulent. E.g there exist millions of possible location and e-commerce sites where credit card can be used, matching become extremely difficult.

- *Unbalanced Dataset*: Every fraudulent transaction posses features similar to legitimate ones and it is usually about 1% of all transaction. So, predicting and detecting about 1% deviant from a card profile pattern become very difficult.

- *Unavailability of Dataset*: Due to the sensitivity of their records in any financial institution, revealing of customer transaction profile to researcher become very difficult or in any cases not possible. Most fraud detection systems are design and evaluated using synthetic data.

- *Concept Drift*: Over time, there are possibilities of changes in the card holder behavioral pattern (profile) or even the fraudsters could conceive new fraudulent strategies. Credit card detection system could thereby classify legitimates (normal) transaction as non-legitimate (fraudulent).

However, several works have proposed the use of different techniques including data mining and machine learning to tackle fraud issues. Most studies focuses on the formulation and modeling of system capable of identifying and predicting those transaction with a high possibility of being illegal considering the historical profile and fraud patterns. These systems are design through learning / training process based on the normal transaction behavior. In most cases, the systems are limited to certain error, not because of bad modeling or inadequate learning but rather inadequate data preprocessing during clustering stage.

To the best of our knowledge, literatures have reported little on the data aggregation prior to feature extraction for fraud detection system modeling. At the preprocessing stage of dataset clustering, K-means clustering approach is usually employed in most studies. The k-means clustering method is popular and has been proven in many studies to be a very effective method of classification of dataset. The only drawback is that it requires that the number of cluster be specify prior to clustering. In credit card fraud detection system, users need to specify the cluster number base on the personal evaluation of dataset and a fixed number is assign for a model. However, different credit card holder will definitely have different spending or behavioral pattern, so the design of a model with fixed number of cluster may not

represent and describe the dataset effectively. In the study, an automatic cluster number selection based of data structure (transaction profile) is suggested. The application of VAT algorithm for the selection of number of cluster is thereby proposed.

VI. CONCEPT OF VAT ALGORITHM

This method has been referred to as Visual Assessment of Tendency (VAT) for data clustering because it involves the visual analysis of dataset [33]. The tool; VAT algorithm is a visual representation of the number of clusters in the large size datasets. The datasets are usually represented by a dissimilarity matrix which is then converted into gray scale VAT image. Every pixel in the grey scale represents the scaled dissimilarities value. The white pixels which indicate high dissimilarities and the black pixels which indicate low dissimilarities are used to identify the number of clusters. Since every object in each cluster is closely similar and largely different from object of other clusters, hence, the diagonal element of the dissimilarity matrix is always zero (0) [34].

In VAT algorithm, data can be represented in two forms, either as vector (column) or matrix (symmetric). When each object in cluster is represented by a vector, it is referred to as object data representation and VAT tool is widely applicable since the reordered form of dissimilarity data can be easily displayed and can easily be obtained from the original data. Moreover, if the original dataset consists of a matrix of pair wise (symmetric) similarities then, the elements in the matrix will be reordered to obtain a reordered dissimilarity matrix that accumulate every smaller dissimilarity values closer to the diagonal of the matrix in square contiguous regions. Plotting the image of the reordered dissimilarity matrix with its elements as pixel intensities, a darker square blocks will then appear along the diagonal line of the image which are closely related to particular cluster in the min dataset. With this, the number of clusters can be visually observed looking at the underlying data without actually clustering it. However, with VAT algorithm, this visual observation and interpretation is not needed since VAT tool will pre-compute the ordered dissimilarity matrix without any specific relation with actual partitioning [34, 35].

One major merit of VAT algorithm is it application to virtually all numerical datasets as shown in literatures [34,36,37,38]. It has been successful in determining potential cluster structure of various data sets. However, it can be computationally expensive for large data sets especially when considering data in financial domain like the credit card. To address this problem, variants of VAT algorithm have been developed for feasibility of large dataset. Among the notable extension and modifications of VAT approach includes; scalable VAT algorithm [36], improved VAT (iVAT) algorithm [37] and automated VAT (aVAT) algorithm [38]. With the success recorded in other area of application, VAT tool is expected to proffer a better solution in the automatic selection of number of cluster in aggregation of credit card behavioral pattern.

VII. CONCLUSION AND FUTURE WORK

This paper presents the concept of fraud detection in credit card and the different approaches towards the detection of illegitimate transaction have been reviewed. The limitation and challenges to these techniques lies in the inadequate analysis and understanding of the dataset due to large data size and it unavailability. This paper proposes to address this problem by introducing a credit card VAT based clustering approach for better description of the dataset during aggregation into cluster. The focus of this method is to automatically determine the number of cluster that will closely describe and represent each element in the dataset. In this approach, the numbers of clusters is not predefined but are determined based on the data structures which conversely result to more adequate data clustering.

Further work will focus on modeling a system that will be efficient in the description of a very high correlative attributes of the extracted features. Similarly, a feedback mechanism will also be considered for the adaptiveness of the system to address concept drift.

REFERENCES

[1] Sahin, Y. and Duman E. (2010). An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey,

[2] C. Jiang., J. Song., G. Liu., L. Zheng., and W. Luan. (2018). Credit card fraud detection: A novel approach using aggregation and feedback mechanism. *IEEE Internet of Things Journal*. Pp 1-11.

[3] S. Suman and M. Bansal (2014). Survey Paper on Credit Card Fraud Detection. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 3(3), pp827-832.*

[4] M. Krivko (2010). A Hybrid Model for Plastic Card Fraud Detection Systems. *Expert Systems with Applications*, 37(8), pp. 6070-6076, 2010.

[5] Y. Sahin and E. Duman (2011). Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. *Proceedings of the International multi conference of engineers and computer scientists (IMECS 2011) vol.1 p 1-6.*

[6] M. Zareapoor and P. Shamsolmoali (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia Computer Science 48, 679-685.

[7] V.Dheepa and R.Dhanapal (2009). Analysis of Credit Card Fraud Detection Methods. *International Journal of Recent Trends in Engineering*, 2 (3), pp 126-128.

[8] V. Dheepa and R. Dhanapal (2012). Behavior Based Credit Card Fraud Detection using Support Vector Machines. *Journal on soft computing*, 02(4), pp 391- 397.

[9] N. Soltani, M. K. Akbari and M.S. Javan (2012). A New User-Based Model for Credit Card Fraud Detection Based on Artificial Immune System .*16th CSI International Symposium on*

[10] M. Masud, J. Gao and L. Khan (2015). Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints. *IEEE Transactions on Knowledge & Data Engineering*, 23 (6), pp. 859-874.

[11] S. Sorournejad, Z. Zojaji, R. E. Atani and A. H. Monadjemi (2016). A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.
[12] R.C. Chen, S.T. Luo, X. Liang, and V.C.S. Lee (2005). Personalized Approach Based on SVM and ANN for Detecting Credit Card Fraud," *in Proc. IEEE International Conference on Neural Networks and Brain*, pp. 810-815.
[13] Tung-shou Chen and chih-chianglin (2006). A new binary support vector system for increasing detection rate of credit card fraud.
International Journal of Pattern Recognition and Artificial Intelligence. 20 (2), Pp. 227–239.
[14] A. Krenker, M. Volk, U. Sedlar, J. Bester and A. Kosh, (2009). Bidirectional Artificial Neural Networks for Mobile-Phone Fraud Detection. *Journal of Artificial Neural Networks*, 31(1), pp. 92-98.
[15] R. Patidar and L. Sharma (2011). Credit Card Fraud Detection Using Neural Network. *International Journal of Soft Computing and Engineering (IJSCE)*, 1(1),pp 1-7.
[16] E. Aleskerov, B. Freisleben and B. Rao (1997). CARDWATCH: A Neural Network-Based Database Mining System for Credit Card Fraud Detection. *The International Conference on Computational Intelligence for Financial Engineering*, pp. 220-226.
[17] M. Zareapoor, K.R. Seeja, M.A. Alam, (2012). Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria. *International Journal of Computer Applications*. 52(3), pp 23.
[18] A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar (2008). Credit Card Fraud Detection using Hidden Markov Model. IEEE Transactions on dependable and secure computing. 5, pp37-48).
[19] A. Singh and D. Narayan (2012). A Survey on Hidden Markov Model for Credit Card Fraud Detection. *International Journal of Engineering and Advanced Technology (IJEAT),* 1(3), pp49-52.
[20] B. S. Gandhi , R. L. Naik, S.G. Krishna and K. Lakshminadh (2011). Markova Scheme for Credit Card Fraud Detection. *International Conference on Advanced Computing*, *Communication and Networks*. pp 144-147.
[21] L. Breiman (2001). Ramdon Forest. Machine learning. 45(1). Pp5-32.
[22] A. Brabazon, J. Cahil, P.B. Keenan and D. Walsh (2011)., "Identifying Online Credit Card Fraud using Artificial Immune Systems", IEEE Congress on Evolutionary Computation (CEC). Pp 1-7.
[23] S. Panigrahi, A. Kundu, S. Sural and A.K. Majumdar (2009). Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Elsevier, Information Fusion* 10(1), pp 354–363.
[24] L. S. Santos and S. R. Ocampo (2018). Bayesian Method with Clustering Algorithm for Credit Card Transaction Fraud Detection. *Romanian Statistical Review, pp103-120.*
[25] P. J. Bentley, J. Kim, G. Jung and J. Choi (2000). Fuzzy Darwinian Detection of Credit Card Fraud. *14th Annual Fall Symposium of the Korean Information Processing Society*. Pp 1-8.
[26] M.R. Harati Nik (2012). "FUZZGY: A hybrid model for credit card fraud detection. *Sixth International Symposium on Telecommunications (IST),*pp. 1088 – 1093.
[27] M. Syeda, Y. Zhang and Y. Pan (2002)"Parallel Granular Neural Networks for Fast Credit Card Fraud Detection". In: Proceedings of the IEEE international conference, Vol. 1; pp572–577
[28] Muggleton, S. and DeRaedt L. (1994). Inductive Logic Programming: Theory and Methods. *Journal of Logic Programming*,19(20), pp 629- 680.
[29] R. Wheeler and S. Aitken (2000). Multiple algorithms for fraud detection. *Knowledge- Based Systems*, 13; pp. 93-99.
[30] E.B. Reategui. and J. A. Campbel (1994). Classification System For Credit Card Transactions, Advances in Case-Based Reasoning, *Proceedings of the Second European Workshop EWCBR-9*, pp. 280– 291.
[31] A. Gulati, P. Dubey, M. FuzailC, J. Norman and R. Mangayarkarasi (2017). Credit card fraud detection using neural network and geolocation. *Materials Science and Engineering* pp263-270.
[32] D. kaushik , I. kashyap and S. Sharma (2017). A review: Credit card fraud detection using various machines learning algorithm.
*International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 6, Issue 6,*pp
[33] S. Asadi, C. D.V.S. Rao, O.Obulesu and P.S. K. Reddy (2011). Finding the Number of Clusters in Unlabelled Datasets Using Extended Cluster Count Extraction (ECCE). *International Journal of Computer Science and Information Technologies*, Vol. 2 (4), 2011,1820-1824.
[34] M. K. Pakhiraa (2012). Finding Number of Clusters before Finding Clusters. *Elsevier; Procedia Technology* 4 (2012) 27 – 37.
[35] C. Subbalakshmia, G. R. Krishnab, S. K. Mohan Raoc, P. Venketeswa- Raod (2014). A Method to Find Optimum Number of Clusters Based on Fuzzy Silhouette on Dynamic Data Set. *Elsevier International Conference on Information and Communication Technologies (ICICT 2014).* pp*346 – 353.*
[36] T. B. Iredale, S. M. Erfani and C. Leckie (2017)**.** An Efficient Visual Assessment of Cluster Tendency Tool for Large-scale Time Series Data Sets. IEEE.
[37] T.C. Havens and J.C. Bezdek (2012). An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm. I*EEE Transactions on Knowledge and Data Engineering*. 24(5), 813–822.
[38] L. Wang, T. V. Uyen, J. C. Nguyen, C. Bezdek, A. Leckie, and R. Kotagiri (2010). iVAT and aVAT: Enhanced Visual Analysis for Cluster Tendency Assessment. Pp 1-12.

Investigation of Frequency Dynamics on the Nigerian Power System Using Frequency Disturbance Recorder (DFR)

S. K. Mohammed, U. O. Aliyu, G. A. Bakare, Y. S. Haruna, M. Ahmed and I. T. Balami
Department of Electrical and Electronics Engineering, Faculty of Engineering and Engineering Technology, Abubakar Tafawa Balewa University, Bauchi, Nigeria
mskumo@atbu.edu.ng, gabakare@atbu.edu.ng, uoaliyu@atbu.edu.ng, ysharuna@atbu.edu.ng, amohammed3@atbu.edu.ng, and tbishaku@atbu.edu.ng

***Abstract*—The frequency of an electrical power system indicates its wellbeing and performance. To that effect frequency relays are put in place in power substations for local frequency control. This research investigates the level of frequency control in an electrical power substation. A Frequency Disturbance Recorder (FDR), manufactured by Virginia Tech Institute University for the purpose of dynamic frequency recording was installed at the Bauchi PHCN 330/132kV substation for the recording of raw frequency data for a seven-day period. The data recorded was preprocessed using data processing tools implemented in MATLAB 7.1 environment. The relevant statistical analysis showing the nature of the frequency dynamics were presented. Despite the presence of noise, spikes and outliers, the processed frequency conformed to the Nigerian frequency regulation band of ± 2.5 % of the nominal frequency of 50 Hz indicating the compliance of the local frequency control.**

Keywords—Frequency disturbance recorder, preprocessed, statistical analysis, sampling period

I. INTRODUCTION

This The electrical power system frequency plays a vital role in the determination of any power system performance [1][2]. Therefore, it's of significance to collect and analyze the frequency data in order to determine the systems condition and reliability [3]–[5]. A collaborative research between the Abubakar Tafawa Balewa University, Bauchi, Nigeria and Virginia Tech University, USA suggests that the Nigerian power system frequency is unstable and requires improvement in terms of frequency control [6] based on the frequency data collected using FDR of Fig. 1 in 2006 and 2007. In this paper, an FDR was installed and configured at the Power Holdings of Nigeria (PHCN) bulk power 330/132 kV substation, Bauchi for the purpose of collecting daily twenty-four hours data [7][8] to be used to determine the current extent of the system's frequency control. The FDR was sited at the power substation to ensure maximum electrical power availability. The raw frequency data collected was preprocessed and analyzed.

II. IMPLEMENTATION OF FDR MEASUREMENTS NETWORK

The measurement system is made up of an FDR, a Computer system, an Ethernet LAN Switch, and an NPort 5110 Moxa. The NPort Moxa is an innovative single port server that enables serial-interfaced devices to belong to a LAN network as shown in Fig. 2. The Virginia Tech Small Server Application, also known as FNET Receiver and the Moxa application are installed on the computer workstation. All the network devices were configured to belong to a single TCP/IP LAN for data acquisition and monitoring [9][6].

a. FDR measurement and data analysis

The FDR is configured to generate frequency data at a sampling rate of ten samples per second [10][8]. The FNet Receiver Application on the workstation converts the generated frequency data into Microsoft Access (MSA) database files saved in a well-organized and easy to use format which are then converted into Matlab binary format for the performance of analyses, such as plotting the data points, signal processing, etc [11]. Fig. 3 shows the procedure adopted for the conversion of MSA format into that of Matlab.

Fig. 1. Frequency disturbance recorder (FDR)

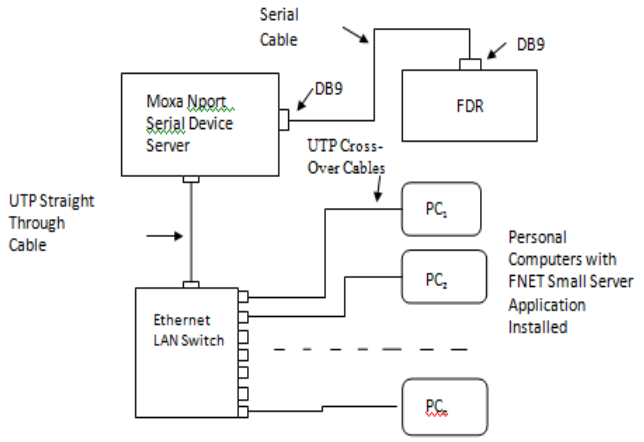


Fig. 2. The FNET structure)

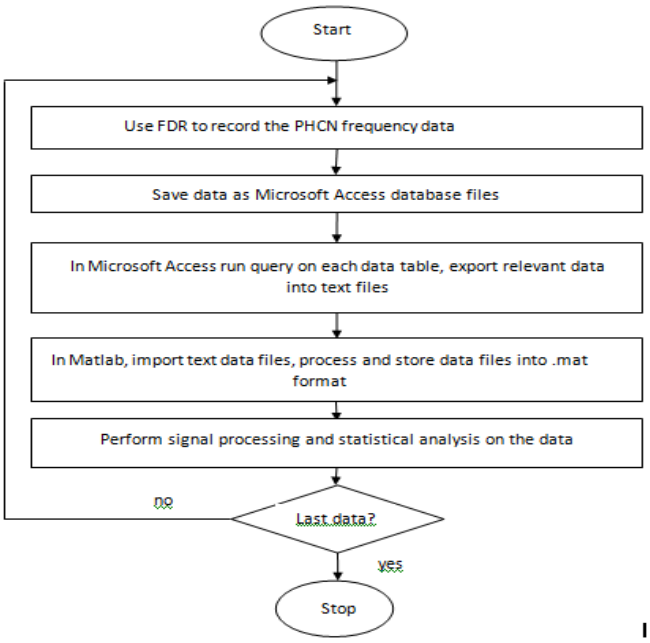


Fig. 3. The Algorithm for processing the generated frequency data

b. FDR data and signal processing

The electrical power system frequency is supposed to be steady, as it characterizes the system electromechanical nature. Therefore, the frequency data points that are significantly above or below the allowable limits in the measurements are removed. The measured data variability comes, among others, as a result of sudden variations in the power system local loads resulting into noise, outliers and spikes [6][11]. Outliers refer to frequency over 52 Hz or below 48 Hz.

III. RESULTS, DISCUSSION AND ANALYSIS

The FDR data acquisition and monitoring system has the FNET small server application as its main component. This application establishes FDR communication link, collects raw data, logs it in the MSA database, as well as displays it graphically and numerically as shown in Fig. 4. [6] On the graphical interface, the status of the communication link and the FDR location information are displayed as in Fig. 4. The real time waveform of the measured server frequency can also be seen.

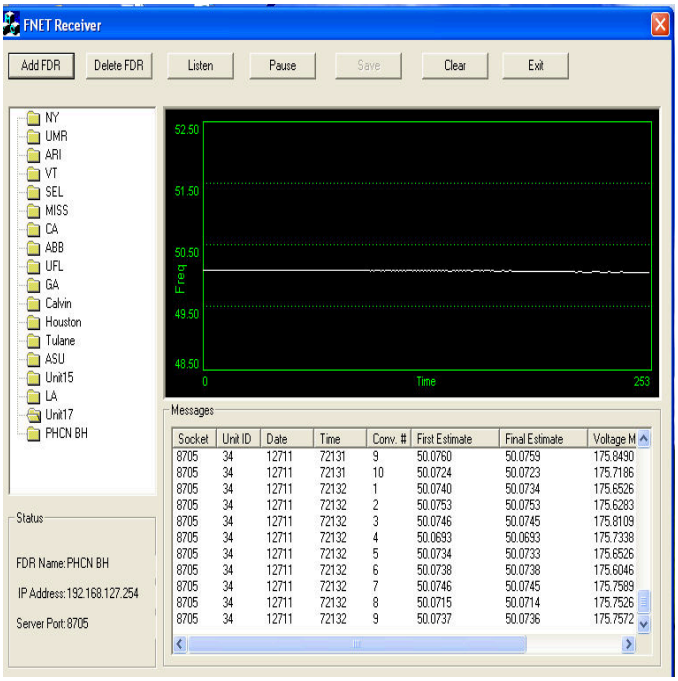


Fig. 4. Data Collection with FDR FNET small server program at Bauchi PHCN bulk power substation

The Moxa Serial server device, Nport 5110, is responsible for establishing communication link with the FDR unit alongside the backend database. The server communicates with the FDR unit via TCP/IP and socket communication. The FDR connects with the server through socket communication and TCP/IP.

a. Results and analysis

The tables that represent the raw frequency data contain very lengthy rows of the necessary columns at the sampling rate of ten rows per second. At this sampling rate, to represent a twenty-four hours data requires a table of 864000 rows which could make the least sense when compared with graphically visualized data. As such the daily recorded FDR frequency data are processed as in Figs. 7-10 and plotted against time to observe the dynamics and also to see if the frequency values are within the allowable tolerance of 2.5 % from the system's nominal value of 50 Hz. The relevant statistical indices were also computed and recorded from the processed signal..

b. Frequency dynamics observed from prolonged FDR recordings

Samples of different FDR raw data taken on various days together with the smoothed data are shown in Fig. 6. As can be seen, the measured frequency is prone to noise, spikes, and in some cases outliers.

Statistical analysis was taken for all the days of the record after necessary signal processing that removes outliers, spikes and smoothen the data. The statistical analysis of the smoothed data is given in Table 1.

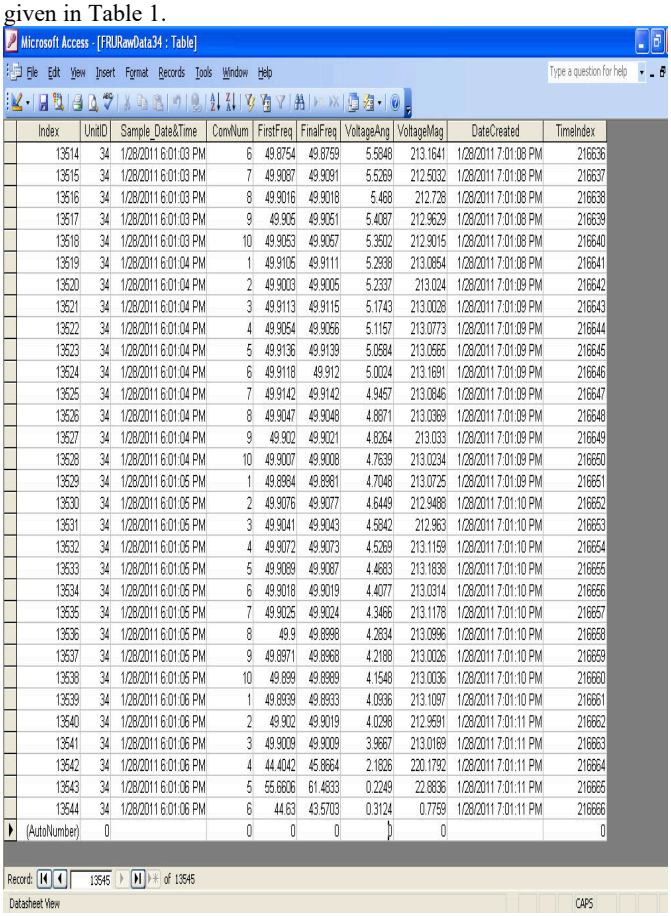


Fig. 5. Sample of stored FDR data in MSA for the FDR recordings

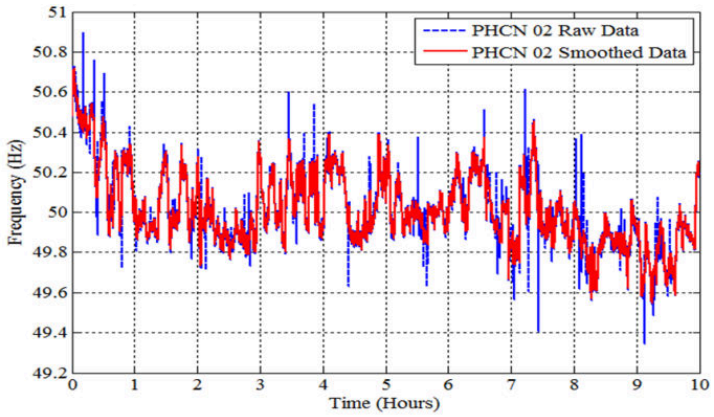


Fig. 6. Raw FDR and smoothed data

TABLE IV. STATISTICAL INDICES FOR THE SMOOTHED FREQUENCY

	Frequency (Hz)								
	Jan 28	Jan 29	Jan 30	Jan 31	Feb 01	Feb 02	Feb 03	Feb 04	Feb 05
Min.	49.7	49.55	49.67	49.48	49.25	49.54	48.86	48.8	48.91
Max.	50.6	50.65	50.61	50.61	50.59	50.72	51.35	51.42	51.42
Mean	50.09	50.11	50.08	50.03	50.02	50.02	50.09	50.05	50.35
Median	50.08	50.07	50.02	49.98	49.96	49.98	50.04	50.00	50.29
Mode	49.81	49.92	49.88	49.92	49.94	49.88	49.65	49.98	49.80
Std	0.1833	0.2006	0.2043	0.1965	0.1795	0.1863	0.5043	0.4286	0.5561
Range	0.9023	1.096	0.9454	1.128	1.345	1.176	2.494	2.617	2.511

Plots of various frequency dynamics meadured in different days are made as shown in the following figures.

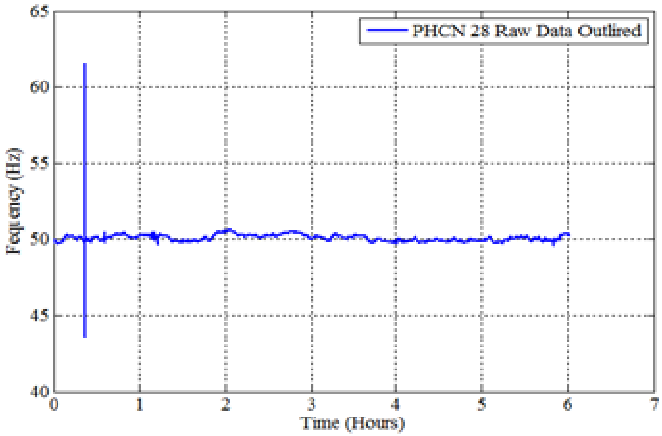


Fig. 7. Frequency dynamics with outliers

In Fig. 7 outliers are seen. These represent frequency values that significantly falls out of the nominal frequency range.

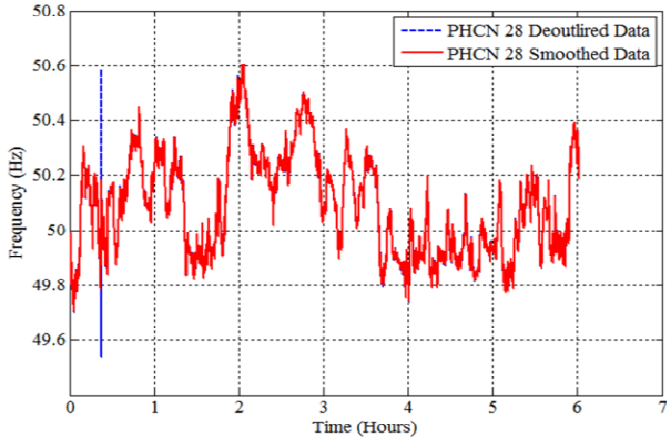


Fig. 8. Raw FDR and smoothed data with spikes

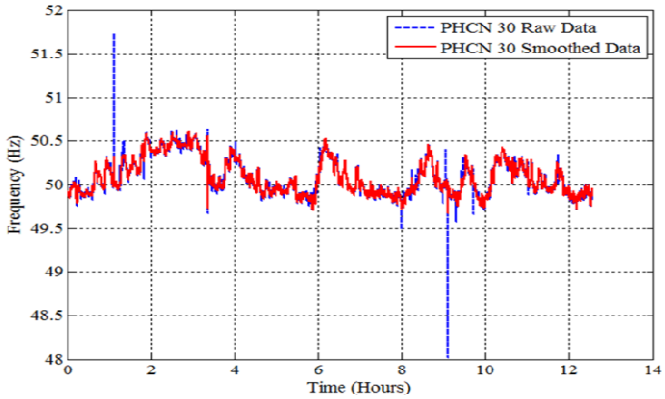


Fig. 9. Raw FDR and smoothed data with spikes and outliers

c. Summary and Conclusion

From the established FDR network and its recorded data, there is presence of noise, spikes and outliers in the Nigerian power system frequency, characterizing the local loads rapid variations, which were conditioned and preprocessed. From the results presented it can be observed that the frequency regulation band of the electrical power system is within the allowable range of ± 2.5 %, that is, between 48.75 Hz and 51.25 Hz despite the load dynamics. This demonstrate frequency regulation improvement regulated by the frequency relays in the power substation.

REFERENCES

[1] *IEEE Standard Conformance Test Procedure for Equipment Interconnecting Distributed Resources with Electric Power Systems*, no. July. 2005.

[2] E. Oleka, S. Ndubisi, and G. L. Lebby, "Synchrophasor for Enhancing Operational Reliability of Nigerian Electric Power Grid," pp. 5–10, 2017.

[3] S. O. Oyedepo, O. P. Babalola, S. C. Nwanya, O. Kilanko, A. O. Abidakun, and O. L. Agberegha, Towards a Sustainable Electricity Supply in Nigeria : The Role of Decentralized Renewable Energy System, vol. 2, no. 4. 2018.

[4] O. A. Somefun, "The Erratic Electric Power Supply in Nigeria : Causes and Remedy The Erratic Electric Power Supply in Nigeria : Causes and Remedy," no. May, 2016.

[5] V. Emodi and K.-J. Boo, "Sustainable energy development in Nigeria: Current status and policy options," Renew. Sustain. Energy Rev., vol. 51, pp. 356–381, 2015.

[6] L. Vanfretti et al., "Estimation of the Nigerian Power System Electromechanical Modes using FDR Measurements," in the 9th International Conference on Power System Operations and Planning (ISCPSOP), 2010.

[7] Z. Lin et al., "Application of Wide Area Measurement Systems to Islanding Detection of Bulk Power Systems," vol. 28, no. 2, pp. 2006–2015, 2013.

[8] Y. Zhang *et al.*, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 159–167, 2010.

[9] Y. L. Y. Liu, "A US-Wide Power Systems Frequency Monitoring Network," *2006 IEEE PES Power Syst. Conf. Expo.*, pp. 1–8, 2006.

[10] T. Xia et al., "Wide-area Frequency Based Event Location Estimation," Power Eng. Soc. Gen. Meet. 2007. IEEE, pp. 1–7, 2007.

[11] *IEEE, IEEE Standard Conformance Test Procedure for Equipment Interconnecting Distributed Resources with Electric Power Systems, no. July. 2005*...

INDUSTRIALIZATION OF THE NIGERIAN ECONOMY: THE IMPERATIVES OF IMBIBING ARTIFICIAL INTELLIGENCE AND ROBOTICS FOR NATIONAL GROWTH AND DEVELOPMENT

Akang, V. I., Afolayan, M.O., Iorpenda, M. J., Akang, J. V.

Corresponding email: akangvictor2@gmail.com

Abstract—**This paper examines the potential impact which artificial intelligence (A.I.) and robotics will have on the Nigerian economy with a view to corroborating its enhancing of positive growth. Artificial Intelligence is the latest and best form of automation, which will enhance a better product output and win global work challenge. Rapid technological progress in AI has been predicted to lead to mass unemployment, rising inequality, and higher productivity growth through automation. A novel approach is used to outline endogenous growth theory, but modified to allow for demand-side constraints. The paper provides a theoretical explanation of this in the context of rapid progress in AI. Safer and a better power outage, inculcating inventive ingenuity in the Nigerian populace, growth of the country's gross domestic product, which has been dubbed as near impossible ventures over ages, but A.I. now seems poised to automate many tasks once thought to be out of reach, from driving cars to making medical resolutions and beyond. How will this affect economic growth in the Nigerian economy and foster job creation, and bridge the gap in the division of income between labour and capital? What about the case of job losses that will be followed? How will the linkages between A.I. and growth be mediated by firm-level considerations, including organization and market structure? The goal throughout is to refine a set of critical questions about A.I. and economic growth and to contribute to shaping an agenda for the field of Artificial Intelligence and Robotics. In the pursuit of a more secure, stable and sustainable world, Nigeria should seek to enhance her human, institutional and infrastructural capacity. To do so she needs a solid base of technologically prepared people to effectively improve her economic and quality of life. Such a base will facilitate the infusion of high level technologically driven products of a high standard, with a comparative advantage above that of a lot of others, without the adoption of AI. In this paper we critically re-assess these predictions by; surveying the recent literature and incorporating AI-facilitated automation into a product variety-model, frequently used for academic interest and economic growth theory, but modified to allow for demand-side constraints.**

Keywords—Industrialization, industrial revolution, artificial intelligence, robots, digitization, economic growth.

I. INTRODUCTION

Industrialization is the creation of industries, which is used for the production of goods and services, it can be visualized as a product of revolution. It is actually the development of locations in readiness for production which is intended to lead to increased productivity. One of the developed nations, the United States of America, is a model of great economies. Though, England was the first to industrialize, other parts of Europe followed; France and Germany, in the course of the industrial revolution. The industrial revolution took place more than a century ago, as production of goods moved from home businesses, where products were generally made by hand, to machine-aided production in factories which were sites for industrialization. This revolution, which involved major transformation in areas such as banking, transportation, manufacturing, and communications, transformed peoples' daily lives especially that of the average American as much as any single event in U.S. history. This land mark trajectory was also witnessed in Europe. These are classed as milestones. The Milestones in Science and Discovery set is based on a simple but powerful idea—that science and

technology are not separate from peoples' daily lives. Technology is a basic part of that culture (Henderson, 2006). This milestone has led to the emergence of artificial intelligence (AI).

Artificial Intelligence (AI) is the hub of today's innovation which is a technology to behold. The definition of AI coined in the 1950s and still in use is "the capability of a machine to imitate intelligent human behaviour." AI gets more interesting when the machine cannot just imitate, but match or even exceed human performance—it gives us the opportunity to offload repetitive tasks, or even to get computers to do jobs more safely and efficiently than we can. Practically speaking, when people think of AI today, they almost always mean *machine learning*: training a machine to learn a desired behaviour. It is highly encouraged by the increasing availability of reprogrammable computational power and increased productivity at super high speed and pin-point accuracy. AI applications offer intriguing possibilities for promoting economic growth and tackling a wide variety of longstanding problems in the universe. The resulting disruptive impact will be massive and could well be revolutionary. The aspect of AI which propels it as branch is the automation, simulation, robotics, mechatronics and modeling.

Automation, which could also be associated with Industrial Control Systems is often referred to as ICS, has an interesting and fairly long history, though its emergence in Africa is relatively new. Today, it is quite common to see discussions of industrial controls paired cyber/physical security; however, that is a relatively recent phenomenon. In Nigeria establishments (private and or public) make use of this AI in a variety of forms (though underutilized). This is not reckoned as AI, but it is, as for instance the issuance of the bank verification number (BVN) which is linked to multiple platforms and which also helps to detect ownership of multiple bank accounts, the detection of the location of someone calling with a mobile phone by triangulation, opening and shutting of doors and windows by the insertion of smart cards, self serving vending machines just to mention a few. This paper will discuss the implication of adopting AI in the Nigerian state with the ripple advantages in today's control systems and provide a sound basis as a significant concern to their reliability and predictability (Hayden *et al,* 2014).

Computer simulations are routinely used in fundamental research to help understand experimental measurements, and to replace {for example} growth and fabrication of expensive samples/experiments where possible. In an industrial context, product and device design can often be done much more cost effectively if carried out virtually through simulation rather than through building and testing prototypes. This is in particular so in areas where samples are expensive such as nanotechnology (where it is expensive to create small things) and aerospace industry (where it is expensive to build large things).(Fangohr, 2015).

The main objectives of this paper is;

- To give a detailed range of analytical treatise about the contexts under which organizations deploy AI implementations, and how they manage these.
- To consider AI's economic impacts at the level of an organization, especially on performance, results, and processes, in addition to the national macroeconomic view so often discussed.
- To inform exigent foray about AI's productivity and national economic impacts.

The Industrial Revolution is one of the most celebrated watersheds in human history. It is no longer regarded as the abrupt discontinuity that its name suggests, for it was the result of an economic expansion that started in the sixteenth century. Nevertheless, the eighteenth century does represent a decisive break in the history of technology and the economy. The famous inventions such as the spinning jenny, the steam engine, coke smelting, and so forth, for they marked the commencement of a process that has carried the West, at least, to the mass prosperity of the twenty-first century. (Allen, 2006). The term "Robot" can be traced back to Karel Capek's play "R.U.R. Rossum's universal robots" (in 1921) that comes from the Czech word for "corvee".

The fear which has been in existence since time immemorial still forms a major reason behind the resistance to full adoption of automation which is seen as making machines more intelligent, i.e. application of AI in national economy. The lack of industrialization has made Nigeria not to be unique among the comity of nations, considering its abundant human resources. Not unique in the sense of being acclaimed the giant of Africa, by mere population and land mass. At present, most countries attempting to industrialize adopt well defined strategies, such as import-substitution, adoption of backward Engineering in the production of equipment from new technologies. Employing a strategy for industrialization, explains H. J. Bruton, implies belief in one (or both) of two things: first:

"that by some twist or torment of the economy, or by some gimmick, the underdeveloped stagnant economy can be set in motion", or secondly that the entire nature of the economy must be modified before sustained growth is possible. It should become clear in the ensuing discussion that these concepts have applied to the Nigerian economy. As for the Nigerian case the Industrial Revolution was a catastrophic failure as she did not achieve the target of Industrialization.

One of the greatest challenges of the Sub - Saharan African economies today is the high rate of unemployment that has maintained a rising trend over the years, bedeviling the continent. The problem of unemployment has been of great concern to the economists and policy makers in Nigeria since early 1980s. The effect of financial crisis on public and private sectors has led to renewed attention on the phenomenon. It is a widely accepted view in economics that the growth rate of the Gross Domestic Product (GDP) of an economy increases employment and reduces unemployment. The three most significant elements for the economy overall are productivity, income distribution and unemployment. The theoretical proposition relating output and unemployment has been proposed by Okun (1962). This relationship is among the most famous in macroeconomics theory and has been found to hold for several countries and regions mainly, in developed countries (Christopoulos, 2004; Daniels and Ejara, 2009).

Manufacturing in 2050 will look very different from today, and will be virtually unrecognizable from that of thirty years ago. This will largely be as a result of the adoption of AI. Successful firms will be capable of rapidly adapting their physical and intellectual infrastructures to exploit changes in technology as manufacturing becomes faster, more responsive to changing global markets and closer to customers. (Era, Opportunity, For, & Uk, n.d.)

The world is on the cusp of a digital revolution that will change fundamentally the way we live, work, and communicate. The transformation happening inside the telecommunications industry has a big impact on the outside world, with the emergence of new promising digital technologies. Artificial intelligence (AI) is one of the emerging technologies with special reference to that of the future. It is in it's early days, but the family of technologies that encompass AI have narrowed the gap with, matched, or in some cases exceeded human performance in different areas. Computer vision systems are getting more accurate, detecting objects at large scales better than the average human performance. Speech recognition systems can now identify language from phone calls and voice records with accuracy levels that surpasses human abilities. AI solutions have the potential to transform areas as diverse and critical as education, healthcare, finance, mobility, and energy − promising to help accelerate progress towards the future Agenda for Economic Growth and Sustainable Development.

Nigerian Economic Growth

Nigeria is a middle income, mixed economy and emerging market, with expanding financial, service, communications, and entertainment sectors. It is not having a marked rating in the area of automation and robotics as at 2018. It is ranked 30th in the world in terms of GDP as of 2011, and its emergence, though currently with an underperforming manufacturing sector is the third-largest on the continent, producing a large proportion of goods and services for the West African region. Nigerian GDP at purchasing power parity more than doubled from $170.7 billion in 2005 to $413.4 billion in 2011, although estimates of the size of the informal sector (which though is not included in official figures) put the actual numbers closer to $520 billion. Correspondingly, the GDP per capita doubled from $1200 per person in 2005 to an estimated $2,600 per person in 2011 (again, with the inclusion of the informal sector, it is estimated that GDP per capita hovers around $3,500 per person). It is the largest economy in the West African Region, 3rd largest economy in Africa (behind South Africa and Egypt). Unemployment has been a major problem for most countries across the world. The USA for example has increased from 5 percent in 2007 to 9 percent so far in 2011. Spain increased from 8.6 percent to 21.5 percent; UK from 5.3 percent to 8.1 percent. Ireland currently stands at 14.3 percent from 4.8 percent, Latvia from 5.4 percent to 16.5 percent, Greece from 8.1 percent to 18.4 percent, and Italy from 6.7 percent to 8.3 percent. The average for the Euro area is 10.7 percent. Even within the African continent, unemployment has risen with South Africa, Africa's largest economy having a higher rate than Nigeria at 25 percent, Angola at 25 percent, Botswana at 17.5 percent, Egypt at 11.8 percent, Kenya at 11.7 percent, and Namibia at 51 percent (Akeju &Olanipekun, 2015)

The developing or less developed countries (LDC) are suffering from one crises to the other, from which the former UN secretary-General (of blessed memory), Mr. Kofi Annan in 2002, had stated: *"Let me challenge all of you to help mobilize global science and technology to tackle the interlocking*

crises of hunger, disease, environmental degradation and conflict that are holding back the developing world."

Scholars and policy makers alike remain fascinated by the development experiences of a group of emerging economies – Brazil, Russia, India, China and South Africa – often referred to as the BRICS (O'Neill, 2001). These economies have a sizeable impact on the global economy. In terms of combined GDP they are already larger than USA and the European Union. Whereas BRICS only accounted for less than 4 per cent of world exports during the early 1980s, by 2010 their combined share reached 13 per cent of world exports. They are regional economic leaders and they are attracting substantial amounts of foreign direct investment. This makes them interesting and important global decision-makers. But how much do their development experiences, particularly their success or failures in industrialization and manufacturing compare, and what can other countries, aspiring to industrial development learn from them?

Our concern with industrialization in general, and manufacturing specifically, is based on the recognition that, a better and more precise product acceptable by the generality of the populace can be made which will ultimately gain the comparative advantage in demand over a lot of other economies. It can be agreed that within the traditions of endogenous growth theory, evolutionary economics and institutional economics that manufacturing (industrialization) is important for economic development (Szirmai, 2012). Technological progress is in turn, necessary for successful industrialization (Von Tunzelmann, 1997; Fu *et al*, 2010). Hence in this paper, a critical look into the prospect of the projection of the vision 2020:20, now extended to vision 2030:20 will be examined, it also examines the potential economic and labour market effects of the adoption of AI.

II. LITERATURE REVIEW

Technological progress and its diffusion were important for the First Industrial Revolution in Britain in the 18th century. It was also vital for the (second revolution) industrialization of Continental Europe and the USA in the 19th century, and for the (third revolution) industrialization of Japan and the East Asian new investment enterprises in the 20th century (O'Brien, 2001; Veloso & Soto, 2001). In each case catch up resulted from lagging countries accessing technology developed in leading nations, adapting it effectively to local circumstances, and subsequently relying more on indigenous innovation. Technology has become more important as ever, in the current modular and flexible production systems that has come to characterize the world economy. Indeed Marsh (2012) claims that the world is experiencing a new industrial revolution, led by networked production, mass customization and new technologies, such as 3-D printing.

Multinational Enterprises can diffuse technologies to developing countries in three ways: Firstly, by directly transferring technology to affiliate or joint ventures; Secondly, through spillover effects, and/or thirdly, through doing Research and Development within a developing country (Lloyd, 1996).

Production technology

The technological innovations accompanied simultaneously the social and economic changes, and this may be considered as a reciprocal process. The driving force of the Industrial revolution was the textile industry and particularly the cotton products. Cotton industry was not considerable through much of the 17th century but exploded one century later. Most cotton was produced in British colonies namely America and India. Because it was labour intensive, it fueled the traffic in African slaves to the American plantations. The cotton produced in America imported to the homeland for the use in textile industry, and the finished product was marketed in Europe and in other countries. At the beginning, having the technological superiority of production, England had almost the monopoly of cotton textiles. This position altered when America became independent and entered into the market as a competitor. From history, we also observed this competition which ended by the victory of America, so why since 1830's, the ordinary cotton textile is named "Amerikan bezi" in Turkey.

Humanoid robots were being developed since 1975 when Wabot-I was presented in Japan. The current Wabot-III already has some minor cognitive capabilities. Another humanoid robot is "Cog", developed in the MIT-AI-Laboratory in 1994. Honda's humanoid robot became well known in the public when presented back in 1999. Although it is

remote controlled by humans it can walk autonomously (on the floor and stairs).

In the late 1940s, Norbert Wiener led the development of cybernetics which is a combination of control theory, information science and biology that seeks to explain the common principles of control and communication in both animals and machines. In 1953, W. Grey Walter applied these principles in the creation of a robotic design termed 'Machina Speculatrix', which was later transformed into hardware as Grey Walter's tortoise. The tortoise consisted of two sensors, two actuators and two 'nerve' cells or vacuum tubes and it could demonstrate behaviors like seek light, head toward weak light, back away from bright light, turn, and recharge battery. Three decades after Walter, Valentine Braitenberg revived his behaviour ideas (Braitenberg, 1984). Braitenberg's vehicles used inhibitory and excitatory influences, directly coupling the sensors to the motors. Seemingly complex behaviour resulted from relatively simple sensor-motor combinations. However, just like Walter's tortoise, the Braitenberg vehicles were inflexible and could not be reprogrammed.

The history of genetic algorithms

Around the year 1850 Gregor Mendel developed his theory of genes, also called the theory of genetics. Genes are tiny bits of hereditary information found in the DNA of all living beings. Each gene describes an aspect of the organism to which it belongs. Every cell of an organism contains DNA and each DNA string contains all the genes of an organism. This means that every cell in its body contains all the necessary information for a complete description of the organism. During its development, the organism's DNA is used as a blueprint for the production of new cells. Usually, new organisms are created by combining DNA parts from two parents. In 1859 Charles Darwin published his book 'Origin of Species'. In this work he explains his theory of evolution, which describes how the emergence of life as we know it, can be explained as a natural phenomenon. A very important aspect of Darwin's theory is the 'survival of the fittest', also called 'natural selection'. This part of his theory explains that an organism that is well fitted for survival has a greater chance to reproduce and pass along its genes to future generations. Winners stay alive and losers become extinct. Just by surviving, an organism proves its superiority. As the environment slowly changes, the organisms can adapt gradually and evolve with it. Nowadays, there are still many opponents to Darwin's theory, mostly religious people, but in the scientific world the basics of Darwin's theory are widely accepted.

Genetic algorithms (GAs) are a way to use Darwin's natural selection theory and Mendel's theory of genes to solve certain problems on a computer. GAs are most commonly used for optimization problems that are difficult to analyze. Many of the terms used with GAs are derived from the biological phenomenon. The GA as it is known today was first described by John Holland in the 1960s and further developed by Holland and his students and colleagues at the University of Michigan (Holland, 1992).

a. HISTORY OF AI

The birth of the computer took place when the first calculator machines were developed, from the mechanical calculator of Babbage, to the electromechanical calculator of Torres-Quevedo. The dawn of automata theory can be traced back to World War II with what was known as the "codebreakers". The amount of operations required to decode the German trigrams of the Enigma machine, without knowing the rotor's position, proved to be too challenging to be solved manually. The inclusion of automata theory in computing conceived the first logical machines to account for operations such as generating, codifying, storing and using information. Indeed, these four tasks are the basic operations of information processing performed by humans. The pioneering work by Ramón y Cajal marked the birth of neuroscience, although many neurological structures and stimulus responses were already known and studied before him. For the first time in history the concept of "neuron" was proposed. McClulloch and Pitts further developed a connection between automata theory and neuroscience, proposing the first artificial neuron which, years later, gave rise to the first computational intelligence algorithm, namely "the perceptron". This idea generated great interest among prominent scientists of the time, such as Von Neumann, who was the pioneer of modern computers and set the foundation for the connectionism movement. The name behind the idea of AI is John McCarthy, who began research on the subject in 1955 and assumed that each aspect of learning and other

domains of intelligence can be described so precisely that they can be simulated by a machine. Even the terms 'artificial intelligence' and 'intelligent human behaviour' are not clearly defined, however. Artificial intelligence describes the work processes of machines that would require intelligence if performed by humans. The term 'artificial intelligence' thus means 'investigating intelligent problem-solving behaviour and creating intelligent computer systems'.

b. CHALLENGES AND RISKS OF AI

Chen (2016) takes a different approach and predicts much lower impacts, estimating the cumulative economic impact of AI from 2016 to 2026 as lying between $1.5 and $3 trillion (0.15 to 0.3 percent of global GDP). They generate four ranges for estimated impacts, using the results of previous studies to calculate the economic impacts of AI as if they were the same as computers in general, broadband internet, mobile phones or industrial robotics, and then triangulating to land on a central range. Because of limitations in the previous studies they rely on, some of the estimated economic impacts are for high income economies only. They also estimate the economic impacts of AI based on the expected economic effects of assumed future industry and private capital investment in AI technologies. This led to a markedly lower figure for economic impact of only around $500 billion. They presented reasons why this method would only capture a portion of the total economic effects of AI. It is common to distinguish between the 'production' effects of technology (the economic impacts from growth in the firms that make the technology) and the 'use' effects (the economic impacts from firms using the technology). The used effects are much larger than the production effects in ICT, i.e. how firms outside the ICT sector used the technology is more significant for economic performance than how quickly firms in the ICT sector itself grew. This is consistent with the previously mentioned analysis. In fact, absolute investments in AI technologies are currently quite small, even if they are growing quickly.

c. THE IMPERATIVE FOR THE INEVITABLE

For decades, the potential for AI in automation in particular, in the form of smart devices such as a wheelchair (taking a life threatening case) to aid those with motor, or cognitive, impairments had been recognized. It is a paradox that often the more severe a person's motor impairment, the more challenging it is for them to operate the very assistive machines which might enhance their quality of life. These devices can be likened to the confounding issues plaguing the Nigerian economy by incorporating robotics and intelligence into assistive machines, turning the machine into a kind of robot, and offloading some of the control burden from the user. Robots already synthetically sense, act in and reason about the world, and these technologies can be leveraged to help bridge the gap left by sensory, motor or cognitive impairments in the users of assistive machines.

The capability of AI to perform increasingly more tasks that were previously considered the sole preserve of humans and expected ongoing continuous improvement has a range of consequences. There is no shortage of reports discussing the potential economic impacts of AI at national and global level or by sector.

III. CASE STUDIES

Studies carried out from three organizations, spanning different industries, geography, sizes, and use cases of AI and machine learning (ML) techniques. AI, a notoriously broad and nebulous term, refers to the science and engineering of systems that could be described as making machines intelligent, able to function appropriately and with foresight in an environment. ML, presently the most successful sub-field of AI, enables computer systems to learn from data to improve performing a task rather than designing a static algorithm to perform a specific task. Of note, each case study varied in the AI/ML techniques used to achieve organizational ends, and organizations often used a combination of approaches.

The Tata Steel Europe case, based in the Netherlands, describes how an incumbent steel manufacturer, facing significant industry headwinds, launched an internally-developed AI strategy (referred to internally as "Advanced Analytics") to optimize its production processes with both ML models and workforce retraining to build and integrate them. The company reports significant economic benefits and productivity gains with a lower investment than alternative, higher cost capital investments. Notably, the company reports avoiding immediate layoffs by retraining many in its workforce to develop and implement the AI models - though this may not always remain the case, nor may it be the equilibrium that emerges across the entire industry.

The Axis Bank case, based in India, explores how an AI chatbot addressed a growing need to improve customer service in a cost-effective and scalable way

through automation. In addition to examining AI's impact on Axis Bank's internal workforce, the case discusses how labour impacts cascaded outside the bank and across sectors to third-party customer service providers and technical vendors.

The Zymergen case examines how a biotech startup in the San Francisco Bay Area is applying machine learning and automation as an alternative to conventional R&D and scientific experimentation practices. The case includes observations on the productivity impact of AI in that setting, as well as the labour implications for a highly educated workforce (e.g., scientists with Ph.D.'s).

Each of the organizations profiled has different strategic motivations and thus different approaches to developing AI/ML-related applications. Tata Steel Europe is an incumbent trying to build an AI capability in-house through training its existing labour resources. Axis Bank, another incumbent, chose to bring in a third-party AI vendor to help develop its chatbot capability. Finally, Zymergen is an "AI-native" startup developing material science R&D services.

IV. FINDINGS OR OBSERVATION

Despite the variety in sector, geography, size and AI applications of the firms we studied (as reported), three findings emerged:

1. Successful adoption of new AI systems required buy-in from management and the workforce alike. In doing so, having intelligible ML models was of paramount importance.
2. Each firm reported productivity or Return On Investment (ROI) growth in the short term, but in all cases, these effects could not be attributed solely to AI, as other process changes always accompanied the introduction of AI technologies.
3. There were varying impacts on the local workforce due to the adoption of AI systems and no reported layoffs to date, but the effects were difficult to measure internally and, in some cases, cascaded beyond the firms studied into their third-party vendors and broader business ecosystems.

A. Benchmarking AI using Sector Investments

A prerequisite for the widespread adoption of any technology is the large amount of research, development, and investment that it takes to bring that technology to market. Therefore, the high levels of current investment by industry players and venture capital firms in AI provide a telling sign of the current state of, and impending advances in, AI. In 2014 and 2015 alone, eight major global technological firms made at least 26 acquisitions (totaling over $5 billion) of companies developing AI technology. Private investment in AI has also taken the form of in-house spending in addition to startup acquisitions. For example, Facebook's AI Research laboratory, Google's Machine Intelligence laboratory, and Microsoft's Machine Learning and Artificial Intelligence research division are all making advances in AI technology and investing in the industry's top talent. Additionally, between 2010 and 2015, nearly $5 billion in venture capital funding was invested in firms across the globe developing and employing AI technology.

B. Summary of Robotics Sale as at 2016

To buttress how the developed nations are carrying on the business of maintaining a high level of industrialization, the business of manufacture and sale of robots will give credence to the discuss. 2015 is by far the year where the highest volume of sale was ever recorded. In 2015, robot sales increased by 15% to 253,748 units, by far the highest level ever recorded in one year. The main driver of the growth in 2015 was the general industry with an increase of 33% compared to 2014, in particular the electronics industry with an increase of (+41%), metal industry (+39%), the chemical, plastics and rubber industry (+16%). The robot sales in the automotive industry only moderately increased in 2015 after a five-year period of continued increase.

China has significantly expanded its leading position as the biggest market with a share of 27% of the total supply in 2015.

Since 2010, the demand for industrial robots has accelerated considerably due to the ongoing trend toward automation and the continued innovative technical improvements in industrial robots. Between 2010 and 2015, the average robot sales increase was at 16% per year (CAGR). The number of robot installations had never increased so heavily before. Between 2005 and 2008, the average annual number of robots sold was about 115,000 units. 2009 is excluded because the economic and financial crisis in 2008/2009 caused an exceptional plunge in robot sales. Between 2010 and 2015, the average annual supply rose to about 183,000 units. This is an increase of about 59% and a clear sign of the significant rise in demand for industrial robots worldwide.

Asia, the most important region

Asia (Australia and New Zealand included in the category) is still the world's strongest growth market. This region saw a total of 160,600 units sold in 2015 – a rise of 19%. This was the highest sales level ever

recorded for the fourth year in a row. Industrial robot sales to the second largest market, **Europe**, increased by 10% to 50,100 units (again a new peak like in 2014). About 38,100 industrial robots were shipped to the **Americas**, 17% more than in 2014, establishing a new peak for the fourth year in a row.

75% of the global robot sales in five countries

There are five major markets representing 75% of the total sales volume in 2015: China, the Republic of Korea, Japan, the United States, and Germany. Sales volume increased from 70% in 2014. Since 2013 China is the biggest robot market in the world with a continued dynamic growth.

With sales of about 68,600 industrial robots in 2015 – an increase of 20% compared to 2014 - China alone surpassed Europe's total sales volume (50,100 units).

Chinese robot suppliers installed about 20,400 units according to the information from the China Robot Industry Alliance (CRIA). Their sales volume was about 29% higher than in 2014. Foreign robot suppliers increased their sales by 17% to 48,100 units (including robots produced by international robot suppliers in China). The market share of Chinese robot suppliers grew from 25% in 2013 to 29% in 2015.

Between 2010 and 2015, total supply of industrial robots increased by about 36% per year on average. About 38,300 units were sold to the Republic of Korea, 55% more than in 2014. The increase is partly due to a number of companies which started to report their data only in 2015. The actual growth rate in 2015 is estimated at about 30% to 35%.

In 2015, robot sales in Japan increased by 20% to about 35,000 units reaching the highest level since 2007 (36,100 units). Robot sales in Japan followed a decreasing trend between 2005 (reaching the peak at 44,000 units) and 2009 (when sales dropped to only 12,767 units). Between 2010 and 2015, robot sales increased by 10% on average per year (CAGR).

Increase in robot installations in the United States continued in 2015, by 5% to the peak of 27,504 units. Driver of this continued growth since 2010 was the ongoing trend to automate production in order to strengthen American industries on the global market and to keep manufacturing at home, and in some cases, to bring back manufacturing that had previously been sent overseas.

Germany is the fifth largest robot market in the world. In 2015, the number of robots sold increased slightly to a new record high at 20,105 units compared to 2014 (20,051 units). In spite of the high robot density of 301 units per 10,000 employees, annual sales are still very high in Germany. Between 2010 and 2015, annual sales of industrial robots increased by an average of 7% in Germany (CAGR).

Other important Asian markets

Since 2013, Taiwan has ranked sixth among the most important robot markets in the world with regard to the annual supply. Robot installations increased considerably between 2010 and 2015, by 17% on average per year (CAGR). In 2015, robot sales increased by 4% to about 7,200 units, a new peak. However, the number of units in Taiwan is far below the number in Germany, which ranked fifth with 20,100 units.

Thailand is also a growing robot market in Asia. In 2015, the supply of industrial robots declined by 30% to about 2,600 units. The peak was reached in 2012 with about 4,000 units. Robot sales to India slightly decreased to about 2,100 units in 2015. This is slightly lower than the peak level of 2014. Robot supplies to other Southeast Asian countries like Malaysia, Singapore and Vietnam increased considerably in 2015.

Other important European markets

Italy is the second largest robot market in Europe after Germany. Worldwide, it ranked 7th in 2015 like in 2014. Robot investments continued to increase in 2015. Total sales of industrial robots were up by 7%, to almost 6,700 units, which is a new peak. Between 2010 and 2013, annual robot sales to Italy were rather weak due to the critical economic situation. The French robot market continued to increase in 2015, by 3% to almost 3,045 units. In Spain, sales of industrial robots surged by 63% to about 3,800 units, the highest number ever recorded. Sales of industrial robots to the United Kingdom further decreased in 2015 to 1,645 units. Robot installations in all other Western European and Nordic countries as well as in the Czech Republic and in Poland were considerably up in 2015. Sales to Turkey continued to increase in 2015.

Other important American markets

Mexico has become an important emerging market for industrial robots. Robot sales more than doubled, to about 5,500 units in 2015, by far the highest sales volume ever registered for the country. In Canada, robot sales surged by 49% to about 3,500 units in 2015, a new peak. In 2015, robot sales to **Brazil** recovered by 11% to 1,400 units.

Main drivers of the growth: electrical/electronics industry, metal industry and rubber and plastics industry

Between 2010 and 2014, the **automotive industry** – the most important customer of industrial robots – had considerably increased investments in industrial robots worldwide. In 2015, robot sales increased by 4% to about 97,500 units establishing again a new peak for the fifth time in a row. The share of the total supply in 2015 was about 38%. Between 2010 and 2015, robot sales to the automotive industry increased by 20% on average per year (CAGR). Since 2010, investments in new production capacities in the emerging markets as well as investments in production modernization in major car producing countries have caused the number of robot installations to rise. Using new materials, developing energy efficient drive systems, as well as high competition in all major car markets, pushed for investments despite the existing overcapacities.

Robot sales to the **electrical/electronics industry** (including computers and equipment, radio, TV and communication devices, medical equipment, precision and optical instruments) were significantly up in 2015, by 41% to 64,600 units, establishing again a new peak. This was more than double the sales volume in 2010. Share of the total supply in 2015 was about 25%. The rising demand for electronic products and new products, as well as the need to automate production (particularly in low wage countries), were the driving factors for the boost in sales.

In 2015, sales to the **metal and machinery industry** were significantly up by 39% to 29,450 units, a new peak, for the third year in a row. The share of the total supply was 12%. Since 2010, sales of all subsectors (basic metals, metal products, industrial machinery) have followed an increasing trend. Between 2010 and 2015, the average annual growth rate was 26%.

The **rubber and plastics industry** has continuously increased the number of robot installations since 2009 from about 5,800 units to 17,300 units in 2015, again a new peak. Share of the total supply in 2015 was about 7%. Between 2010 and 2015, sales were up by 14% on average per year.

Sales to **all industries, except for automotive and electrical/electronics,** increased by 27% in 2015. Between 2010 and 2015, the average growth rate per year was 19%. The respective growth rate for the automotive industry was 20% and for the electrical/electronics industry 15%. This is a clear sign that not only the main customer industries (automotive industry and electrical/electronics industry) have increased robot installations considerably in recent years, but other industries have done so as well.

Continued considerable increase of worldwide operational stock

The total worldwide stock of operational industrial robots at the end of 2015 increased by 11% to about 1.6 million units. Since 2010, the stock has been increasing considerably by 9% on average per year.

Value of the global market was up to US$11.1 billion

In 2015, the sales value increased by 9% to a new peak at US$11.1 billion. It should be noted that the figures cited above generally do not include the cost of software, peripherals and systems engineering. Including the mentioned costs might result in the actual robot systems' market value to be about three times as high.

The worldwide market value for robot systems in 2015 is therefore estimated to be
US$35 billion.

High potential for robot installations in many countries

When comparing the distribution of multipurpose industrial robots in various countries, the robot stock, expressed in the total number of units, can sometimes be a misleading measure. In order to take into account, the differences in the size of the manufacturing industry in various countries, it is preferable to use a measure of robot density. One such measure of robot density is the number of multipurpose industrial robots per 10,000 persons employed in manufacturing industry or in the automotive industry or in the "general industry" (which is all industries excluding the automotive industry). The average global robot density is about 69 industrial robots installed per 10,000 employees in the manufacturing industry in 2015. The most automated markets are the Republic of Korea, Singapore, Japan and Germany. The Republic of Korea has by far the highest robot density in the manufacturing industry. 531 industrial robots were in operation in 2015 per 10,000 employees. The rate increased from 241 units in 2009 due to continued installation of a large volume of robots since 2010, particularly in the electrical/electronics industry and in the automotive industry. It is followed by Singapore with a rate of 398 robots in 2016. The robot density was calculated for the first time in the 2015 survey. Due to a very low number of employees in the manufacturing industry and the increasing number of installed robots, the robot density is very high. The robot density in Japan further decreased to 305 units, and in Germany it continued to increase to

301 units. The United States which is one of the five the biggest robot markets regarding annual supply has a robot density of 176 units in 2015. The robot density in China, the biggest robot market since 2013, reached 49 units in 2015 unveiling the huge potential for robot installations in this market.
In 2015, the average robot density in the following regions was: 92 units in Europe, 86 in the Americas and only 57 in Asia.

The considerable high rate of automation of the automotive industry compared to all other sectors is demonstrated in the evaluation of the number of industrial robots in operation per 10,000 employees in automotive industry and in all other industries. Despite its shrinking robot density, Japan had still the highest robot density in the automotive industry. 1,276 industrial robots were installed per 10,000 employees in Japan's automotive industry. It is followed by the Republic of Korea with 1,218 units, the United States also with 1,218 units and Germany with 1,147 units. The robot density in the automotive industry in China has increased significantly since 2007 but compared to countries like Korea, Germany, and the United States, it is still on a rather moderate level (392 units). The reason is the large number of employees working in the automotive industry. According to the China Statistical Yearbook about 3.5 million people worked in the automotive industry (including automotive parts) in 2014. The number increased from 3.4 million in 2013. In 2015, about 21 million cars were produced in China, the highest volume of cars produced in a country, accounting for about 30% of the global car production. The need to modernize and to further increase capacities will boost robot installations in the coming years. The robot density in the automotive industry in the United States increased only moderately between 2011 and 2015 (from 1,103 to 1,218 units), while the operational stock rose considerably in the same period. The reason is the remarkable rise in employment in the automotive industry in the same period. The employment rate in the automotive industry increased by 27% in 2015 compared to 2011.

The robot density in the general industry (all industries excluding automotive) is still comparatively low. However, countries with an important electronics industry have a higher rate. The Republic of Korea is on top with 411 robots installed per 10,000 employees. It is followed by Japan with 213 robots, Germany with 170, Taiwan with 159, and Sweden with 154 units. Germany and Sweden do not have any important production sites regarding the electronics industry. The comparatively high rate in both countries is due to a more diversified distribution of industrial robots in all industries. The increasing automation in the production of electronic devices will push robot installations within the related production hubs, particularly in Asian countries. Italy ranks sixth with a robot density of 126 units. It is followed by Austria with 95 units and the United States with 93 units. Most of the emerging robot markets have a robot density rate below 30. The United Kingdom also has a rather low rate in the general industry; only 33 robots are installed per 10,000 employees.

The overall conclusion indicates that in almost all the surveyed countries, the potential for robot installations in the general industry is still tremendous. It is also considerably high in the automotive industry among the emerging markets and in some traditional markets as well. Continued necessary modernization and retooling also guarantee further robot investments in already highly automated countries. Relocation of productions may result in declining investments in that country. However, robot investments will be shifted to the new production base in another country.

V. AI'S PERCEPTION PROBLEM: AI AND THE RISING TECHNO-PANIC

Innovative new technologies often arrive on waves of breathless marketing hype. They are frequently touted as "disruptive!", "revolutionary!", or "game-changing!" before businesses and consumers actually put them to practical use. The research and advisory firm Gartner has dubbed this phenomenon the "hype cycle." But there is a corollary to the hype cycle for new technologies that is less well understood and far more pernicious. It is the cycle of panic that occurs when privacy advocates make outsized claims about the privacy risks associated with new technologies. Those claims then filter through the news media to policymakers and the public, causing frenzies of consternation before cooler heads prevail, people come to understand and appreciate innovative new products and services, and everyone moves on. Call it the "privacy panic cycle." Many individuals and organizations jump on the bandwagon during the Rising Panic, knowing that making outrageous claims about privacy and other issues is a sure path to recognition. For example, not content with repeating the already vastly exaggerated claims by Oxford University researchers that AI and robotics will destroy 47 percent of U.S. jobs in 20 years, one Silicon Valley pundit had claimed that it will destroy 80 to 90percent of U.S. jobs in the next 10 to 15 years. And not to be outdone, Kevin Drum writes in Mother Jones that all jobs will be gone in 40 years. As a result of this sort of unquestioning hysteria, the

public is bombarded with overblown fears and a false sense of urgency. Because of the crowded field of opinion and analysis, the media tends to recognize those with the most outrageous claims, setting a pattern whereby it continuously escalates the perceived implications, challenges, and threats brought by the new technology. This has been the pattern with AI. Skeptics and antagonists have engaged in hyperbolic and emotional rhetoric that the media then repeats and amplifies. This phase of panic has been marked by apocalyptic and dystopian imagery for AI, including Elon Musk's warning that it could be "summoning the demon" that destroys the human race (Hearing on Economic and Labour Force Implications of Artificial Intelligence) [Testimony of Robert D. Atkinson President Information Technology and Innovation Foundation].

Techno panic cycles typically end at what we call the "Point of Practicality," at which apocalyptic concerns fade and people move on. At this stage, the majority of the public no longer believes the dystopian claims that antagonists make, and the technology has reached a sufficient level of maturity that most people no longer express concerns about its misuse. The technology is just part of life. And we move on to a new techno-panic cycle for the next big technological innovation (Atkinson & Wu, 2017).

A. CAUSES FOR THE AI TECHNO-PANIC

AI has been swept up in the techno-panic cycle for at least three major reasons. First, AI is what economists call a "general purpose technology" that can and likely will affect many different aspects of the economy. As such, it is easy to offer doomsday scenarios in which it could affect all occupations, all industries, and all workers. Second, AI is extremely complicated and opaque. While science fiction writer Arthur C. Clarke wrote that "Any sufficiently advanced technology is indistinguishable from magic," this is even more true with AI because it is not tangible. Even if people in the past were not mechanical engineers, they could get at least a rudimentary sense of what a lathe, truck, or assembly line could and couldn't do. But unless someone has a computer science degree, ideally with a specialization in machine learning, they have virtually no understanding of AI. As such, it can and does take on mysterious and ominous powers. As a result, when an AI dystopian suggests that we are only a few short steps away from artificial general intelligence (a computer with intelligence equivalent to human intelligence) or even artificial superintelligence (a computer with vastly superior intelligence), such that Elon Musk can call it our biggest existential threat, the vast majority of people have no common-sense way to judge the validity of his claim.

Third, AI has a perception problem because of its very name. The term "artificial intelligence" implies that the technology has or soon will have intelligence akin to human intelligence. And, ominously, that this will quickly transform into artificial super-intelligence that is beyond human control. But this is wrong. AI has very limited intelligence—it can figure out a game of GO or that a picture of a cat is not a dog, but it can't and won't be able to make the kinds of complex decisions that a three-year-old child can make. Computers don't really think, and they certainly are not conscious. While a child might yell at Apple's Siri that she is stupid, Siri isn't conscious of this. As philosophy professor John Searle wrote about IBM's Watson, "IBM invented an ingenious program—not a computer that can think. Watson did not understand the questions, nor its answers, not that some of its answers were right and some wrong, not that it was playing a game, not that it won—because it doesn't understand anything."9 Yet many AI skeptics just don't want to believe this. James Barrat, a documentarian and author who wrote the anti-AI book Artificial Intelligence and the End of the Human Era, blithely writes, "As for whether or not Watson thinks, I vote that we trust our perceptions."10 By this logic, we should believe the earth is flat. Put this all together, and it is not surprising that much of what has been written about the social and economic impacts of AI is so ludicrous. Many claims are so comical that it is surprising that people take them seriously (Atkinson & Wu, 2017).

Double-digit growth between 2016 and 2021

- Industry linking the real-life factory with virtual reality, will play an increasingly important role in global manufacturing.
- Human-robot collaboration will have a breakthrough in this period.
- Compact and easy-to-use collaborative robots will drive the market in the coming years.
- Global competition requires continued modernization of production facilities.
- Energy-efficiency and using new materials, e.g. carbon-composites, require continued retooling of production.
- Growing consumer markets require expansion of production capacities.
- Decline in products' life cycle and an increase in the variety of products require flexible automation.

- Continuous quality improvement requires sophisticated high tech robot systems.
- Robots improve the quality of work by taking over dangerous, tedious and dirty jobs that are not possible or safe for humans to perform.
- Continued strong demand from the automotive industry.
- Accelerating demand from the electrical/electronics industry.
- Increasing demand from the metal and machinery industry, the rubber and plastics industry, the food and beverage industry.
- Small and medium sized companies will increasingly use industrial robots.
- China will remain the main driver of the growth and will expand its dominance.

By 2021, almost 40% of the global supply will be installed in China.

- Continued increase of robot installations in all major Asian robot markets:
- Korea, Japan, Taiwan and other Southeast Asian markets.
- Continued growth in North America – recovering sales in the near future in Brazil.
- Accelerating growth of robot sales in Central and Eastern Europe.
- Moderate growth of robot sales in Western Europe.

Global robot installations are estimated to increase at least by 14%, to about 290,000 units in 2016. Robot supplies in the Americas will increase by 5%, in Asia/Australia by 18%, and in Europe by 8%. From 2017 to 2019, robot installations are estimated to increase by at least 13% on average per year (CAGR): 8% in the Americas and in Europe and 15% in Asia/Australia. Total global sales will reach about 413,000 units in 2019. Between 2016 and 2019, it is estimated that more than 1.4 million new industrial robots will be installed in factories around the world.

The global robotics industry is prepared for this challenge. Production capacities have been expanded and some suppliers have established or will establish robot production assemblies in the most important market, China, or in the United States. In terms of units, it is estimated that the worldwide stock of operational industrial robots will increase **from about 1,631,600 units at the end of 2015 to 2,589,000 units at the end of 2019,** representing an average annual growth rate of 12% between 2016 and 2019. In 2015, the stock increased by 12% to about 1.8 million units. The operational stock of robots in 2016 is estimated to increase by 9% in the Americas, by 16% in Asia/Australia, and by 6% in Europe. Between 2017 and 2019, the stock will increase by 9% on average per year in the Americas, by 16% in Asia/Australia, and by 6% in Europe.

A. EFFECTS OF ARTIFICIAL INTELLIGENCE

ECONOMIC IMPACT: The capability of AI to perform increasingly more tasks that were previously considered the sole preserve of humans, and its expected ongoing continuous improvement has a range of consequences.

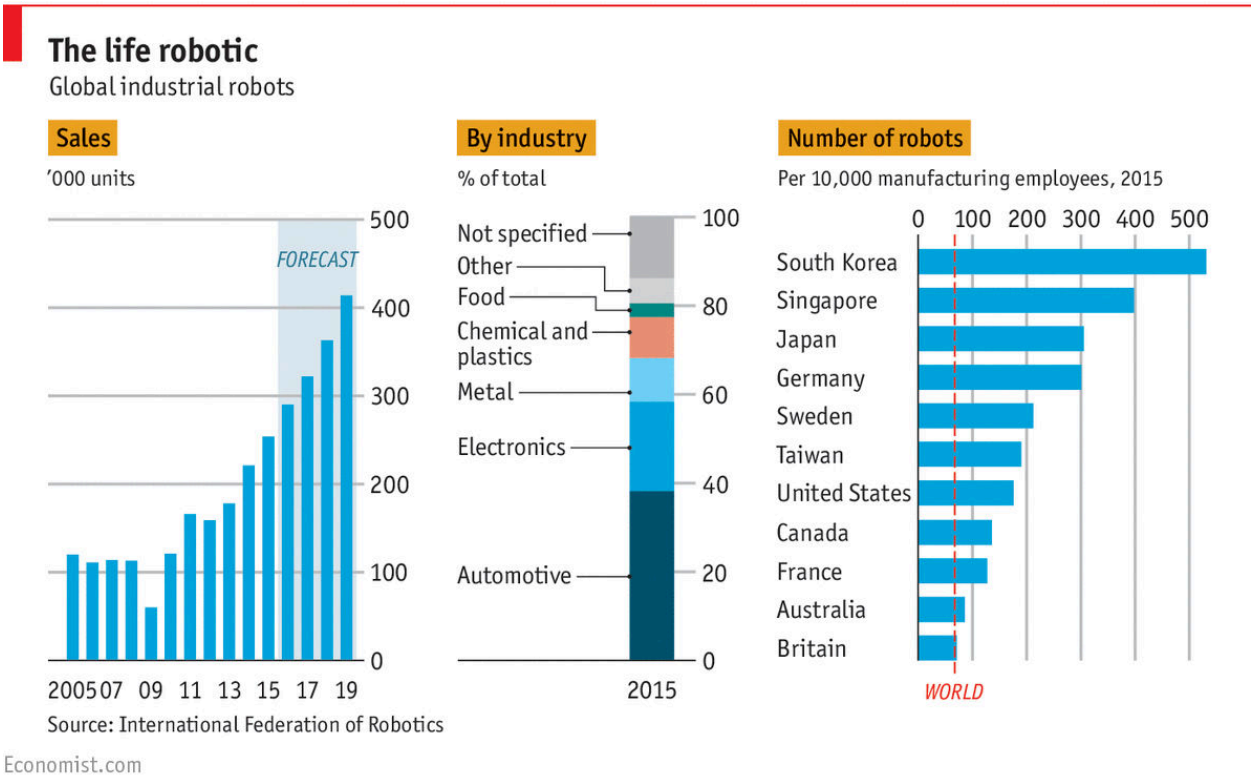
Hanson (2001) models the wage, population and economic growth consequences of what he calls 'machine intelligence'. The model is best considered as an 'ideal' or 'pure' state, abstracted from reality but helpful to show the main economic effects.

Hanson et al uses a standard economic growth model, modified to introduce computers, which complement human workers and machine intelligence, which eventually replaces them. Machines complement human labour by making them more productive at the jobs they perform, but machines also substitute human labour by taking over human jobs.

B. EVIDENCE ON THE IMPACT OF EXISTING TECHNOLOGY ON WORK

Technological change is a key driver of economic growth. However, the invention, diffusion and effective use of new technology are in turn likely to be influenced by other factors, including economic conditions, institutions, and social conditions. For example, the adoption of labour-saving technology in the first industrial revolution may have been driven by the specific economic conditions of 18th century. Moreover, sociological research has shown that patterns of work organisation vary across countries even between establishments that use very similar technologies. It is broadly accepted that the first industrial revolution eventually led to improving standards of living in society and in particular for the working class, but there is evidence to suggest that it took some time for these improvements to materialise. In the context of the first industrial revolution, technological change was also linked to changes in the nature of work: the mechanisation of textile production involved work moving from artisans' homes to the factories, from rural to urban areas, and from independent work often filling downtime in rural work to full-time, predictable work in a hierarchical structure. Recent automation does not seem to have led to overall decline in employment levels but there have been income losses

for low-educated workers employed in the manufacturing sector. Employment losses in manufacturing have typically been compensated by increasing employment in services, leading to stable or growing overall employment levels. It is not clear from the literature included in this review whether the aggregate figures conceal employment losses for specific demographics and how the new service jobs compare to the manufacturing jobs lost in terms of opportunities for progression, security, quality of working environment.



VI. RESULTS

As new technologies are adopted, the effects on consumer demand become more important. PwC notes that, "AI is still at a very early stage of development overall." PwC recognises that the productivity impacts it expects are very large: "The impact on productivity could be competitively transformative – businesses that fail to adapt and adopt could quickly find themselves undercut on turnaround times as well as costs. They stand to lose a significant amount of their market share as a result." The biggest productivity gains are in capital intensive sectors like manufacturing and transport because their operational processes are very susceptible to automation. This includes the productivity impacts of autonomous vehicles. When it comes to later years and the impacts of new product development, healthcare, the automotive industries and financial services are the sectors with the greatest economic potential. The overall income effects for humans need not be negative. In Hanson's model, computers are considerably cheaper than human labour, the world can produce the same output at much lower cost. The question becomes how to allocate resources if there is little work to do, or a question of what the displaced humans will do instead? The model is built on exogenous growth, which means that the rate of technology improvement is determined by assumptions outside the model. Hanson et al shows that an endogenous growth model (where growth builds on growth and rates are determined by equations inside the model) show much bigger impacts. At the extreme they lead to the economic equivalent of the singularity where growth rates skyrocket as intelligent machines learn faster and work more. Hanson et al does not consider any impacts of AI creating new jobs for humans. Nor does he look closely at the pace at which these effects

might take place, or exactly which jobs the computers will do.

VII. DISCUSSIONS

As many have justifiably pointed out, these are not new anxieties. Concerns that technology will adversely impact jobs and `social justice' have been around. As a result, economic theory has built up a significant scholarship dealing with the relationship between technology, growth and development, employment, inequality and productivity. In this paper, it is well appreciated that technology can possibly be a cause of unemployment and inequality. In theories of skill based technological change (SBTC) technology and capital, raise the demand for highly skilled workers, and hence their wage premium, causing increasing wage inequality (Autor et al., 2003). In curtailing unemployment, economic growth has been recognized as a key variable that can address the menace of kidnapping, insurgency, vandalization of public property. The contributions of industrialization through adoption of AI will facilitate economic growth to economic development which cannot be overemphasized as it has been recognized as one of the necessary conditions for economic development. It is a medium through which unemployment can be reduced, inflation can be stabilized, innovation can be promoted and a panacea for poverty reduction and eradication in a nation. Industrialization via AI provides the platform for entrepreneurs to emerge who on the other hand reduces unemployment rate through the creation of jobs. Lewis (1954) in a theoretical discourse which was further proved by empirical studies (Lee, 2000) predicted a declining rate of unemployment as the economy experience significant growth. Despite Lewis (1954) prediction, the Nigerian economy has remained largely underdeveloped, the huge human and natural resources has remain untapped, the per capita income is low, unemployment also is high despite growth experienced in previous years. This shows that various macroeconomic policies by government have been unable to achieve sustained reduction in unemployment and sustained growth as the economy has plunged into recession in recent time. This might not be unconnected to the failure in the macroeconomic management. The essence of macroeconomic management underlines the rationale for the existence of government as a vital economic agent. In the 1960s and early 1970s, the Nigerian economy provided jobs for most Nigerians and absorbed considerable imported labour while inflation rates were low. The wage rate compared favourably with international standards and there was relative industrial peace in most of the years. Following the oil boom of the late 1970s, there was mass migration of people, especially the youth, to the urban areas seeking for jobs (Yelwa et al., 2015). In a research environment (both in academia and research on new products/ideas etc. (in industry) one often reaches a point where existing packages will not be able to perform a required simulation task, or where more can be learned from analysing existing data in new ways etc. at that point, programming skills are required. It is also generally useful to have a broad understanding of the building blocks of software and basic ideas of software engineering as we use more and more devices that are software controlled.(Fangohr, 2015) Subsequent the downturn in the economy in the early 1980s, which caused the problems of unemployment, hasty the introduction of the Structural Adjustment Programme (SAP) in 1986. As AI continues to progress, one of the challenges we face is to ensure that its benefits are widely distributed and fairly shared. We have prepared this paper to assess the impact of AI on the world economy − by modeling how AI technologies could be adopted and absorbed by different players, and by simulating the economic disruptions that countries, companies, and individuals could experience as they transition to greater use of AI.

The introduction or the adoption of AI and Robotics, will increase productivity and churnout a productive output comparable anywhere in the world as quality control measure will be addressed since the machines are precise. Robotics are based on two enabling technologies: Telemanipulators and the ability of numerical control of machines.

Telemanipulators are remotely controlled machines which usually consist of an arm and a gripper. The movements of arm and gripper follow the instructions the human gives through his control device. First telemanipulators have been used to deal with radio-active material.

Numeric control allows to control machines very precisely in relation to a given coordinate system. It was first used in 1952 at the MIT which led to the first programming language for machines (called APT: Automatic Programmed Tools). The combination of both of these techniques led to the first programmable telemanipulator. The first industrial robot using this principle was installed in 1961. These are the robots one knows from industrial facilities like car construction plants.

The development of mobile robots was driven by the desire to automate transportation in production processes and autonomous transport systems. The former led to driver-less transport systems used on factory floors to move objects to different points in

the production process in the late seventies. New forms of mobile robots have been constructed lately, like insectoid robots with many legs modeled after, for example, what nature gave us like autonomous robots for underwater usage. For a few years, wheel-driven robots are commercially marketed and used for services like "Get and Bring" (for example in hospitals), “pick and place”, in factories.

VIII. CONCLUSION

The Industrial Revolution is one of the most celebrated watersheds in human history. It is no longer regarded as the abrupt discontinuity, but an expansion in production methods and techniques which led to greater productivity. As discussed, this led to the invention of Artificial Intelligence which can be incorporated into machines and devices that makes it carryout human functions with precision as the its name suggests. The recent successes of AI have captured the wildest imagination of both the scientific communities and the general public. Robotics and AI amplify human potentials, increase productivity and are moving from simple reasoning towards human-like cognitive abilities. Current AI technologies are used in a set area of applications, ranging from healthcare, manufacturing, transport, energy, to financial services, banking, advertising, management consulting and government agencies. The global AI market is around 260 billion USD in 2016 and it is estimated to exceed 3trillion by 2024. To understand the impact of AI, it is important to draw lessons from it's past successes and failures and this paper provided a comprehensive explanation of the evolution of AI, its current status and future directions. We believe that history, logic, and economic analysis all strongly point to the conclusion that the next technology wave, powered by artificial intelligence and robotics, will not lead to above average unemployment levels and that we will not run out of work. What it could do, however, is significantly improve labour productivity growth rates, making society better off, and boosting per-capita incomes for virtually all sectors. As such, policymakers should not give in to the rising techno-panic over AI or take steps to slow down AI progress. Rather, they should take steps to support AI, including by using AI much more extensively within government operations. Finally, while the next wave of innovation won’t create mass unemployment, it will likely increase labour market churn, making it essential that state governments do a much better job equipping workers with the support, tools and skills they need to navigate a more turbulent labour market. The key term is Industry. It describes the integration of all value-adding business divisions and of the entire value-added chain with the aid of digitalisation. In the “factory of the future”, information and communication technology (ICT) and automation technology are fully integrated. All subsystems – including non-producing ones such as R&D as well as sales partners, suppliers, original equipment manufacturers (OEMs) and customers – are networked and consolidated into one system. This makes it possible to simulate the results in full by changing the parameters. In other words: all relevant requirements concerning manufacturing and production capacity are already confirmed during product development. The entire process can be considered and managed in real time holistically from the very first step – including seamless quality assurance in production. Economic growth on the continent occurred when these innovations were adopted there. This paper brings to the fore the need to totally adopt AI as this imperative will lead to increased productivity as studies which indicated that the pace of economic growth in France was not very different from that in England despite the differences in economic structure–hence, the thesis of O’Brien and Keyder (1978) that there were “two paths to the twentieth century.” This critique has gathered force with the recent emphasis on the Scientific Revolution, a pan European phenomenon, as the cause of the Industrial revolution. With increased capabilities and sophistication of AI systems, they will be used in more diverse ranges of sectors including finance, pharmaceuticals, energy, manufacturing, education, transport and public services. The next stage of AI is the era of augmented intelligence, seamlessly linking human and machine together.

REFERENCES

[1]. Akeju, K. F. Olanipekun, D. B. (2015): Unemployment and Economic Growth in Nigeria. International Journal for African and Asian studies. Vol 11

[2]. Allen, R. C. (2006): The British Industrial Revolution in Global Perspective: How commerce created The Industrial Revolution and Modern Economic Growth.

[3]. Atkinson, R. and Wu, J. (2017). False Alarmism: Technological Disruption and the U.S. Labor Market, 1850-2015. Information Technology and Innovation Foundation, May:1-28.

[4]. Autor, D., Levy, F., and Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. Quarterly Journal of Economics, 118(4):12791333.

[5]. Braitenberg, V. (1984) *Vehicles: experiments in synthetic psychology*. The MIT Press, Cambridge Massachusetts.

[6]. Chen, Nicholas, Lau Christensen, Kevin Gallagher, Rosamond Mate, Greg Rafert (2016), Global Economic Impacts Associated with Artificial Intelligence.

[7]. Christopoulos D. (2004), "The Relationship Between Output and Unemployment: Evidence from Greek Regions Study", Regional Science, 83: 611-620, DOI: 10.1007/s10110- 004-0198

[8]. Daniel, K. and Ejara D. (2009), "Impact of information asymmetry on municipal bond yields: An empirical analysis. *Am. J. Econ. Bus. Admin., 1: 11-20.*

[9]. Era, A. N. E. W., Opportunity, O. F., For, C., & Uk, T. H. E. (n.d.). THE FUTURE OF MANUFACTURING :

[10]. Fangohr, H. (2015). Introduction to Python for Computational Science and Engineering (A beginner ' s guide).

[11]. Fu, X., Pietrobelli, C. and Soete, L. (2010). 'The Role of Foreign Technology and Indigenous Innovation in the Emerging Economies: Technological Change and Catching-up', World Development, 39 (7): 1204-1212.

[12]. Hanson (2001), Economic Growth Given Machine Intelligence.

[13]. Hayden, E., Assante, M., Conway, T. (2014): An Abbreviated History of Automation and Industrial Controls and Cybersecurity. Sans analyst white paper

[14]. Henderson, Lh. (2006). Modern robotics, building versatile machines. Chelsea House.

[15]. Holland, J.H. (1992) Adaptation in natural and artificial system. (2nd edition), The MIT Press, Cambridge, Massachusetts, (1st edition 1975).

[16]. Jones, R. (2007). AC 2007-926 : ENGINEERING CAPACITY BUILDING IN DEVELOPING Russel Jones , World Expertise LLC Engineering Capacity Building in Developing Countries.

[17]. Lloyd, P.J. (1996). 'The Role of Foreign Investment in the Success of Asian Industrialization', Journal of Asian Economics, 7 (3): 407-433.

[18]. Marsh, P. (2013). The New Industrial Revolution. Yale: Yale University Press.

[19]. O'Brien, P.K. (2001). 'Industrialization; Typologies and History of', International Encyclopedia of the Social and Behavioural Sciences, pp. 7360 – 7367.

[20]. O'Neill, J. (2001). 'Building Better Global Economic BRICs', in Global Economics Paper No: 66, ……GS Global Economics Website

[21]. Okun A. M. (1962), Potential GNP: Its Measurement and Significance, American Statistical Association, Proceedings of the Business and Economics Statistics Section, pp. 98–104.

[22]. PWC (2017), The Economic Impact of Artificial Intelligence on the UK Economy.

[23]. Summit, G., & Engineering, M. (n.d.). 2028 Vision for Mechanical Engineering. The Factory of the Future. (n.d.).

[24]. Szirmai, A. (2012a). 'Industrialization as an Engine of Growth in Developing Countries, 1950-2005', Structural Change and Economic Dynamics, , 23 (4), December 2012, pp. 406-20, http://www.sciencedirect.com/science/article/pii/S0954349X11000 18X

[25]. Veloso, F. and Soto, J.M. (2001). 'Incentives, Infrastructure and Institutions: Perspectives on Industrialization and Technical Change in Late-Developing Nations', Technological Forecasting and Social Change, 66: 87-109.

[26]. Von Tunzelmann, G.N. (1997). 'Innovation and Industrialization: A Long-Term Comparison', Technological Forecasting and Social Change, 56: 1-23.

[27]. World Bank (2018). The Changing Nature of Work. World Development Report 2019. Washington DC: The World Bank.

[28]. Yelwa, M., David, O. O. K. and Omoniyi, A. E. (2015). Analysis of the relationship between inflation, Unemployment and economic growth in Nigeria: 1987-2012. Applied Economic and Finance, 2(3): 102-09.

Improved Trajectory Planning for FES-Induced STS Movement Restoration in Paraplegia

Mohammed Ahmed
Department of Electrical and Electronics Engineering
Abubakar Tafawa Balewa University
Bauchi, Nigeria
inunugoloma@yahoo.com

M. S. Huq
Department of Mechatronic Engineering
Institute of Technology
Sligo, Ireland
saif.huq@gmail.com

B. S. K. K. Ibrahim
School of Mechanical, Aerospace and Automotive Engineering
Coventry University
Coventry, United Kingdom
Ad1465@coventry.ac.uk

Abstract*—The work presented an enhanced trajectory planning for better handling of Functional Electrical Stimulation (FES) actuated revival of sit-to-stand (STS) manoeuvre. Currently there the problem of lack of FES aided devices as results of the failures of the proposed equipment to pass clinical tests. Employing control systems techniques have been yielding fruitful results. Literature indicated that further enhancement of trajectory planning could help in achieving milestones towards better performances. Continuous and coupled with bump-free trajectories are some essential properties for better control systems designs for such systems. The proposed trajectory planning utilises the six-order polynomial. Comparisons are made with the five and seven order polynomials. The fifth order polynomials give an insight of the performance of the system with lower order polynomials while the seventh order gave that of the higher order counterparts. It was employed to enhance the desired movement trajectories. Obtained results portrayed that the trajectory planning achieved are continuous and the tendencies of the appearances of jerks or spikes are eliminated. Others are an improvement on execution time by 11%, reducing the upper and lower bound terminal velocities by 16.9% and 20.9% accordingly. Thus, these indicate the possibilities of realisations of higher performance control systems for the induced movement tasks.

Keywords*—induced sit-to-stand, functional electrical stimulation, trajectory planning, paraplegia

I. INTRODUCTION

In robotics, paths describe the motion geometrically; it maintains silence regarding the dynamics of the system. Paths give loci of points within a specified operational domain. Trajectory portrays time-dependent description of movement in robotics systems. System dynamics were as well considered when necessary, e.g. velocity and acceleration. Hence, it is a time domain representation of the path, and it is usually, planned. The trajectories are generally described by algorithms which also considered the constraints of the path together with that of the dynamics. Therefore, this resulted in trajectories that are functions time and are used for attaining the desired positions, velocities and accelerations [1-4].

Among the earliest works on trajectory planning for robotics applications using polynomials was the works of Paul and Zhong [5]. The trajectories proposed are characterised by being parabolic and joints velocities that linearly change with time. Several authors also made similar presentations [6-11]. A major shortcoming of the methods was the presence of a high number of spikes in jerks. It makes it necessary to use three different functions when dealing with the derivative of acceleration. Similar polynomials of order three were proposed [4, 6, 9-11] and also of order five [4, 6, 9, 11]. In the works of Fu et al. [12], presented a slightly different approach. The single joint motion trajectory was created using either three polynomials having orders of four, three and four or orders of three, five and three. Another option was using five polynomials of orders three. The use of two third order polynomials was proposed in the works of Craig [9]. In the works of Angeles [13], four polynomials having different orders of four, five, six and seven were used to represent the trajectory of the joint. After which improvement was made on the proposed trajectory by approximating the associated seven order polynomial using a cubic spline. These methods are complex are complex. In order to solve the issues both jerking and using multiple polynomials, Williams II [14] proposed a single four order and six order polynomials for the single joint trajectory planning. The four order polynomial trajectory presented still have shortcomings, but the six order polynomial trajectory solves the issue of unbounded spikes in the jerk. Hence, producing continuous, smooth, reliable and robotics systems with greater life longevity. These desired properties are required as well during FES-induced movements. The experimental illustration of the fact recently was as in the works of Nasir [15]. Fatigue was reduced considerably by 35.89% which was achieved by lowering or smoothening the tracking error and reducing the amount of stimulation current.

Similarly in the works of Ibrahim [16] although the result was regarding energy consumption rather than current. All the same, the energy is a function of the current, and a proportional relationship exists between the parameters. Energy consumption was lowered by 10% which improves the fatiguing time from 26s to 42s (by about 57%). In the works of Kirsh [17], was regarding the effectiveness of reducing the effect of fatigue based on how well the control system maintains the response close to the desired value (reduced tracking error). Results show that both the fatigue compensation (FC) and proportional derivative (PD) controllers maintain almost the same level of stimulation currents. The FC controller produces angular tracking errors less than half of that of the PD controller. In the works of Lynch et al. [18-23], improved fatigue, spasm, and tremor compensation were associated with lowering the control signal/stimulation current, smoothening of the stimulation current/control signal.

II. METHODOLOGY

Three angles were involved in the modelling process. Obtaining the time required/elapsed for the FES induced sit-to-stand movement was achieved by taking the average acquired experimentally earlier works. These works include that of Nuzik et al. [24], Kamnik et al. [25], Fattah et al. [26] and Dolan et al. [27]. The duration of four (4) seconds emanated as the time elapsed for the movement. The polynomial was used to approximate the trajectory of the movement as used by Fattah et al. [26] and Dolan et al. [27]. Six order polynomial proposed by Williams II [14] was selected for the study and was used to approximate the trajectories due to its relative advantages as given by equation (1). Its benefits as mentioned earlier include generating movements with relative continuity, smoothness, and reliability, and robotics systems having better life longevity. The polynomial parameters are established and evaluated by considering the entire STS dynamics during selection even though not clearly shown.

$$\theta(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5 + a_6 t^6 \quad (1)$$

INITIAL AND FINAL VALUES OF ANGLES/POTITIONS

Angle/Position	Initial Value ((rad/°) / m)	Final Value ((rad/°) / m)
Ankle Joint Angle θ_A	1.8326/105	1.5708/90
Knee Joint Angle θ_K	4.8526/278	6.2857/360
Thigh Joint Angle θ_T	0.0000/0	0.0000/0
Thigh Joint Position d	0.6601	0.9430

TABLE V.

Table I shows the various angles as well as their initial and final values. The initial t_i, middle t_m and final values t_f of times are zero (0), two (2) and four (4) seconds respectively. The initial velocity, final velocities, initial and final accelerations are all zero (0). The various trajectories of the system are as given by equations (3)-(14). The parameters a_1 and a_3 are all zero (0). Obtaining the various trajectories was achieved using these mentioned conditions as given by equation (2) for the ankle, knee and thigh joints respectively.

$$\begin{bmatrix} 1 & t_m & t_m^2 & t_m^3 \\ 1 & t_f & t_f^2 & t_f^3 \\ 3 & 4t_f & 5t_f^2 & 6t_f^3 \\ 6 & 12t_f & 20t_f^2 & 30t_f^3 \end{bmatrix} \begin{bmatrix} a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} \left(\frac{\theta_m - \theta_i}{t_m^3}\right) \\ \left(\frac{\theta_f - \theta_i}{t_f^3}\right) \\ 0 \\ 0 \end{bmatrix} \quad (2)$$

$$\theta_A(t) = 105 - 2.34375t^3 + 0.87891t^4 - 0.08789t^5 \quad (3)$$
$$\dot{\theta}_A(t) = -7.03125t^2 + 3.51564t^3 - 0.43945t^4 \quad (4)$$
$$\ddot{\theta}_A(t) = -14.0625t + 10.54692t^2 - 1.7578t^3 \quad (5)$$
$$\theta_K(t) = 278 + 12.8125t^3 - 4.80469t^4 + 0.48047t^5 \quad (6)$$
$$\dot{\theta}_K(t) = 38.4375t^2 - 19.21876t^3 + 2.40235t^4 \quad (7)$$
$$\ddot{\theta}_K(t) = 68.875t - 57.6562t^2 + 9.6094t^3 \quad (8)$$
$$\theta_T(t) = 0 \quad (9)$$
$$\dot{\theta}_T(t) = 0 \quad (10)$$
$$\ddot{\theta}_T(t) = 0 \quad (11)$$
$$d(t) = 0.6601 + 0.04420t^2 - 0.01658t^4 + 0.00166t^5 \quad (12)$$
$$\dot{d}(t) = 0.1326t^2 - 0.06632t^3 + 0.0083t^4 \quad (13)$$
$$\ddot{d}(t) = 0.2652t - 0.1989t^2 + 0.0332t^3 \quad (14)$$

Equations (3), (4), and (5) are descriptions of the desired angular postion, angular velocity and angular acceleration trajectories of the ankle joint during the sit-to-stand movement. Equations (6), (7), and (8) are representations of the desired trajectories for the angle, angular velocity and acceleration trajectories of the knee joint during the movement respectively. Equations (9), (10), and (11) are depictions of the desired angular, angular velocity and angular acceleration trajectories of the hip joint during the sit-to-stand transition. Equations (12), (13), and (14) are descriptions of the desired vertical position, vertical velocity and vertical acceleration trajectories of the hip joint during the sit-to-stand movement. The developed desired trajectories resulted after careful observation on earlier works and some of which were experiment based. Hence, the resulting trajectories indirectly harnessed real data which are expected to be close as possible to that of exact/natural human biomechanics.

III. RESULTS AND DISCUSSION

Fig. 1 – Fig. 4 show the graphical comparisons of the angular positions, angular velocities, angular accelerations, and derivatives of angular accelerations trajectories. The legend in Fig. 1 also applies to Fig. 2, Fig. 3 and Fig. 4.

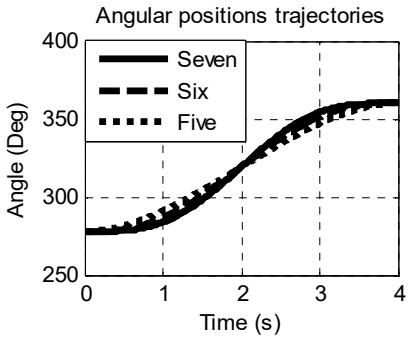


Fig. 1 Angular Positions Trajectories Comparison

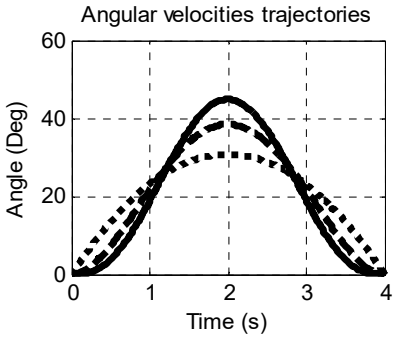


Fig. 2 Angular Positions Velocities Comparison

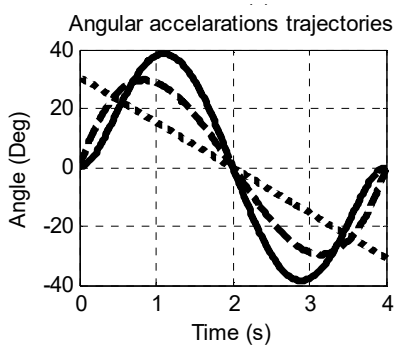


Fig. 3 Angular Accelerations Trajectories Comparison

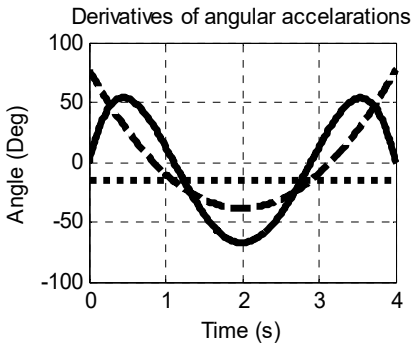


Fig. 4 Derivatives of Angular Accelerations Trajectories Comparison

Considered for the analysis was a single joint, and evaluated based on three parameters; continuousness, velocities at terminal points, and cost or computational time of the trajectories. Hence, utilised are the trajectories associated with the knee joints. Beginning with the first item, portrayed from the various angular acceleration trajectory plot of the five order polynomial is the likelihood of occurrence of jerking or spikes. Observed as well are discontinuities on the derivative of the angular acceleration. These shortcomings made the five order polynomial trajectory at a disadvantage. The determined computational times for five, six and seven order polynomials are 0.022617 s, 0.024018 s, and 0.026512 s respectively. The six and seven order polynomial trajectories, execution times are 6.2% and 17.2% higher than that of the five order polynomials. Hence, six order polynomial has 11% faster execution time. The terminal velocities at the beginning for the five, six, and seven order polynomial trajectories are determined as 23.6^0/s, 19.6^0/s, and 19.0^0/s respectively. And at the end of the transitions are 29.2^0/s, 23.1^0/s, and 16.7^0/s respectively. These indicate improvements of 16.9% and 20.9% at the starting velocity with the six and seven order trajectories respectively over the five order. And 19.5% and 42.8% at the stopping speeds respectively. These reasons made the six-order polynomial desirable for the study.

IV. Conclusion

The successfully conducted appraisal on five, six, and seven order polynomials confirm some of the claims by Williams II [14]. Utilised for comparison are the five order because it is the closest to the six, and seven order because of its application in an earlier study by Fattah et al. [26]. The five order trajectory has the characteristics of discontinuity and likelihood of jerking occurrence. Also, the proposed trajectory indicates an execution time which was 11% faster than that of the seven order polynomial trajectory. It also meant improvements over the five order polynomial trajectory regarding the starting and stopping velocities as 16.9% and 20.9% respectively. Applying the method for the trajectory planning resulted in outcomes were close to that of previous works as mentioned in the discussions. The work is novel in the sense no previous work had proposed six order polynomial trajectory by Williams II for the FES induced sit-to-stand movement. And again incorporating an experimental finding that shows that upper limbs support some percentage of the overall weight of the subject. It also gives elaborations or clear stages to accomplish the tasks which are lacking in previous studies.

Acknowledgement

Profound appreciation to the Faculty of Engineering and Engineering Technology, Abubakar Tafawa Balewa University Bauchi, University Tun Hussein Onn Malaysia and Tertiary Education Trust Fund Abuja Nigeria for their support.

References

[1] L. Biagiotti and C. Melchiorri, *Trajectory planning for automatic machines and robots*: Springer Science & Business Media, 2008.

[2] A. Gasparetto, P. Boscariol, A. Lanzutti, and R. Vidoni, "Trajectory planning in robotics," *Mathematics in Computer Science,* vol. 6, pp. 269-279, 2012.

[3] M. Zhihong, *Robotics*. Singapore: PEARSON, Prentice Hall, 2005.

[4] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: modelling, planning and control*: Springer Science & Business Media, 2010.

[5] R. P. Paul and H. Zhang, "Robot motion trajectory specification and generation," in *2nd International Symposium on Robotics Research, Kyoto, Japan*, 1984.

[6] A. Koivo, *Fundamentals for control of robotic manipulators*: John Wiley & Sons, Inc., 1989.

[7] R. E. Parkin, *Applied robotic analysis*: Prentice-Hall, Inc., 1991.

[8] L. Sciavicco and B. Siciliano, *Modelling and control of robot manipulators*: Springer Science & Business Media, 2012.

[9] J. J. Craig, *Introduction to robotics: mechanics and control* vol. 3: Pearson/Prentice Hall Upper Saddle River, NJ, USA:, 2005.

[10] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control* vol. 3: Wiley New York, 2006.

[11] R. N. Jazar, *Theory of applied robotics: kinematics, dynamics, and control*: Springer Science & Business Media, 2010.

[12] K. S. F. R. G. C. S. G. Lee, *Robotics: Control Sensing. Vis*: McGraw-Hill, 1987.

[13] J. Angeles, *Fundamentals of Robotic Mechanical Systems: Theory, Methods, and Algorithms*, 4 ed.: Springer International Publishing, 2014.

[14] R. L. Williams II, "Simplified Robotics Joint-Space Trajectory Generation with a via Point using a Single Polynomial " *Journal of Robotics,* 2013.

[15] N. B. M. Nasir, "Intelligent Modelling and Control with Fatigue Reduction for FES Induced Knee Joint of Hemiplegic for Rehabilitation," Doctor of Philosophy, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 2017.

[16] B. S. K. K. Ibrahim, "Modelling and control of paraplegic's knee joint (FES-swinging)," PhD, University of Sheffield, 2011.

[17] N. A. Kirsch, "Control methods for compensation and inhibition of muscle fatigue in neuroprosthetic devices," University of Pittsburgh, 2016.

[18] C. Lynch and M. Popovic, "Sliding mode control of FES-induced quiet standing," in *Proceedings of the 17th Annual Conference of the International Functional Electrical Stimulation Society (IFESS), Banff, Canada*, 2012.

[19] C. L. Lynch, "Closed-Loop Control of Electrically Stimulated Skeletal Muscle Contractions," PhD, University of Toronto, 2011.

[20] C. L. Lynch, G. M. Graham, and M. R. Popovic, "Including nonideal behavior in simulations of functional electrical stimulation applications," *Artificial organs,* vol. 35, pp. 267-269, 2011.

[21] C. L. Lynch, G. M. Graham, and M. R. Popovic, "A generic model of real-world non-ideal behaviour of FES-induced muscle contractions: simulation tool," *Journal of neural engineering,* vol. 8, p. 046034, 2011.

[22] C. L. Lynch and M. R. Popovic, "A comparison of closed-loop control algorithms for regulating electrically stimulated knee movements in individuals with spinal cord injury," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 20, pp. 539-548, 2012.

[23] C. L. Lynch, D. Sayenko, and M. R. Popovic, "Co-contraction of antagonist muscles during knee extension against gravity: Insights for functional electrical stimulation control design," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 1843-1846.

[24] S. Nuzik, R. Lamb, A. VanSant, and S. Hirt, "Sit-to-stand movement pattern," *Physical Therapy,* vol. 66, pp. 1708-1713, 1986.

[25] R. Kamnik, J. Q. Shi, R. Murray-Smith, and T. Bajd, "Nonlinear modeling of FES-supported standing-up in paraplegia for selection of feedback sensors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 13, pp. 40-52, 2005.

[26] A. Fattah, M. Hajiaghamemar, and A. Mokhtarian, "Design of a Semi-Active Semi-Passive Assistive Device for Sit-to-Stand Tasks," presented at the 16 th Annual (International) Conference on Mechanical Engineering-ISME2008, May 14-16, 2008, Shahid Bahonar University of Kerman, Iran, 2008.

[27] M. J. Dolan, B. J. Andrews, and P. Veltink, "Switching curve controller for FES-assisted standing up and sitting down," *IEEE Transactions on Rehabilitation Engineering,* vol. 6, pp. 167-171, 1998.

Detection and Sentiment Analysis of Hate Speech on Twitter in Nigerian Politics

Aliyu Garba
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
garba.aliyu@aun.edu.ng

Sandip Rakshit
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
sandip.rakshit@aun.edu.ng

Fatima Rilwan
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
fatima.rilwan@aun.edu.ng

***Abstract*— Politics in Nigeria is becoming an unusual thing due to the heterogeneity of different political factions, races, and faiths. This leads to growing a consequential menace in securing unity in the country. Social media is the real channel where information is effectively scattered to different users across the world within a short period. In this way, politicians during campaigns post some abhor and dangerous hearsays to tarnish or denounce race or rival (running mate) without considering what in the long run will be the results. Consequently, causes an unconstrained crisis which relies upon its dimension (politics, religion, and tribe). This research tends to analyze social media posts to detect hate speech and decide its impact on individuals by examining the polarity. Various classification algorithms including Random Forest, XGBoost, Logistics Regression, Gradient Boosting, and Multi-Layer Perceptron were used to train the model. However, Random Forest is found to outperforms all the other algorithms with harmonic mean of the precision and recall (F-measure) of 90%. The polarity of tweets, is found to be 53% are for those who are neutral, 26% positive and 21% negative**

Keywords*— *Politics, Hate speech, Social Media, Polarity, Analysis

I. INTRODUCTION

Social networks have been the significant root of producing unstructured data since its inception. This turns into the trends of creating real-time big data in connection to Customers Business Relationship (CBR), National and International Issues, Politics, among others. Besides, social media sites have become a principal incentive of hate speech [1]. Thus, analysts in the scholarly world and business organizations tend to spotlight on big data analytics through social media to take research on sentiments [2]. Hate speech is a statement that points to a precise group of individuals in an instigated way by snubbing their ethnic, opinion, believes as well as religion [3] [4]. Hate speech in Nigeria leads to a dangerous threat to social security where partisans use it to influence people to carry out an illegal act in the circumstances whereby they could not achieve or succeed in what they are aiming at. The vice president of Nigeria related hate speech as "hate speech precedes genocide, the greatest tragedy in human history"[5]. There is a lack of perception of hate speech in the countryside, and there is no constitutional law to back it up when infringed. There could be extreme charges on freedom of speech by over-enthusiastic security agencies If it allows being interpreted. To guarantee harmony, we cannot bear to be quiet regarding incitement upon all groups [6]. Nigeria needs to come up with some policy to diminish such public abuse. Consequently, security forces get it challenging to regulate the flow of information that steers to deviate from the law of the land. Unlike countries, France, Canada, and the United Kingdom where the law to regulate hate speech are enforced, and anybody exposed in the violation might be imprisoned or to pay some fines [7].

In this paper, we proposed a technique for identifying and analysing hate speech and its impact. Sentiment analysis is investigated to ascertain the polarities on the hate speech and how emotionally and enthusiastically citizens are in sharing. However, this research is constrained to just hate speech that is identified with partisan politics in Nigeria, especially the 2019 general election.

This paper is arranged in the following order: Section II intensively study the related works, Section III describes the approach used in detecting and analyzing hate speech and the system model, Section IV discusses the results, lastly, Section V, conclusion and future work.

II. RELATED WORKS

The use of online social media is not curbed to only a particular gender and or a specific age boundary to possess full access to circumvent hateful speech. Twitter and Facebook give attention to hate speech, and they remarkably devise means of regulating public insult [6]. The consequence of hate speech for which depends on the dimension, it varies across countries, states, and geopolitical zones. The people of a particular country or state might not have diverse ethnic groups but rather may have comparative cultures. This is not the fact in Nigeria with various ethnic groups and religious faiths, and with several cultures. The consequence of hate speech concerning politics might not be identical to some different countries. Nevertheless, because of this heterogeneity, our study precisely concentrates on the detection of hate speech and sentimental analysis that is associated with politics in Nigeria.

[7] carried out a study on automatic detection of hate speech specifically on Twitter. This is concentrated on women of the Turkish country. Supervised machine learning technique was employed viewing various classification models including Naïve Bayes (NB), J48, Random Tree (RT), Support Vector Machine (SVM), and Random Forest (RF). The efficiency of the result reveals that NB and linear SVM implied significantly more trustworthy than the other models. Furthermore, the tweets signified often recognized to be about sexual and homophobic slurs.

[3] identified hate speech using deep learning architecture. Firstly, the authors researched how deep learning techniques could be employed in the detection of hate speech. The twitted data is categorized as sexist, racist, or neither. Be that as it may, because of the multifaceted nature of language constructs, a far-reaching experiment with several deep learning architectures (Convolutional Neural Networks, FastText, and Long Short-Term Memory Networks) was used. Besides, many classifiers such as Logistic Regression, Random Forest, SVMs, Gradient Boosted Decision Trees (GBDTs) and Deep Neural Networks (DNNs) were used to get the most favorable result. The study proved that while deep neural networks and gradient boosted decision trees are joined, it follows in the best accuracy and more trustworthy than the present techniques in hate speech detection.

[4] carried and presented preliminary research on the machine learning technique. Several classification models were applied and weighed. The results revealed that Random Forest Decision Tree outperformed Naïve Bayes, Support Vector Machine and Bayesian Logistic Regression. The futures included religion, hatred, race, ethnicity, and gender. However, politics has not been examined in their study and how hate speech influences people were not investigated.

According to the research by [2], This concentrated on building the Artificial Intelligence (AI) model that performs the capacity of detecting emotions that are observed on unstructured texts. The work tried to learn the opinions of several users based on retweets and likes of tweets and the use of natural language processing (NLP) toolkit to analyse the sentiment of texts. The tweets that are liked most and the most retweeted tweets that emotionally influence users are analysed. This is related to our study concerning the influence of tweets to the users. However, our study is continued to detect hate speech before making a sentiment analysis.

[8] carried research that classifies hate speech in social media. This work employed classification techniques and used the dataset to build lexical baselines for their task. The system augments the existing dataset with emotional information by applying the Natural Language Processing (NLP) techniques and passes to machine learning classification. The findings showed 80.56% accuracy, which gives a 100% increase while compared to the original study adopted by [9].

III. METHODS

In this section, we explain how to secure the dataset and the process to use in taking the study using a machine learning technique, where numerous classification models are applied and compared to achieve the most commendatory result.

A. Data Collection

Data is the part and parcel of every data mining. Tremendous data produced every day by social media can be used to make or infer meaningful decision out of it. Data mining is the process of using data to drive mining. To obtain this data, REST APIs is made available for academic research [10], which is required as an initial stage. for this research, we applied through Tweeter and permission was granted alongside the API keys. While the election campaigns were ongoing in Nigeria, politicians are posting offensive and contentious statements or comments which regarded as hate speech that may emerge as a crisis. This research centered on detecting hate speech, implementing sentiment analysis, and determining the weight of its impact (influence) on the users. We extracted the data using the hashtag (#NigeriaDecides2019). The hashtag was used to group all related posts in regards to February 2019 general election.

In our research, the Tweepy API was used. The Tweepy API is a Python library that can give access to posts (extract, search, post stream) from Twitter API[11]. Because of the large size of data from the posts, we aimed to use NLP. Python provides nltk library for processing textual data. nltk provides access to its algorithms and allows us to process textual data known as NLP. It is quite famous and helpful as it gives string tokenization, sentiment analysis, tagging part of speech, n-grams, correction of spelling, extraction of a noun phrase, etc.[2].

B. System Model

The system model in this study is formed based on major components. A program written in Python will communicate with the Twitter platform by API to retrieve all data associated 2019 general election. The extracted data will be stored in a repository. The program is now load the data from repository into memory for pre-processing. The cleaned data is immediately transferred to a machine learning classification model. This is in charge of learning from the data and predict whether the statement tweeted by a specific user is offensive (Hate speech or offensive speech). If and only is the statement is offensive and marked as hate speech, the sentiment analysis will be taken out to ascertain its impacts on people and how it is virally shared and make a positive or negative or neutral remark. Figure 1 below shows the System Model.

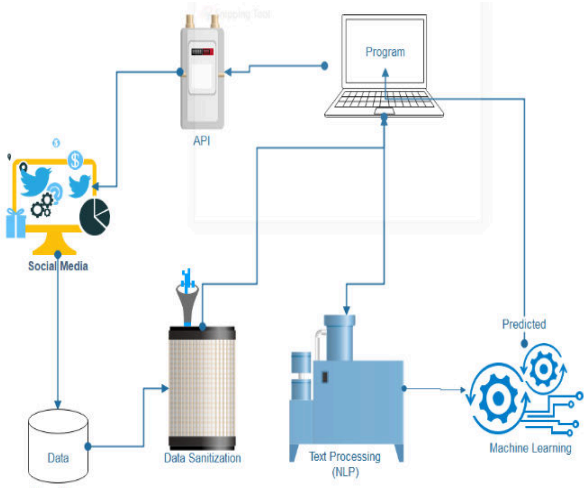


Figure 12: System Model

IV. Results Discussion

The detection of hate speech from various classification models, the result showed people using the social media are actually posting statements that are aggressive, tagged as hate speech or offensive (not as bad as hate speech that can cause much trouble). However, there are people who are just neutral, that are not posting aggressive statement at all. At all classification thresholds, the performance classification model is plotted based on two parameters, True Positive Rate (TPR) known as recall against True Negative Rate (TNR) as shown in the figure 2 below. Depressing the classification threshold rates more items as positive, consequently rising both False Positives and True Positives. The XGBoost receiver operating characteristic (ROC) curve tends to show a better recall of 87% than Base Model and Logistics Regression. Considering the harmonic average of the precision and recall (f-score or f-measure) as performance metric, Random Forest is found to outperformed all other classification algorithm with 90% shown in figure 3 below.

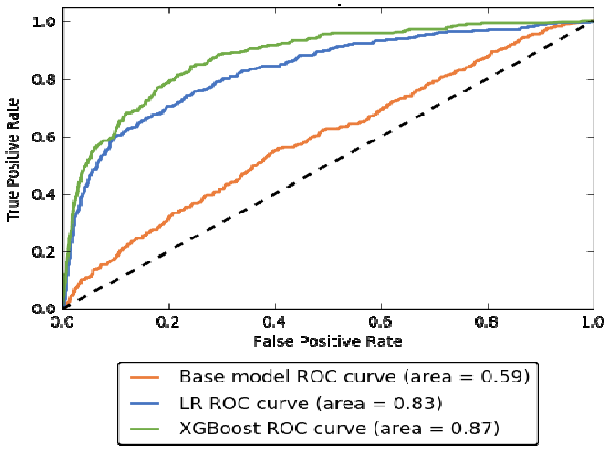


Figure 13: ROC for Hate Speech Detection

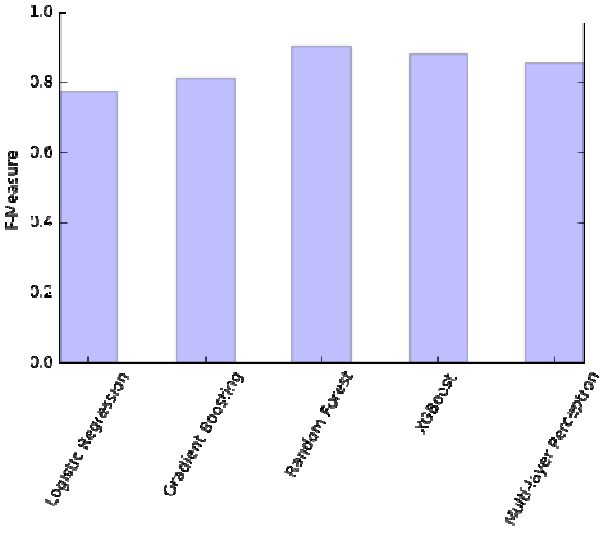


Figure 14: F-Measure for Five Classification Models

When a hate speech is detected, sentiment analysis is carried out to know the influence of it among users. This is to gather people's opinion about a particular post. After the analysis, the result showed that 53% are for those who are neither making a positive or negative comment (neutral), 26% made positive and 21% negative comments as shown in figure 4 below. This shows how hate speech has a negative impact on human even though a high number of people seemed to be neutral. Figure 5 shows the distribution of the length of text people tweeted for detected hate speech. The highest length of tweeted text is 144 characters, minimum is 94 while the average is 133 characters.

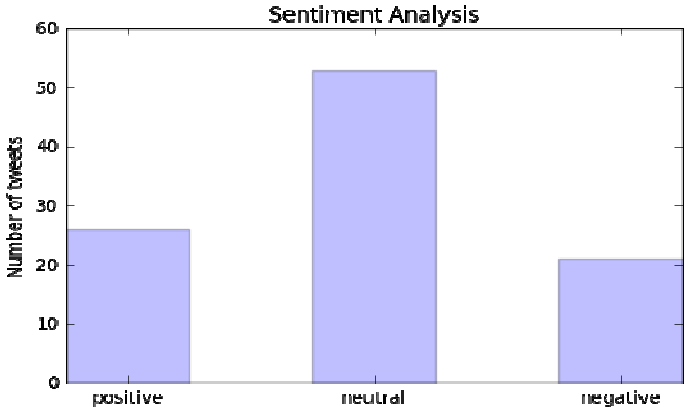


Figure 15: Sentiment Analysis

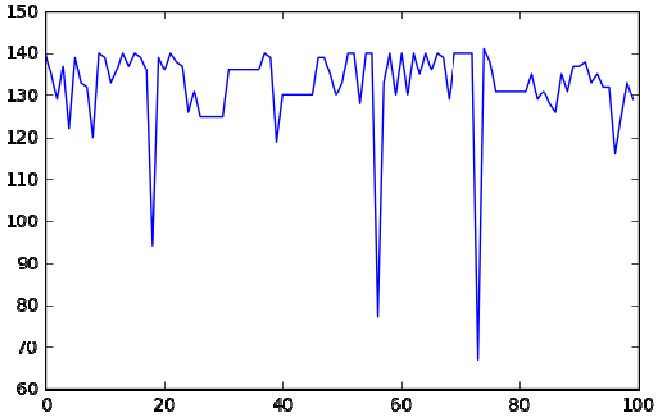


Figure 16: Distribution of Length of Text

V. Conclusion

The findings of this study will help the authority to know the degree to which hate speech is spread over social media and presumably provide a rule of law to control. This work centered on February 2019 general election. The data is obtained from Twitter using a hashtag #NigeriaDecides2019.

We used machine learning classification models, Random Forest, XGBoost, Logistics Regression, Gradient Boosting, and Multi-Layer Perceptron. the results of the models showed that Random Forest outperformed better than the other algorithms.

In the future, this work can be enhanced by concentrating on major native languages in the nation, as this research concentrated on English. Again, in this research, data collected contains a mixture of somewhat we called in Nigeria, pigeon English. This insufficiency to be analysed to attain a more trustworthy result.

REFERENCES

M. Mondal, L. A. Silva, and F. Benevenuto, "A Measurement Study of Hate Speech in Social Media," in *Proceedings of the 28th ACM Conference on Hypertext and Social Media - HT '17*, Prague, Czech Republic, 2017, pp. 85–94

Y.-P. Huang, N. Hlongwane, and L.-J. Kao, "Using Sentiment Analysis to Determine Users' Likes on Twitter," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, Athens, 2018, pp. 1068–1073

P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep Learning for Hate Speech Detection in Tweets," in *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, Perth, Australia, 2017, pp. 759–760.

I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekanata, "Hate speech detection in the Indonesian language: A dataset and preliminary study," in *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Bali, 2017, pp. 233–238.

A. Dan, "Hate Speech - Vanguard News Nigeria." [Online]. Available: https://www.vanguardngr.com/2018/12/hate-speech/. [Accessed: 11-Mar-2019].

Published, "But, what exactly is hate speech?," *Punch Newspapers*..

H. Sahi, Y. Kilic, and R. B. Saglam, "Automated Detection of Hate Speech towards Woman on Twitter," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, 2018, pp. 533–536.

R. Martins, M. Gomes, J. J. Almeida, P. Novais, and P. Henriques, "Hate Speech Classification in Social Media Using Emotional Analysis," in *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, Sao Paulo, 2018, pp. 61–66.

T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," p. 4.

"Twitter Developer Platform — Twitter Developers." [Online]. Available: https://developer.twitter.com/. [Accessed: 13-Mar-2019].

"Getting started — tweepy 3.5.0 documentation." [Online]. Available: https://tweepy.readthedocs.io/en/v3.5.0/getting_started.html#introduction. [Accessed: 14-Mar-2019].

A linguistic Approach to Relational Databases Using Fuzzy Logic

R. Salahudeen
SITC
American University of Nigeria.
Yola.
ridwan.salahudeen@aun.edu.ng

G. Aliyu
SITC
American University of Nigeria.
Yola.
garba.aliyu@aun.edu.ng

A. F. D. Kana
Department of Computer Science
Ahmadu Bello Univeristy.
Zaria.
donfackkana@gmail.com

***Abstract* - Database system is an elegant way of storing knowledge about a domain of interest in a real-world domain. It is however undoubtedly clear that real-world knowledge is mostly inherently imprecise due to the ambiguity prominence in human natural languages. This problem made many relational databases to only capable of returning records that completely satisfy it query predicate in a Boolean fashion, however, users sometimes not only need the records that completely satisfied a query but also need records in the close boundary of their query for good decision making. This paper seeks to address the imprecision of relational databases using the concept of fuzzy logic.**

Keyword - Relational database system, SQL, Fuzzy logic

I. INTRODUCTION

A relational database is a logical organization of data in a collection known as table or sometimes called relation. A relation consists of two dimensions, a column dimension and a row dimension. Each record in a database table has an atomic value corresponding to a particular attribute in the column dimension [1]. Relationship are created through the data linkages among multiple table by enforcing key and reference constraints, this is in contrary to earlier generations of databases like the Network Model and Graph Databases, that uses some algebraic mechanisms to maintain relationship among collective data structures through the usage of links and pointers.

Structured Query Language (SQL) is an effective programming language used to manipulate and/or query out information stored in a relational database, sometime, to derive data meaning and/or relationship among other data stored in the same database, conjunctions of connected tables are done using the enforced primary-foreign columns of the corresponding tables.

Aside data query and manipulation, SQL is also used in defining the storage structure of a database. SQL is an information manager that is based on relational calculus. It belong to the declarative family of formal languages that are focused on descriptive structure of what to be done and abstract out the details of how records are retrieved or manipulated [2].

Table 3: Company Project Table [1]

Project Id	Budget (×00)	Duration	StaffLvl
P01	55	108	29
P02	48	115	33
P03	57	90	10
P04	81	87	40
P05	78	92	16
P06	63	102	25
P07	91	125	41
P08	90	88	27
P09	65	117	37
P10	73	102	30
P11	93	93	45
P12	60	110	31
P13	59	121	37
P14	89	92	18
P15	51	101	23

When a selection query is made in SQL, all record satisfying the where clause is filtered out and when there is nowhere clause associated to the selection, then all the records in the selected table(s) are returned. At other times, the where clause can also be used for the joining of multiple table that has inherent primary-foreign key relationship. The mechanism of a where statement requires some set of predicate(s) that must be satisfied by the table being queried, then a set of rows in the table that satisfy the predicate would be returned. Literally, a where clause can take as much predicates as possible to refine its filtering to a more specific user desire, so when the predicates' complexity grows, the efficiency of record retrieval becomes more challenging and sometimes leads to a total loss of precision.

II. CLASSICAL SQL LIMITATIONS

Information in a real-world domain are highly infused with uncertainties and most decision to be made based on the information derived from the data storage requires some level of precision in order to make reliable and accurate decisions. In the contrary, queries in databases are intended for the retrieval of decisive information to aid decision-making, but with the advent of vagueness in the concepts of the real world that produce these information, conclusive decision are always challenging [1].

This challenge of classical SQL in determining the precision of vague concepts and its lack of accuracy is been described as follow:

Table 1 illustrated a database table containing a company's various projects along with their financial budget and time duration. The columns of the table are: the project ids, budget in hundreds, duration and the staff assigned to each project.

It could be vividly seen that retrieving projects with say, high budget that required short duration to be ran will require some definitive value of "high budget" and "short duration" when dealing with the classical SQL. Now, let assume the choose of $80 and above is what is meant by high budget and a time duration of 90 or below is what is meant as short duration, then, the SQL query is:

```
SELECT * FROM Project_table WHERE duration
                < 90 AND Budget > 80;
```

A resultant table of this query can be seen in Table 2 below.

Table 4: Resultant Table

ID	Budget	Duration	StaffLvl
P04	81	87	40
P08	90	88	27

Based on the intension of the query, the only two records returned by the SQL is not very efficient, since there still good project that might serve the intension of high budget and short duration that were supposed to be presented to a human decision maker for further consideration. A graphical illustration of the company's project table with the selection boundary of the query with the high budget and short duration is shown in Figure 1.

It could be easily seen from Fig. 1 that some projects are highly close to the are close to the area of intension are could be potential candidate for the intended objective but they were obviously not in Table 2. The effect of this would be even greater as the database records increases or when the number

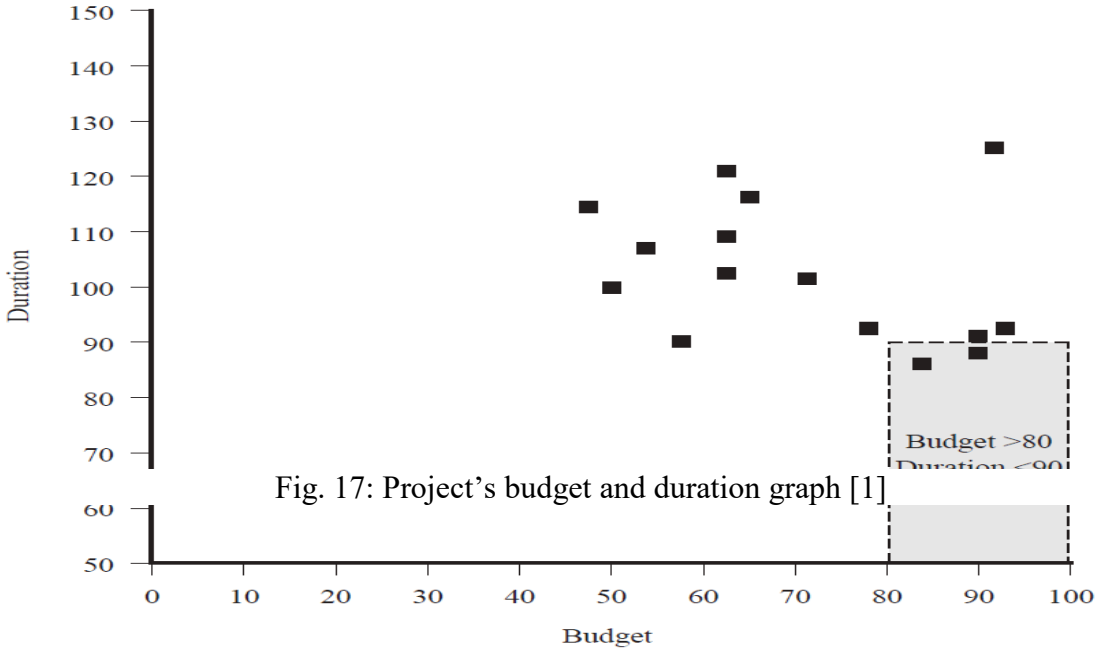


Fig. 17: Project's budget and duration graph [1]

of predicates for the where clause filtering increases.

One of the common practise by database users to solve the imprecision is by adjusting the predicates to accommodate the close boundaries shown in Figure 1. For instance, the following SQL showed a relaxed version of the previous SQL:

SELECT * FROM Project_table WHERE duration < 100 AND Budget > 75;

Approach like this, is effective in realtively small tables. Unfortunately, it introduces additional challenges of differentiating the original intended filters from the extended related records, and in most cases, as the query get to be refined iteratively, the original query meaning might be lost entirely. Then, a person reading the refined query much later can hardly what the original idea conceived was. The more the number of iterations for refining the original query the less accurate the query becomes.

This paper is therefore proposing an alternative approach in addressing the challenge of imprecision in classical SQL by using fuzzy logic to quantify and rank the degree of closeness of RDBMS records from its queries.

j.

III. LITERATURE REVIEW

[3] extensively reviewed the models of the fuzzy database and highlighted the need for modelling complex object with imprecision and uncertainty in fuzzy relational and object-oriented databases.

[4] explored the limitations of data selection mechanism in SQL when linguistic expression and degrees of truth to an investigation is involved.

[5] developed a model to enable users to a linguistic oriented queries in a relational database. They were able to prove that the limitation of classical SQLs on dealing with imprecision can be addressed by extending the SQL and not alteration need to been done on the relational database management systems.

[6] demonstrated some situations where Boolean logic is not suitable for querying relational databases. They used the query builder tool of MATLAB and GEFRED model to demonstrate the results.

k.

VI. THE PROPOSED APPROACH

The imprecision limitation of classical SQL can be addressed by fuzzy logic. Fuzzy logic was introduced by Zadeh in 1965 [7] and has been in used to extend various data models since 1980s to date. A Fuzzy set is effective in mapping elements of a particular set to a certain degree of belongingness to a vague concept. For instance, let

Z be a set of individuals, say, Z = {"Hanna", "Mercy", "John"}. Then a concept of Tallness of these individuals can best be measured in degree as oppose the sharp boundary of classifying the individuals as either tall or not tall. Hence, a fuzzy set of Tall over the set of individuals "Z" could be {.69, .78, .97}, which can be simply read as Hanna, Mercy and John are tall to the degrees of .69, .78 and .97 respectively.

Information retrieval based on wide ranged linguistic concepts rather than sharp boundary crisp values are made possible by an approach known as Fuzzy. This approach tends to be more effective and accurate when dealing with natural languages' concepts in a real world domain, especially when the concept(s) involved are infused with vagueness and imprecision. This is a necessary solution to the classical SQL as the forceful coercion of natural language concepts with inherent vagueness has an implication of losing confidence in knowledge discovery and engineering.

To address this, a linguistic predicate can be adopted in SQL queries rather than using crisp values. For instance, the query for the company's database to filter high budget and short duration project could be conceptually specified as follow:

SELECT * FROM Project_table WHERE duration is Short AND Budget is High;

Some terms in natural language are obviously continuous in nature, so terms like High and Short have their boundaries very wide and also imprecise in nature. Such terms are more effectively described in degrees of closeness to the concept being examined in other to make an efficient decision.

Fuzzy logic can be used to model such imprecise terms into vague concepts, such that when such concepts appears in SQL queries, the SQL filtering can retrieve only records satisfying that term and presents it to a user that make decision on it in a descending degree of truthiness to the objective.

Here, an algorithm for the proposed approach is given as thus:

Step 1: Sort the affected table in descending order of the column(s) in the where clause.

Step 2: Create a new column for the membership values of each of the column(s) in the where clause.

Step 3: Compare (divide) each record's value for the column(s) mentioned in the where clause of the query against the highest value of that same column in the table.

Step 4: The comparison value should be entered in the corresponding column created in step 2.

To achieve these, a membership function is used to map the fuzzy concept to each record in the specified table. The resultant selections would be ranked based on the value of each record from the membership function that consequently represent the semantic meaning of the concept(s) used in the query statement.

For instance, a selection based on the concept of high could be specified as follow:

SELECT * FROM Project_table WHERE Budget is High;

On comparing the value of each record's membership with the conceptualized term of High, a ranked record of projects would be returned, where the record with the highest budget tops the list then follow by the next runner up, the list follows an ascending order of budget value until the project with the least budget value.

As shown in Table 3, P11 has the highest Budget of 93, so the comparison value would be 93/93 which is equals to 1, then P07 Budget value is 91, hence, the comparison value would be 91/93 = 0.98. This procedure will be continued on all the records of the table as shown thus:

Table 5: Fuzzy selection based on high budget

ID	Budget	Duration	$\mu(HighBudget)$
P11	93	93	1
P07	91	125	0.98
P08	90	88	0.97
P14	89	92	0.96
P04	81	87	0.87
P05	78	92	0.84
P10	73	102	0.78
P09	65	117	0.7
P06	63	102	0.68
P12	60	110	0.65
P13	59	121	0.63
P03	57	90	0.61
P01	55	108	0.59
P15	51	101	0.55
P02	48	115	0.52

This linguistic query approach has more desiring benefits when an SQL query with multiple predicates is used with the conjunction and disjunction operators of fuzzy logic on multiple fuzzy sets. The records satisfying the characteristics of the fuzzy concept(s) based on the properties of the con(dis)junction will be efficiently filtered out with their relative degree of closeness.

V. CONCLUSION

SQL is a database language used to manipulate data in a relational database management system (RDBMS), the classical SQLs are only effective when crisp values are used in their WHERE clause' predicates. However, data stored in relational databases are meant to be highly interrelated in order to achieve am elegant information representation and consequently making intelligent decision out what is been stored. Since these data

stored in these RDBMS are often form natural languages, then there are always the possibility of vagueness in some of the concept represented, as such, creating a challenge of imprecision when classical SQL are used in manipulating such concepts.

This challenge of classical SQL is therefore necessitating an investigation on how vagueness can be accommodated in our database languages. As such, this paper intends to investigate this challenge of classical SQL, then explore and demonstrate the potentials of a linguistic querying approach when fuzzy set and logic are employed in our traditional SQLs.

m.

REFERENCES

[1] C. Earl, Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration, Elsevier, 2005.

[2] W. Hai, X. Zeshui and P. Witold, "An overview on the role of fuzzy set techniques in big data processing: Trends, chanllenges and opportunities," *Knowledge-Based Systems,* vol. 118, pp. 15-30, 2017.

[3] Z. M. Ma and L. Yan, "A Literature Overview of Fuzzy Database Models," *JOURNAL OF INFORMATION SCIENCE AND ENGINEERING,* pp. 189-202, 2008.

[4] M. Hudec, "Fuzzy Improvement of the SQL," *Yugoslav Journal of Operations Research,* vol. 21, no. 2, pp. 239-251, 2011.

[5] M. Hudec, "An approach to fuzzy database querying, analysis and realization," *Computer Science and Information Systems,* vol. 6, no. 2, pp. 127-140, 2009.

[6] Poonam, "Fuzzy to SQL Conversion using Gefred Model with the help of MATLAB," *International Journal of Computer Applications,* vol. 104, no. 17, pp. 32-37, 2014.

[7] P. K. Maji, R. Biswas and A. R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Math,* vol. 9, no. 3, pp. 589-602, 2001.

[8] L. A. Zadeh, "Fuzzy Sets," *Information and Control,* vol. 8, no. 3, pp. 338-353, 1965.

Implementation of a Data Transmission System using Li-Fi Technology

Ifada Emmanuel E.
Department of Electrical and Electronics Engineering
University of Ilorin
Ilorin, Nigeria.
emmanuelifada7@gmail.com

Surajudeen-Bakinde N.T.
Department of Electrical and Electronics Engineering
University of Ilorin
Ilorin, Nigeria.
deenmat1211@gmail.com

Faruk Nasir
Department of Telecommunication Science
University of Ilorin
Ilorin, Nigeria.
faruk.n@unilorin.edu.ng

Abubakar A.
Department of Electrical and Electronics Engineering
Ahamadu Bello University
Zaria, Nigeria.
abkzarewa@gmail.com

Mohammed Olatunji O.
Department of Electrical and Electronics Engineering
University of Ilorin
Ilorin, Nigeria.
reacholaabdul@gmail.com

Otuoze A. O.
Department of Electrical and Electronics Engineering
University of Ilorin
Ilorin, Nigeria.
otuoze.ao@gmail.com

Abstract - Over the years, the overdependence on Wireless Fidelity (Wi-Fi) for data transmission necessitated the need for an alternate and more reliable means of communication, hence, Light Fidelity (Li-Fi). It involves the use of Light Emitting Diode to transmit data by blinking (i.e. switching them On and Off) at a speed not noticeable to the eye. This paper proposed the development of the Li-Fi system using off the shelf electronic components. The proposed system utilizes an embedded system with dual-core Advanced Virtual RISC (AVR) microcontroller (ATmega16L) interfaced to input/output circuits comprising of the Light Emitting Diode (LED), LM358N Operational Amplifier and a photodiode. Also, by developing a user (Receiver PC) interface using JAVA programming, the sample data (text) transferred was monitored and the speed, efficiency, security and capacity of the system was examined and discovered to be top-notch. This would make the system an indispensable means of communication in the nearest future. This data transmission system is different from those in existence because expensive components were not in the design, invariably reducing the overall cost of the implementation.

Keywords - Light Fidelity (Li-Fi), Wireless Fidelity (Wi-Fi) Optical communication, Telecommunication

I. INTRODUCTION

Light Fidelity (Li-Fi) is a fast, remote correspondence utilizing visible light. It falls under the classification of optical remote communications. Information transmission happens through Light Emitting Diode (LED) bulbs whose intensity changes (varies). During this variation in light intensity, communication takes places digitally. This innovation has huge applications where the utilization of Wi-Fi is restricted or prohibited. It likewise takes out the unfavorable wellbeing impacts of utilizing electromagnetic waves. Except light is seen, information can't be hacked; thus, data transmission is secure. The utilization of light as a way to transmit information has been authored Li-Fi. The high-speed innovation is like Wi-Fi but is quicker, enabling you to send and get more data in less time. By swapping glowing bulbs with LEDs - which have electronic properties - Li-Fi could bring Internet access to more regions and could reform the media communications industry.

As of late, wireless innovation has blossomed, as it were such that, wireless technology requires huge amount of data to transmit each day. These days, wireless communication have turned out to be significant in the communication process. The fundamental methods for transmitting wireless information is by utilizing electromagnetic waves for example radio waves. In any case, radio waves can bolster less data transfer capacity (bandwidth) as a result of conservative spectrum accessibility and intrusion. The solution to this is data transmission using Visible Light Communication (VLC). Wi-Fi deals with wireless coverage within premises, whereas Li-Fi is perfect for high compactness wireless data coverage in a defined area and for mitigating radio interference issues. Li-Fi basically focuses on transmitting multimedia data between two terminals using LEDs [2].

Two fundamental functionalities: illumination and wireless data transmission are combined by fitting a small microchip to every potential illumination device. This relationship between these functionalities could solve the four essential problems namely, capacity, cost, efficiency, and security, that we face in wireless communication nowadays.

In this work, the issues with unpredictability, various contrary standards, and the subsequent costs in these current systems are tended to by providing a simplified Li-Fi data transmission system using off-the-shelf electronic components.

II. REALATED WORKS

Harald Haas coined the term "Li-Fi" at his 2011 TED (Technology, Entertainment, and Design) Global Talk where he introduced the idea of "Wireless data from every light bulb" [3]. Through an organization with the French lighting maker Lucibel, pureLiFi manufactured a Li-Fi framework for places of business. Incorporated into the Li-

Fi framework are luminaires that hold the LED bulbs just as the balancing and demodulating circuits and computerized signal processors that run the correspondence protocols as firmware. LED bulbs installed in the ceilings can be networked into the company's IT architecture, so people using a computer or mobile device outfitted with the LiFi-X dongle can access data from office LEDs anywhere in the building [3].

PureLiFi introduced its first LiFi-X dongle in February 2016. About the size and width of a business card, LiFi-X plugs into a computer's USB port [4]. An optical device that accepts signals from the LED bulb, the dongle holds a receiver that converts the light-intensity variations of the LED into an electric signal, which is then converted back into a data stream that is transferred to a computer [5]. The dongle likewise contains a computerized data modulator that works with an infrared LED to give full-duplex bidirectional remote access at 43 Mb download and transfer speeds.

In October 2011, companies and industry groups formed the Li-Fi Consortium, to promote high-speed optical wireless systems and to overcome the limited amount of radio-based wireless spectrum available by exploiting a completely different part of the electromagnetic spectrum [1] while a number of companies that offered uni-directional VLC products, which is not the same as Li-Fi - a term defined by the IEEE 802.15.7r1 standardization committee were reviewed in [2]. Recently, the first VLC smartphone prototype was presented at the Consumer Electronics Show in Las Vegas from January in 2014 [6]. In 2013, Lain's work on the Li-Fi bulbs produced showed in an exhibition in China that the speed of data transfer was as high as 150Mbs [7].

However fascinating the use of these implementations in [3,4,6,7], the cost of procurement dissuades readily available consumers. Consequently, a Li-Fi data transmission system has been implemented using off-the-shelf electronic components.

III. MATERIALS AND METHODOLOGY

A. Background

The design of the Li-Fi data transmission system to be controlled was divided into two sections; the hardware section and software section as shown in Fig. 1.

In the system, the Arduino microcontroller creates the interface that was used to encode and then decode the data transmitted via the blinking LED (connected in the transmitter's circuit) and received by the photodiode in the receiver's circuit). The transmitter gets a signal from the PC utilizing USB Cable then from input-output connectors (pins) on the Arduino.

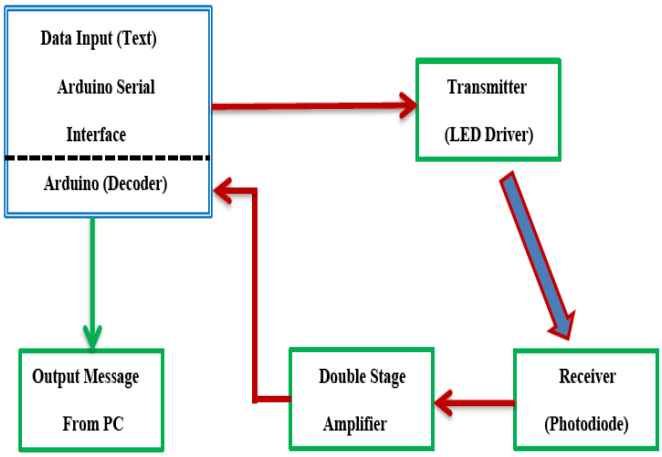


Fig. 1. Li-Fi Data Transmission System block diagram

This sign controls the transistor which functions as a switch to the power supply of the LED. The receiver converts the incoming light (described by the thick arrow in Fig. 1) into an electrical sign utilizing the photodiode. The operational amplifier - which channels and intensifies the signal - then feeds the microcontroller. The signal is in an analog form after amplification; hence, it undergoes an ADC operation, before giving it to the Arduino.

B. Hardware Section

The hardware section consists of the various parts as shown in Fig. 2:

1. microcontroller (Arduino). It also serves as the Power Supply Unit (PSU) for the Li-Fi Trans receiver System,
2. Switching Transistor (2N222),
3. Dual Operational Amplifier (LM358N),
4. Light Emitting Diode,
5. Photodiode,
6. arious resistors and a potentiometer (10 kΩ),
7. umpers and
8. Some connectors.

The circuit diagram of the Li-Fi system is divided into two major parts: the transmitter part placed above while the receiver part is placed below. The circuit diagram shows in details how the hardware components of the system are connected for data transmission. However, during the implementation of the system, a single microcontroller (Arduino board) is used.

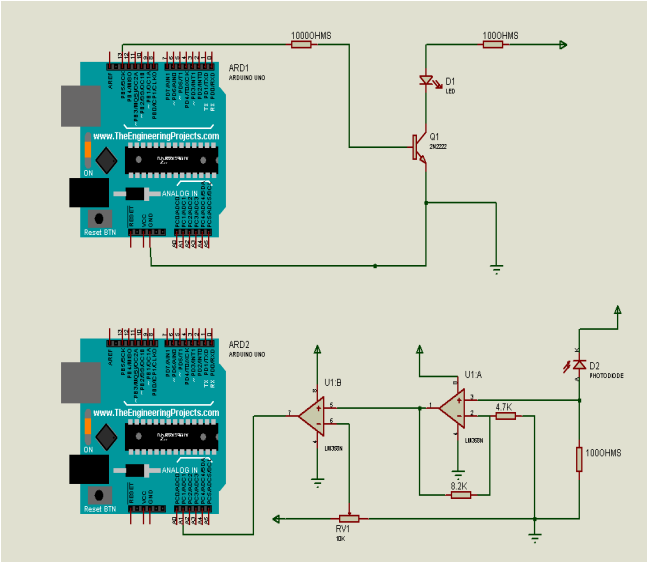


Fig. 2. Circuit diagram of the System designed using Proetus Design Tool

C. *Arduino Uno Microcontroller board*

In the Li-Fi transmission system, the Arduino board acts as both the transmitter pc (using its serial monitor interface) and also the power supply for the whole system. This configuration helps in preventing the use of an external power supply unit and also two microcontroller boards in the transmitter and receiver circuit for the system. The ATmega328P has 32 Kilobytes of flash memory for storing the program code, 2 Kilobytes of static random-access memory (SRAM) and 1 Kilobyte, and an electronically erasable and programmable read-only memory (EEPROM). These features and other facilities enable the Arduino board to interconnect with a computer, another microcontroller as another communication device. The ATmega328P pins configuration is shown in Fig. 3.

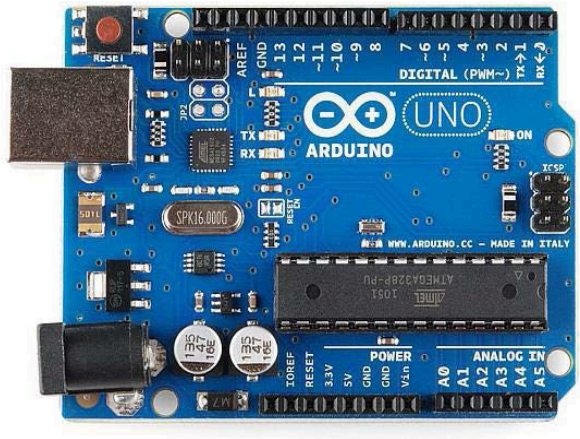


Fig. 3 Arduino Microcontroller board [8]

D. *Switching Transistor (2N222)*

In the proposed design, it is expected that: the switching transistor in the transmitter circuit would operate under hard saturation i.e. in the saturation region under all conditions. Hence, a base resistance that produces a current gain of 10 i.e. (β_{DC} = 10) is selected [9]. This is because there is more than enough base current to saturate the transistor.

Selecting a base resistance,

R_B = 1 kΩ and taking V_{BE} = 0.7 V, from the base loop,

$$V_{BB} = I_B R_B + V_{BE} \quad (1)$$

$$I_B = (V_{BB} - V_{BE}) / R_B$$
$$= (5 - 0.7)\text{ V} / 1\text{ k}\Omega$$
$$= 0.43\text{ mA}$$

Thus,

$$I_C = \beta_{DC} I_B \quad (2)$$
$$= 10 \times 0.43\text{ mA}$$
$$= 4.3\text{ mA}$$

To get the collector resistor: From the collector loop,

$$V_{CE} = V_{CC} - V_{LED} - I_C R_C \quad (3)$$

where V_{CE} = 0.3 V (from the datasheet) and $V_{LED} = 0.7\,V$,
The collector resistor is calculated as:

$$R_C = \frac{(V_{CC} - V_{CE} - V_{LED})}{I_C} \quad (4)$$
$$= (5 - 0.3 - 0.7)\text{ V} / 4.3\text{ mA}$$
$$= 93.023\ \Omega$$
$$\approx 100\ \Omega \text{ (closest standard resistor value)}$$

E. *Dual Operational Amplifier (LM358N)*

To duly implement the Li-Fi Transmission System, certain characteristics guided the choice of the operational amplifier in the receiver circuit. The following were the unique characteristics and features of the LM358N that made it most suitable for the receiver circuit.

i. n the direct mode, the common-mode input voltage range incorporates ground and the voltage at the output can likewise swing to ground, despite the fact that it is operated from just a solitary power supply voltage (of 5V from the Arduino).

ii. liminates the need for dual supplies thereby making the Li-Fi receiver circuit simpler and cost-effective.

iii. A current drain of about 500 µA during low supply which is basically autonomous of the voltage supplied.

iv. Most importantly, it has a large dc voltage gain (of around 100 dB), large output voltage swing (which is very pivotal because of the inconsistently low signal expected to be generated by the photodiode), and extremely low input offset current and voltage (2mV) in relation to other Operational Amplifiers [9].

The first stage of the receiver circuit is set up as a non-inverting DC gain amplifier to amplify the weak signal from

the photodiode. The voltage gain of the non- inverting amplifier having its output at Pin 1 is defined as:

$$A_V = 1 + R_2/R_1 \qquad (5)$$

Since from the datasheet, the output signal from the photodiode is about 1.4V maximum and it is desired to have an output voltage that would be easily detected as a 1 signal when bit '1' is sent, the amplifier is designed to have a DC voltage gain of 3.

Setting the feedback resistor, R_2, to be 8.2 kΩ; using equation 5, R_1 is obtained to be :

$$R_1 = R_2 / (A_V - 1)$$
$$= 8.2\ \text{k}\Omega / (3 - 1)$$
$$= 4.1\ \text{k}\Omega$$

$R_1 \approx 4.7\ \text{k}\Omega$ (next closest standard resistor value)

With the configuration above, the maximum output voltage to be read a bit '1' is 3.84 V and a bit '0' voltage would still be far below - at around 0.6 V after amplification. This ensures the data sent from the LED to the photodiode is decoded properly with high precision and the message is deciphered accurately.

The second stage of the receiver circuit is the gain regulation stage. The second amplifier having input Pins 5 & 6 and output Pin 7 is desired to have a unity gain. At the feedback circuit, an automatic gain controller is implemented using a variable resistor of value 10 kΩ. This variable gain setup helps in boosting the amplified signal when necessary i.e under improper lighting conditions and ensure that the photodiode signal from the transmitter's LED is decoded correctly.

F. Software Section

The Arduino, compared to other families of microcontrollers is a very easy microcontroller to use. One of the reasons is due to the ease of programming the microcontroller [8]. It involves using the Arduino Integrated Development Environment (IDE) downloaded at Arduino.cc/en/Main/Software for free. Codes written for Arduino are very similar to C sometimes called Arduino-C. Programs to be uploaded to the board are called "sketches" and each sketch is divided, mainly into three parts:

1. Library and variable declarations
2. The setup function and
3. The loop function

G. Hardware Design

The Transmitter

The conversion digital data into visible light is the basic task if the transmitter. In view of a LED's moderately linear relationship between its intensity and current, it becomes a suitable component. Modulation of the LED's intensity is the general idea i.e., the symbol transmitted correlates with the light intensity. The Arduino ports are not equipped for conveying the appropriate measure of current to make the light intensity solid and quick enough. To get around this issue, a transistor is utilized as a switch, which made it conceivable to switch a bigger current quicker. In Fig. 2, a schematic is appeared to give the transmitter's overview.

The Receiver

The task of the receiver is to convert the approaching light into current utilizing a photodiode. Since for a signal that is digital, the Arduino cannot receive a voltage that is above 5 V; thus, the electrical circuit between the photodiode and the Arduino needs to process the electrical signal so it tends to be deciphered correctly. The receiver's circuit needs to change over the current to voltage so as to enhance(amplify) and then compare it. The transmitter and receiver's separation can be changed, yet so as to avoid the problem of having too small or excessively high signal, an automatic gain regulator in form of a variable resistor is utilized here. This component enhances or decreases the digital input voltage signal to a desired digital output voltage signal. The LM358 Operational Amplifier comparator was used because it has little or no off-set [9]. The schematic of the circuit diagram as shown in Fig. 2 was tested on a Breadboard along with the transmitter before being transferred to a Vero Board.

IV. METHODOLOGY

A. Arduino Setup

The Arduino board features a serial communication interface (which includes a Universal Serial Bus (USB) controller on most models) which was used for loading programs from a personal computer to the board. The Arduino IDE was used to program the Arduino. The Arduino via the code was made to perform various tasks like switching on or off the LED on the transmitter's circuit and as well reading the analog signal read from the photodiode in the receiver's circuit.

In order not to make the system less bogus and redundant, the Arduino board was used as the micro-controller as well as the Power Supply Unit (PSU) for the circuits. It supplies a voltage of about 5 V from its V_{CC} port when connected to the PC via the serial connector cable and also acts as the virtual ground [8].

B. Hardware Implementation

The Li-Fi data transmission system is set up such that a single Micro Controller - Arduino board is used to encode, decode and supply power to the transmitter and receiver circuits.

After the design and calculations of the resistor values in the transmitter and receiver, all the components of the circuit were implemented on a breadboard carefully with the aid of jumper wires and a multimeter was used to test the continuity of the circuit from point to point.

1) The Hardware for the transmitter was designed and implemented utilizing the accompanying steps:
2) The transmitter circuit design was drawn after design calculations were completed. The software used for drawing the circuit schematic layout is the Proteus 8 Professional Design Tool.
3) After designing the circuit layout, the components were acquired and then connected on the breadboard using jumper wires where necessary.
4) Next, the circuit after a series of tests on the Breadboard is transferred into a Vero board of appropriate size carefully.

Then finally, the components were soldered appropriately and then retested.

In other to ensure the proper testing of the system and circumvent its proneness to errors, the transmitter and receiver circuits were implemented on a single breadboard and the LED and photodiode were in close proximity to one another. After several tests, the components on the breadboard were then soldered carefully using the same circuit configuration on separate Vero boards. A multimeter is used after soldering to detect dry joints, bridging of joints and test for continuity of lines on the Vero board. This helps to reduce the strain of constructing the circuit and detecting causes of error before using the circuit.

C. Software Implementation

The only implementation of software design carried out is the creation of the application that decodes the sent message from the transmitter and displays it on its output window. The application is a Java application created using Net beans Programming interface. The Arduino library was imported to the Net Beans IDE as this is an external library specially designed to ensure the Java application can communicate with the Arduino microcontroller effectively.

The Graphical User Interface (GUI) developed from the program has a connect button and a system-generated port list so as to give the user flexibility in the selection of ports on the Receiver PC. The application is opened through its link stored in the Net Beans IDE used to create it. The application prompts the user to connect the Arduino microcontroller to one of the ports from the drop-down list of ports. The port to be chosen can be determined by carefully viewing the port number from the Arduino IDE used to upload the microcontroller encoding program onto the Arduino board. The port to be connected to is automatically chosen. Any error in connecting to the right port would impede the user from viewing the message sent from the first computer.

After successfully connecting to a port, a message is displayed to acknowledge that a port has been chosen correctly. The 'Connect' button now becomes locked and changes to 'Disconnect' to prevent the user from changing the port when a message is being received and thus disrupting the decoding process.

D. Implemetation Procedure

The practical execution of the work comprises of the following parts:

1) Data packaging and encoding: This is done by the Arduino connected to the transmitter. It converts the text to bits and sends sequentially as a Low or High voltage signal.
2) Hardware control;
3) Transmission synchronization;
4) Transmission decoding: This is done by the photodiode which converts the light signal to an analog signal of varying magnitude and then passes it to the amplifier for amplification and differentiation of the bits.
5) Error handling: This is performed by the automatic gain controller which is immediately after the photodiode. It ensures that a "1" is clearly differentiated from a "0" bit.

At whatever point a user needs to send information (data) to another user, the following steps will occur:

1) The Arduino program is uploaded onto it chip via the serial port and the port number is noted.
2) After successfully uploading the program, the Arduino serial monitor is opened and the text to be sent is typed in the text box of the Arduino serial monitor indicated in Fig. 4 (which was used to avoid making the process tedious). The sent text was "Li-Fi System".
3) The system module will buffer it and begin transmitting it through the LED which acts as the input to the channel.
4) The Arduino connected to the receiving PC will receive the data transferred over the LED to the channel i.e. the photodiode and the double stage amplifier, which in turn decodes it.

The received text is then displayed using the Java GUI application designed as shown in Fig. 4.

V. RESULT AND DISCUSSION

The final design verification was done using the complete system; transmitter, receiver, and software, and the results are shown in Fig. 4.

A. Transmitter

The transmitter gets a signal from the PC utilizing USB Cable then from Arduino's GPIO connecting pins. The transistor which works as a switch to the LED's power supply is regulated by this signal. A Test data sent – a text ("Li-Fi System") encoding involved determining the characters of

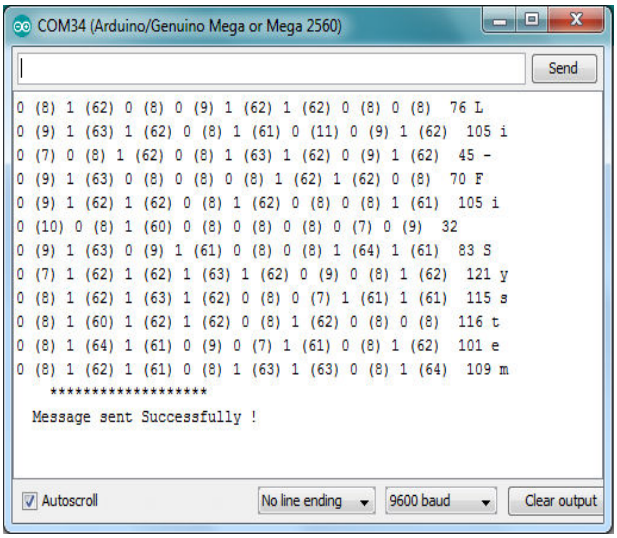


Fig. 4. Encoding and Decoding Monitor

the text individually in ASCII codes and then converting them to binary bits individually before being transmitted as light signals. Here it tends to be noticed that both 5 V and 12 V supply can be utilized in the transmitter for powering the LED yet for straightforwardness and less complexity of the circuits, 5 V supply is favoured for low power operation.

B. Receiver

The incoming signal (light) is converted to a signal (in electrical form) by the receiver using the photodiode and feeds it into the microcontroller via the operational amplifier - which filters and amplifies the signal. ADC operation is carried out on the signal before giving it to the Arduino because it is in analog form after amplification. Fig. 4 shows the measured analog signal in parenthesis at the output of the amplifier when a bit '1' or '0' is received. The generated current by the photodiode is exceptionally low; hence, it is converted to voltage by a high-value resistor. Subsequently, this voltage is additionally enhanced for the comparator to give appropriately transmitted bits. Finally, on the receiver PC, the message (text) decoded is collected and displayed as shown in Fig. 4.

VI. CONCLUSION

The aim of this work was to design a Li-Fi Data Transmission system to send data (limited to text – strings) from a PC to another PC using off-the-shelf electronic components. The Data Transmission system constructed when tested showed satisfactory performances. The Li-Fi data transmission system constructed was very cheap making it satisfy the major aim of the project – incorporation of a Li-Fi medium using off the shelf electronic components and devices.

Consequently, a Li-Fi model has been created which exhibits the essential standard and furthermore underpins the upsides of Li-Fi over Wi-Fi. The system built has a few restrictions moreover. Multi-User access is not bolstered, by the Li-Fi model structured. Additionally, this model isn't bidirectional. Thus, it is utilized for communication purposes only. Higher-end devices utilization could evacuate the impediments of this work. Furthermore, a specially made chip - Integrated Circuit (IC) could be manufactured for the encoding circuit to encode data as the microcontroller and also at the receiver to decode data sent in bits and interpret it correctly without the aid of the whole Arduino board which makes it look more bogus and less portable. This special chip can be connected to the computer system just like a modem and would get its power supply from the USB port of the computer. This set up would greatly improve the portability of the device as well as prevent its proneness to error. The major improvement on this system should focus on the process of transferring data to and from a PC either in half-duplex mode or even simultaneously.

VII. REFERENCES

[1] B. Ekta and R. Kaur, "Light Fidelity (LI-FI) - A Comprehensive Study", *International Journal of Computer Science and Mobile Computing,* Vol.3, Issue 4, pg. 475- 481, ISSN 2320 – 088X, April-2014.

[2] R. Mahendran. "Integrated LiFi (Light Fidelity) for smart communication through illumination", 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2016.

[3] H. Haas, "Wireless Data from Every Light Bulb.", *Internet: http://bit.ly/tedvlc,* Aug. 2011 [Dec 30, 2017]

[4] D. Tsonev, S. Videv and H. Haas, "Light Fidelity (Li-Fi): towards all-optical networking", Proc. *SPIE 9007, Broadband Access Communication Technologies VIII conference*, December 18, 2013.

[5] S. Dimitrov and H. Haas, "Principles of LED Light Communications: Towards Networked Li-Fi", *Cambridge, United Kingdom: Cambridge University Press*, 2015.

[6] M. Watts, "Meet Li-Fi, the LED-based alternative to household Wi-Fi," *Wired Magazine*.

[7] T. Lain, "Forget Wi-Fi, boffins get 150Mbps Li-Fi connection from a lightbulb The Register", *Internet: http://www.theregister.co.uk/2013/10/18/forget_wifi_chinese_boffins_get_150mbps_lificonnection_from_a_lightbulb*, 2013. [Dec 30, 2017].

[8] Badamasi, Y.A., "The Working Principle of an Arduino." *11th International Conference on Electronics, Computer and Computation (ICECCO),* Abuja, 29 September-1 October 2014.

[9] A. Malvino, D. J. Bates, "Electronic Principles", *7th Edition (2007)*, ISBN 0-7-063424-6, New Delhi, India.

ANALYSIS OF A GSM NETWORK QUALITY OF SERVICE USING CALL DROP RATE AND CALL SETUP SUCCESS RATE AS PERFORMANCE INDICATORS

H. A. Abdulkareem
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
ha2zx@yahoo.com

A. M. S. Tekanyi
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
amtekanyi@abu.edu.ng

A. Y. Kassim
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
aykassim@abu.edu.ng

Abstract— Quality of Service (QoS) is an important Key Performance Indicator (KPI) that is used in determining the efficiency of an industry in terms of availability and quality of services rendered. In the telecommunications industry, accessibility, retainability, and connection (voice) qualities are the three major factors used in evaluating QoS of an operator. These affect business and research activities, security, and most importantly, they also affect the daily activities of a common person. For consumers outside the industry, it is expected that the maximum satisfaction should be derived from any services paid for. Over the decades, the increase in the call drop rates and poor network availabilities within the localities, or even while moving from one location to another had become a concern to every network user and researchers. Hence, this paper assesses the QoS of MTN GSM network in four geographical areas of Kaduna State (Kaduna south, Kaduna North, Zaria and Kafanchan), Nigeria. The data collated from the management center of MTN network was used for the evaluation of the measured KPI parameters using the data management tool. The result of this research paper was compared with those specified by the Nigerian Communications Commission (NCC) and were found to be better.

Keywords— Key performance indicators, Percentage Call Drop, Call Setup Success Rate, QoS, Network provider, Users

I. INTRODUCTION

Telecommunication is the exchange of information between two or more entities and it involves the use of network technology that makes use of channels either through signal cables or in the form of electromagnetic waves. One of the major means of telecommunication is with the use of mobile phones. GSM has boosted the economic activities in Nigeria and improved the quality of life of Nigerians [1], [5]. They are now able to enjoy services such as mobile television, affordable internet service, cheaper international calls, and even internet banking. As the number of services and subscribers of GSM in Nigeria increase, the demand for good QoS has become an issue. In finding the lasting solution to this problem, the NCC, a body responsible for the regulation of GSM in Nigeria, on 6th July, 2007 issued out the threshold levels on the KPIs to ensure QoS for all the GSM networks in the country. The KPIs on which the GSM networks were tested include:

i. Call Setup Success Rate (CSSR).
ii. Percentage Call Drop (PDROP) or Call Drop Rate (CDR).
iii. Call Setup Failure Rate (CSFR).
iv. Handover Failure Rate (HOFR).
v. Handover Success Rate (HOSR).

For the purpose of this paper only the first two (CSSR and CDR) were considered in the study. Once a radiotelephone network is designed and operational, its performance should be monitored to improve overall service quality [6], [11]**.** That is where network performance monitoring for QoS assessment, analysis of faults, and corrective actions come in. In addition to monitoring network faults, the operator needs immediate information on how the network performs, especially from the end user perspective. The end user network performance report can be used as source of information for corrective actions or for evaluating the utilization of resources. The globalization of world economy has further amplified the importance of telecommunications to the economy, not only in Nigeria, but anywhere in the world [3]. Nigeria's GSM industry is one of the fastest growing and the largest Telecommunications industry in Sub-Sahara Africa and Africa at large. The focus of the GSM operators is gradually shifting from providing coverage to providing quality service [2]. The Euphoria of having a hand set (mobile phone) is gradually giving way to complaints of bad or poor service quality due to the rate of drop calls and congestion. The longer session of rains in Nigeria and its effect on the Signal strength is a critical factor, in addition to the incessant power outages that also contributes to poor service, which must be considered in the planning and ensuring good and better QoS by any of the telecommunications operators. The increase in the call drop rates and poor network availabilities within the localities or even while moving from one location to another have become a great concern to every network user and researchers in the academic profession. These affect businesses, security in the case of military operation, research activities for the academia, etc. The network operators keep expanding their base stations and capacities but the problem of poor call qualities persist. Hence, the need for network optimization for proper handover is important as a means to improve the quality of service. Three fundamental indices are used in monitoring mobile network performance, which include drive test, customer complaints, and network statistics [12]. The quality of network is, therefore, a key index in the race for users. This can be achieved when the network is adequately optimized to meet the expected QoS from its clients [8, 13].

II. MATERIALS AND METHODS

The main tools used in this paper were:

i. File Transfer Protocol (I Manager M2000) that was used for data collection from the network.
ii. Microsoft Excel tool box was used to plot the results for easy understanding and interpretation.

The collected data from MTN network management center was then analyzed based on KPIs values gathered. The methodology was a step by step procedure that is itemized as follows:

i. Analysis of the data set.
ii. Determination of average KPI (CSSR and CDR) values.
iii. Evaluation of the performance values

a. Analysis of Collected Data

The analysis of the data set for each of the studied locations were investigated base on daily assessment from the Base Station Controllers (BSCs), which were Abuja BSC number 7 (ABHBSC7), Abuja BSC number 13(ABHBSC13), Kaduna BSC number 8 (KDHBSC8), Kaduna BSC number 15 (KDHBSC15), Kaduna BSC number 4 (KDHBSC4), Kaduna BSC number 1(KDHBSC1), Kaduna BSC number 3 (KDHBSC3), and Kaduna BSC number 9 (KDHBSC9), respectively. At the Network Management Switching (NMS) center, a File Transfer Protocol (FTP) tool known as the I Manager M2000 was used to pull out the data from the network.

b. Determination of Average of CSSR

For each of CSSR and CDR the average variable was already evaluated from the raw data set obtained from the Network Management Center (NMC). The total average of CSSR and CDR on a daily basis for a month was obtained by using MS-Excel average tool box. e4rr4

In telecommunications, CSSR indicator measures the ease with which calls are established or setup. The higher the value, the easier it is to set up a call. High CSSR is achieved when Standalone Dedicated Control Channel (SDCCH) seizures and Traffic Channel (TCH) allocation are easily achieved to set up a call. It is calculated as number of the unblocked call attempts divided by the total number of call attempts [10]. The amount of traffic flow that would occur if all call set-ups and call attempts were successful is referred to as offered traffic, that is, traffic successfully offered. CSSR is the number of successful attempts to make a call. In an ideal world, a network should be capable of accepting all the calls attempted to be made. The best value of CSSR is 1 i.e. the network should be capable of accepting 100 % of the calls made. CSSR is found out during a short call [10]. The CSSR is also defined as the fraction of the overall total attempts made to make a call that result in a connection to the dialed number. The CSSR is one of the KPIs used by network operators to assess the performance of their networks. It has direct influence on the customers' satisfaction on the services provided by the network and it is expressed in percentage as:

$$CSSR = \alpha/\beta \times 100\% \quad (1)$$

Where,
α = Number of Call Setup
β = Number of Call Attempt

c. Determination of Average of CDR

In GSM network operation, call drop is a major observed problem frequently complained by subscribers. The call drop rate, however, is an important index that greatly influence the quality of a wireless network. This paper is set to carry out analysis on call drop rates in Kafanchan, Kaduna North, Kaduna South, and Zaria geographical locations in order to establish the frequency of the drop based on the data collected from the operation and maintenance center of NMC. The paper would discuss the cause of call drop and how to locate and resolve the problem that influences call drop in order to improve network quality. CDR, expressed as follows [4]:
It can be expressed as,

$$DCR = \varphi/\mu * 100\% \quad (2)$$

Where,
φ = Number of Drop per Calls Number of Call Setup
μ = Number of Call Setup

Equation (2) can be considered as the rate of calls that are not completely successful due to quality deprivation. In mobile network, DCR gives a quick overview of network quality and revenue lost. It is a Key Performance Indicator to determine service retains ability. It has a direct great control on the subscribers' satisfaction and reliability [4]. The average values were sorted daily for each of the three months (January, February and March 2016).

d. Evaluation of KPI Values

Evaluation and performance analysis of the quality of service results that were obtained on CSSR and CDR. The performance values for Kaduna South, Kaduna North, Kafanchan, and Zaria computed for the months of January, February and March 2016 were plotted on a graph together with those values the NCC benchmark for comparison.

The NCC is empowered by Government to establish minimum QoS standards in service delivery for the telecommunications industry in Nigeria. These QoS technical standards require operators meet basic minimum quality levels for all services provided to ensure that consumers continue to have access to high quality telecommunications services [7]. Table 1 is the set NCC KPI benchmarks followed by all operators of mobile services in Nigeria.

Table 1: NCC KPI Benchmarks [7]

S/NO	QUALITY PARAMETER	TARGET VALUE
1	Call Setup Success Rate (CSSR)	≥ 98 (%)
2	Call Drop Rate (CDR)	≤ 2(%)

From Table1, the QoS was measured base on these NCC standard KPI values for all the key performance parameters influencing the QoS of a network.

III. ANALYSIS, RESULTS, AND DISCUSSION

Performance analysis to evaluate the QoS KPI values of this work against those of NCC was carried out in this Section. Graphical representations of the results for the months of January, February, and March, 2016 on a daily basis, were produced to facilitate comparison and discussion of the results.

a. CSSR Results for January, 2016

From Figure 1, Zaria area suffered low performance with CSSR QoS level below the NCC QoS bench mark of ≥98% through the month, except on 9th, 12th and 14th when it was almost 98%. Other days (4th, 7th, 28th through 31st) experienced poor call accessibility, particularly the 25th when there was a sharp drop of the ratio of call setups to call attempts. The average value for the CSSR in Zaria for the month of January was 97.27%, which failed to meet the NCC standard benchmark minimum of 98%. On average, Kaduna

South suffered an unstable availability of CSSR ratio average value of 97.57% in January, which was below the [7] benchmark for successful SDCCH and TCH, respectively.

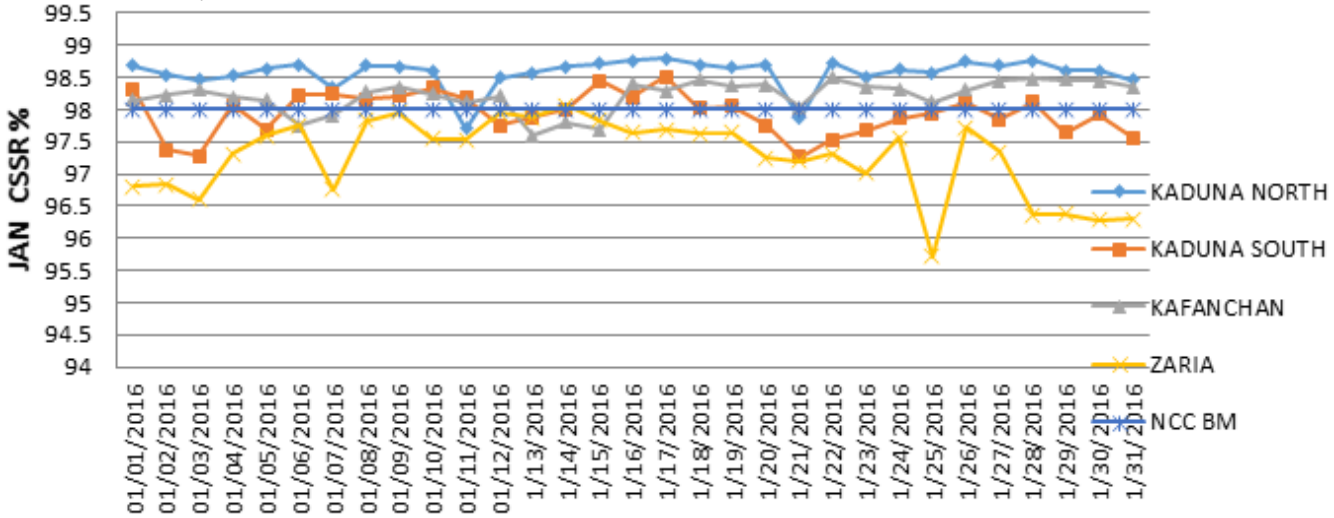


Figure 1: January, 2016 CSSR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Based on the data obtained from the NMC, Kaduna North and Kafanchan were able to meet the stipulated values of ≥98% with average values of 98.46% and 98.361%, respectively, although subscribers were unable to establish or set up calls in Kaduna North successfully on 11thJanuary. While Kafanchan performed below the NCC benchmark on the 6thand 13ththrough 15th of January as reflected in the graphical plots.

These worst short falls in performances for Zaria and Kafanchan were due to power failures during those days of January. On the other hand, Kaduna South had an average performance of 97.57% and Kaduna North had the best call set up rate with an average value of 98.57% base on set NCC standard bench mark of 98%.

b. CSSR Results for February, 2016

From Figure 2 graphical plots of CSSR values for the month of February, Zaria, and Kaduna South performed below the NCC minimum bench mark of 98% during the early stage of the month, specifically on 1st and 2nd of February. The failure to meet [7] benchmark standards was attributed to power failure due to lack of diesel fuel.

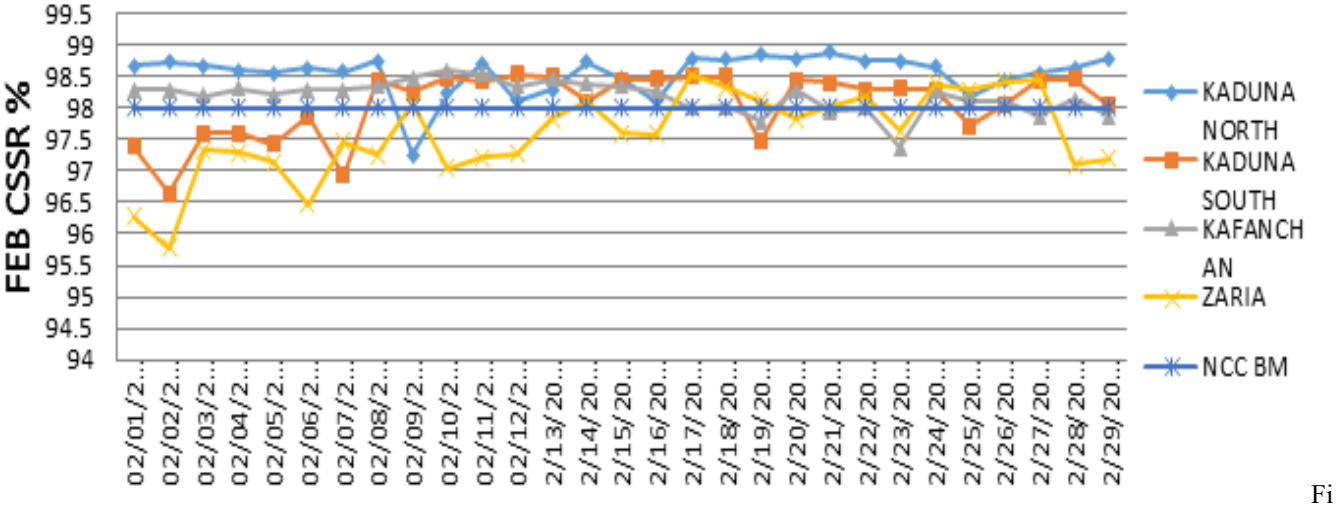


Figure2: February, 2016 CSSR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Failure to meet NCC minimum bench mark standards was again caused by power failure due to lack of diesel fuel. Kafanchan was able to maintain its good performance with an average monthly CSSR value of 98.19%. As for Kaduna North, there was a very good performance of service all through the month except on 9th with a sharp fall in performance as a result of major power outage recorded on the cell site, resulting in zero percentage availability percentage TCH for complete 24hrs. The reasons or causes for not been able to meet the NCC minimum bench mark standard of 98% availability in the four studied locations were majorly due to power failure, as well as other issues such as TCH congestion, interference, poor coverage, and faulty hardware units in those equipment at the base stations or cell sites.It is therefore evident from Figure 2 that Kaduna North had the best CSSR for the month of February, 2016 with an average value of 78.79%.

c. CSSR Results for March, 2016

March plots of Figure 3 shows that Kafanchan had a very poor ratio of successful numbers of set up calls against successful numbers of attempted calls and as a result subscribers had an unpleasant experience throughout the month of March compared to the previous months of January and February. Within Kafanchan location, several sites or base stations had zero percent TCH availability on the 1st, 2nd, 4th, 5th, 6th, and 7th with an average CSSR of 95% (less than minimum NCC benchmark [7] of 98%)

from a hub base station ID number T1874.This affected four other cell sites due to faulty generator that caused power outage, hence, resulting to subscribers' bitter experience of poor QoS from MTN network during the month of March, 2016. Kaduna North maintained its best performance with an average CSSR value of 98.33% in line with the [7] benchmark of ≥ 98%.

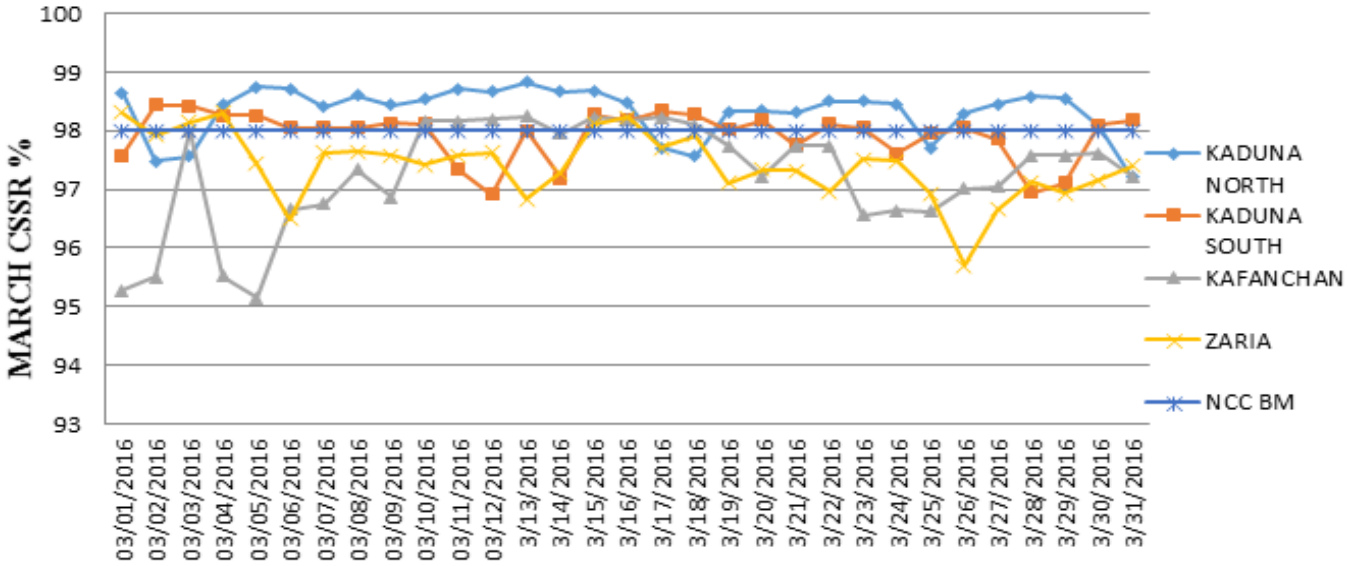


Figure 3: March, 2016 CSSR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

The Zaria plot, however, shows that there was slight average performance improvement over Kafanchan with a value of 97.41% against 97.25% for Zaria. Except for days like 1st, 3rd, and 4thwhen the performance was above 98%, other days it was below the standard QoS level expected by the regulatory body, [7]. In general, these showed that subscribers had a bitter experience throughout the three months considered in this research work.

d. CDR Results for January, 2016

Figure 4 shows the rate of call drop encountered duringthe month of January for the ares studied. It could be observed that Kaduna North experienced more call drop beyond the [7] stiputed standand of ≤2% on 10th, which means that subscribers in these period had a poor QoS after estabilshement of calls. It almost hit the standard benchmark of ≤2% on 23rd January with good QoS.

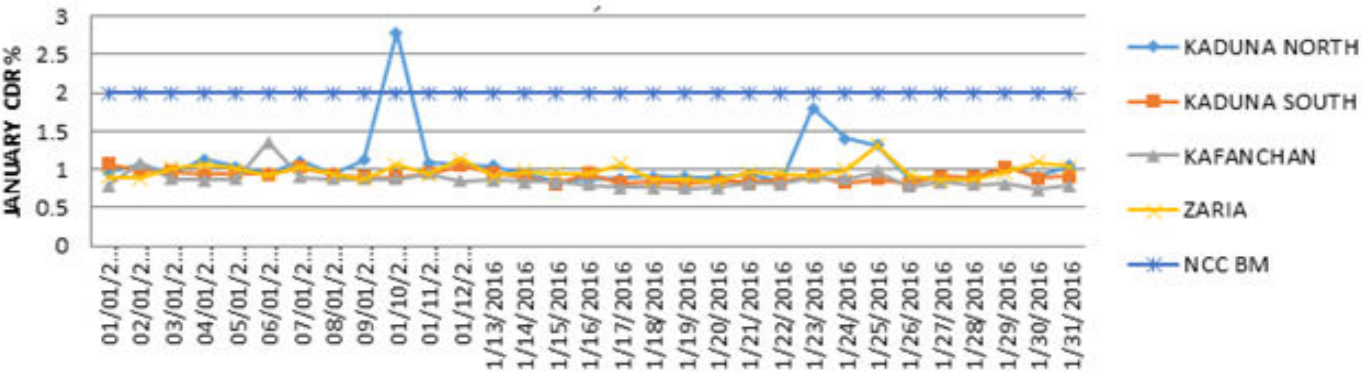


Figure 4: January, 2016 CDR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

On the average, the percentage call drop in January in Kaduan North was as good as 1.08 %, Kaduna South registered 0.911%, Kafanchan had 0.86 %,and Zaria recorded 0.97 %. These implied that there was a minimal drop of established calls after TCH assignment in most of areas studied and it was a good performance with good QoS for all the locations studied.Considering the KPI CDR, Kafanchan had the best QoS, Kaduna South had the second best,while Zaria;srecord was not good enough and Kaduna North had the worst performanceof the four areas.

e. CDR Results for February, 2016

From Figure 5, it could be observed that between 12th and 17th of February, there was an abruptly termination of an established calls almost in all the locations while conversation was ongoing, most especially, in Kaduna North and Kaduna South,where the call drop rates were more than the [7] benchmark of ≤ 2%. These resulted in poor QoS in these locationsdue to major transmision failures between the base station controller and base transceiver stationlink or the link between the base station controller and mobile switching centre.

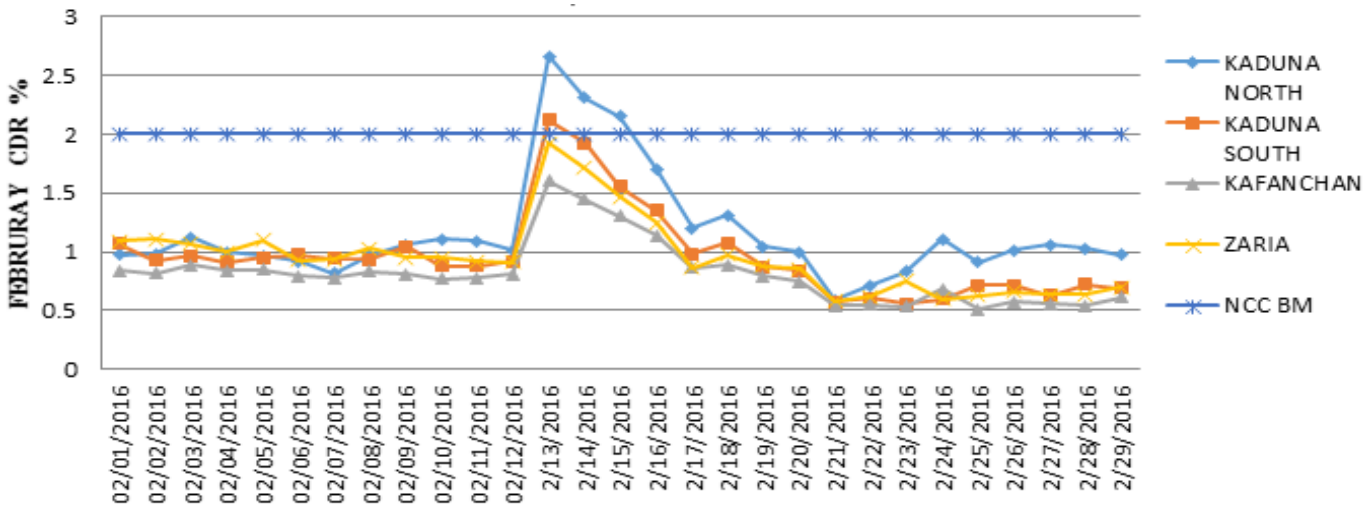


Figure 5: February, 2016 CDR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Low signal strength on downlink or uplink, lack of best server, congestion in neighboring cells, battery flaw, poor quality on down or uplink, missing neighboring cells definitions, unsuccessful outgoing handover, unsuccessful incoming handover, low BTS output power are other issues which could have contributed to the poor performance in these areas. The overall performance for the month showed that Kaduna North achieved an average performance of 1.16%, Kaduna South had an average performance of 0.96%, Kafanchan recorded an average performance of 0.82% and Zaria had an average performance of 0.95%. These implied that Kafanchan with the least average value of CDR had good performance and better QoS than any of the three areas of Kaduna North, Kaduna South, and Zaria. Comparing these values with the NCC standard benchmark values of ≤2%, the performance of each of the three areas studied was very good due to low CDR values recorded by the four locations researched into.

f. CDR Results for March, 2016

Figure 6, it was observed that there was a high rate of call drop beyond the NCC benchmark of ≤2% from 24th to 27th March, 2016 for Kaduna North and for Kaduna South,this occurred between 3rd and 7thof March, 2016. These high call drop rates were as a result of poor quality on downlink or uplink, missing neighboring cells definitions, unsuccessful outgoing handover, unsuccessful incoming handover, low BTS output power.

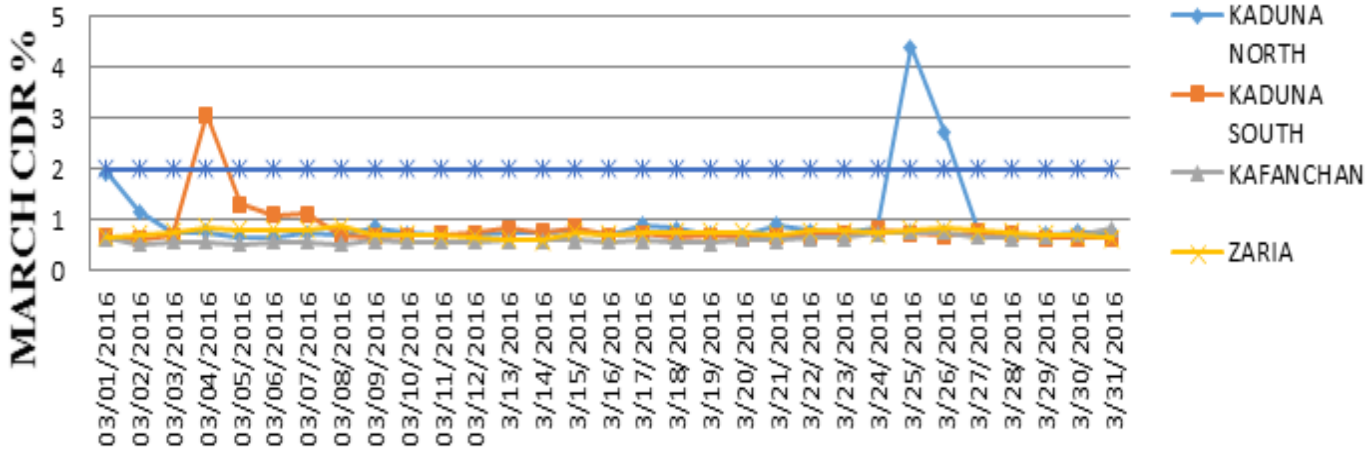


Figure 6: March, 2016 CDR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

The total average value of call drop rate for Kaduna North was0.99 %, Kaduna South achieved 0.83 %, Kafanchan recorded as low as 0.62%, and Zaria had 0.75%. These values for the respective areas studied in March, 2016 were also good because each area registered a value less than the standard benchmark value of ≤2% recommended by NCC, particularly Kafanchan area with the CDR value as low as 0.62%.

IV. RECOMMENDATIONS

In order to improve the network performance and the quality of service, it is recommended that the management of MTN Nigeria and the government regulatory body, NCC should ensure the following:

1. Quality of service for each service rendered to subscribers by network operators such as MTN should be upgraded to 99.50% for successful calls to ensure customers satisfactions.
2. Government should ensure adequate power supply to their site to minimize site downtime due to power failure, also BTS power consumption should be minimized at the design point by manufacturers such as Huawei, Motorola, Ericsson, Alcatel, etc
3. Proper definition of neighbouring BTSs and BSCs should be done for proper and successful handover.
4. MTN should ensure they carry-out constant network upgrade in other to increase the capacities of all existing BTSs in order to be able to handle the subscribers'

demand all the time. This will improve the call set up time and also meet up the stipulated KPIs threshold values for good network performance.

V. CONCLUSION

The quality of service of the KPI results for in the four locations (Kaduna North, Kaduna South, Kafanchan, and Zaria) during the three months' period of January, February, and March, 2016 showed considerable good performance by the MTN network in all the locations in terms of call drop rate because they all performed well within the [7] benchmark. Kaduna North experienced good call setup success rate, but other locations require some improvement in order to meet up NCC benchmark. The results also indicated that it was only in Kafanchan that the subscribers enjoyed good quality of handover of calls, but not in Kaduna North, Kaduna South, and Zaria where improvement in needed. In general, the entire MTN network need to be expanded base on their large number of subscribers in Nigeria (Kafanchan, Kaduna South, and Zaria, in particular). The large traffic caused congestions which in turn result in high call setup failure rate and handover failure rate Therefore, if the large traffic volume problem is addressed it will ensure good quality of service in the MTN network and in turn their subscribers will enjoy unlimited good quality of service.

REFERENCES

[1]. Adegoke A.S. and Babalola I. T. (2011), "Quality of Service Analysis of GSM telephone system in Nigeria," America Journal of Scientific and Industrial Research, Am. J. Sci. Res, 2(5): pp707-712J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2]. Adegoke A.S, I. T. Babalola and W. A. Balogun (2008), "Performance Evaluation of GSM Mobile System in System in Nigeria." The pacific Journal of Science and Technology, Vol 9, Number 2, pp 23-26.K. Elissa, "Title of paper if known," unpublished.

[3]. Adekitan, Rasheed Akanfe (2014), "Performance Evaluation of Global System for Mobile Telecommunication Networks in Nigeria".SCSR Journal of Business and Entrepreneurship (SCSR-JBE) Volume 1, Issue 1 (February, 2014), pp 09 – 21

[4]. Gordon Ononiwu1, Bukola O.H. Akinwole 2, Cosmos Agubor1 and James Onojo (2016) " Performance Evaluation of Major Mobile Network Operators in Owerri Metropolis of Nigeria" ISSN (Print): 2279-0047, ISSN (Online): 2279-0055, pp. 06-13 International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

[5]. Idigo V. E., A.C.O. Azubogu, C. O. Ohaneme and K. A. Akpado(2012), "Real-Time Assessments of Quality of service of Mobile Cellular Networks in Nigeria". International Journal of Engineering Inventions. Vol.1, Issue 6, pp 64-68

[6]. Joseph Isabona and Kingsley Obahiagbon (2014), "A practical Optimisation Method to Improve QoS and GoS-Based Key Performance Indicators in GSM Network Cell.

[7]. Nigeria Communication Commission (NCC, 2015) "Quality of service and QoS Metrics Use by Nigeria Communication Commission." pp 93-107, www.ncc.gov.ng

[8]. Nnochiri Ifeoma.U. (2015), "Evaluation of The Quality of Service of Global System for Mobile Telecommunication (GSM) Operators in Nigeria" Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040, Vol. 2 Issue 7, pp 1686-1693.

[9]. Obota, M.E and Agbo, P.E. (2012), "Strategies for Improving Quality of Service (QoS) of Global System for Mobile Communication (GSM) in Nigeria," Journal of Research and Development. Vol.4. Number 1, pp123- 127

[10]. Ozoveche A.S and A.U Usman(2015), Performance Analysis Of GSM Networks In Minna Metropolis of Nigeria.Nigerian Journal of Technology (NIJOTECH) Vol. 34 No. 2, pp. 359 – 367 University of Nigeria, Nsukka, ISSN: 1115-8443 www.nijotech.comhttp://dx.doi.org/10.4314/njt.v34i2.21

[11]. Sheik Manzoor A. K. and Shahabudeen F., (2018) "A Study on Key Performance Indicators and their Influence on Customer Satisfaction inCall Centers" International Journal for Engineering Research (IJER). Serials Publications 11(2), ISSN: 0972-9380. Pg 304-313

[12]. Ukhurebor K. E., (2018) "Evaluation of the Quality of Service of a Cellular Network using the Network Statistics" International Journal of Advanced Engineering and Technology ISSN: 2456-7655, pp 1-7, www.newengineeringjournal.com

[13]. Ukhurebor Kingsley E, Andikara John and Azi S.O., (2015) "Effects of Upsurge of Human Traffic on the Quality of Service of GSM Network in Eagle Square Abuja, Nigeria" Journal of Scientific and Engineering Research Volume 6, pp 1-6

[14]. VenkataSai Sireesha B. +, Dr.S.Varadarajan++, Vivek and Naresh+++ (2015) "Increasing Of Call Success Rate in GSM Service Area using RF Optimization" Engineering Research and Applications (IJERA). ISSN: 2248-9622. Vol. 1, Issue 4, pp1-7, www.ijera.com

ANALYSIS OF GSM NETWORK QUALITY OF SERVICE USING CALL SETUP FAILURE RATE AND HANDOVER FAILURE RATE INDICES

A. M. S. Tekanyi
Department of Electronics and Telecommunications Engineering,
Ahmadu Bello University,
Zaria, Nigeria.
amtekanyi@abu.edu.ng

H. A. Abdulkareem
Department of Electronics and Telecommunications Engineering,
Ahmadu Bello University,
Zaria, Nigeria
ha2zx@yahoo.com

Z. Z. Muhammad
Department of Electronics and Telecommunications Engineering,
Ahmadu Bello University,
Zaria, Nigeria.
zzmuhammad@abu.edu.ng

***Abstract*— Over the decades, the Global System for Mobile (GSM) communications technology has become one of the fastest growing and most challenging telecommunications technologies. The Call Setup Failure Rate (CSFR) and Hand-Over Failure Rate (HOFR) were the most important Key Performance Indicators (KPIs) used in ascertaining the efficiency of GSM network in terms of the quality of services rendered. For customers, it is expected that maximum satisfaction is derived from any service rendered, which was not the case over the years. The increase of HOFR and poor network availabilities due to increased CSFR became a great concern to all parties (providers, users, and researchers). Hence, this paper assessed the Quality of Service (QoS) of MTN GSM network in four geographical areas (Kaduna south, Kaduna North, Zaria, and Kafanchan) using two Key Performance Indicators (KPIs) of HOFR and CSFR. The data used in the assessment was collated from MTN Network Management Center (NMC) with the aid of the FACTS tool. These KPI results were evaluated against those specified by the Nigerian Communications Commission (NCC) in order to make some important recommendations from the findings (contributions) to improve the QoS of MTN network.**

***Keywords*— GSM technology, Subscribers, QoS, Call setup failure rate, Handover failure rate**

I. INTRODUCTION

Over the years, the mobile phone business in Nigeria witnessed tremendous improvement in terms of better coverage and accessibility, but not with the QoS that was also expected to be good. The poor QoS was mostly due to increased HOFR as well as poor network availabilities within the localities and even while moving from one location to another **[14]**. The GSM (cellular) network comprises of a Mobile Station (MS) that connects to the Base Transceiver Station (BTS) via air interface (A_{bis}). Network performance and QoS estimation are the two most vital aspects serious considered by mobile operators as the revenue base and client satisfaction are both directly related to them **[15]**. GSM communications is the most popular multi-services base network all around the world. One of these services included data communications by packet data transport via General Packet Radio Service (GPRS) and improved data rates for GSM using Enhanced Data for Global Evolution (EDGE). In additional improvements were achieved when the 3GPP developed the 3rd Generation (3G) Universal Mobile Telecommunications System (UMTS) standard followed by 4th Generation (4G), where 4G phones are connected to the Internet for faster and wider coverage **[9]**. Network running is not an easy task despite the equipment that is part of the network. The difficulties of network management are very much manifested when devices and equipment from different manufacturers are used. Detecting problems and issues becomes more difficult when circuits in the network do not have a single console under which they operate **[1]**. A Network Management System (NMS) is installed on the network with an online database that is responsible for the monitoring of all that happens on its network in raw data form. In order to ensure good network performance with better QoS for customers' satisfaction and retention, this data is analyzed and evaluated to spot events, trends, problem areas, and drift from recommended KPIs values **[14]** required for the evaluation of the QoS of a GSM network. There are several KPIs used in network QoS evaluation, some of which are network accessibility, service retainability, connection quality, and network coverage **[6]**. There are three key techniques commonly used to monitor network performance, which are drive test, network statistics, and customer complaints. Client complaints are the most commonly used technique because the client is always ready to give an input whether valid or not. Though performance estimation by this technique is easily achievable, it is not the best alternative because the clients' experiences can be touching and subjective **[13]**. A call centre is often operated through a widespread open workspace for call set/headset connected to a telecommunication switch with one or more monitoring stations. It can be separately operated or networked with extra centers, often connected to a business computer network, including mainframes, microcomputers, and LANs **[11]**. The drive test technique used by NCC in 2005 to determine QoS of the GSM networks in Nigeria, is another technique for performance estimation. Apart from network performance assessment, it can also be used for the detection of network problem areas, validation of effects of optimization changes and analysis of root cause of problems in an operational network **[13]**

II. MATERIALS AND METHODs

Some of the appropriate materials available for use in this paper were:

iii. Manager M2000 File Transfer Protocol (FTP) that was used to facility data collection from the network.

iv. Microsoft (MS) Excel tool box was used to plot the data values for easy understanding and interpretation of investigation results.

To realise the goal of this paper, a step by step procedure was adopted on the data collected from MTN Network Management Center in line with the established methodology, where data was first analyzed based on KPIs values gathered. The procedure was as follows:

- Analysis of the data set.
- ii. Determination of Call Setup Failure Rate (CSFR) and Handover Failure Rate (HOFR)
- Evaluation of the performance values.

A. Analysis of Collected Data

Analysis was based on the daily evaluation of Base Station Controllers (BSCs), which were Abuja BSC number 7 (ABHBSC7), Abuja BSC number 13 (ABHBSC13), Kaduna BSC number 8 (KDHBSC8), Kaduna BSC number 15 (KDHBSC15), Kaduna BSC number 4 (KDHBSC4), Kaduna BSC number 1 (KDHBSC1), Kaduna BSC number 3 (KDHBSC3), and Kaduna BSC number 9 (KDHBSC9). Analysis of the data set was then carried out to investigate the QoS performance for each of the studied geographical locations. At the Network Management Switching (NMS), a FTP tool known as the I Manager M2000 was used to draw out the data from the network.

B. Determination of Average CSFR

From the two indices of CSFR and HOFR, this study was conducted using the average variable that was already evaluated using the raw data set obtained from the NMC. The total average of CSFR and HOFR on a daily basis for a month was obtained by using MS-Excel tool box. There could be lots of reasons for a poor CSSR **[15]**, amongst them are the following:

- Low Signal Strength.
- Standalone Dedicated Control Channel (SDCCH) Congestion.
- CM Service Reject.
- Traffic Channel (TCH) Failure Assignment.
- Hardware Problem.

The problem of QoS is directly affected by the mobility of subscribers and difficulty of the radio wave propagation **[3]**, **[10].** However, most of the network problems are caused by increasing subscribers that may result in congestion in the network and the changing environment that which may have harsh conditions impossible to permit radio wave propagation **[3]**, **[5]**. RF optimization is a continuous process that is required in meeting the best conditions of service in an evolving network **[15]**. CSFR was obtained from the formula of equation (1) and the average values were sorted daily for each of the months (January, February, and March, 2016). The CSFR refers to the number of call setup failure divided by the total number of call attempts. It is also called the blocking probability and it is expressed in percentage **[7]**, **[4]** as:

$$CSFR=\mu\backslash\beta\times100\% \quad (1)$$

Where,

μ = Number of Blocked Calls

β = Number of Call Attempt

C. Determination of Average HOFR

Handover failure rate is an important indicator for monitoring mobility. Handover failure means that mobile station tries to make a handover (inter/intra cell) but for some reasons it fails **[7]**, **[4]**. These values are kept within some precise threshold in order to meet the required QoS criteria accepted by the regulatory bodies like Nigerian Communications Commission (NCC) for subscribers' satisfaction **[3]**.

It is expressed in percentages as:

$$HOFR = \alpha/\ \delta\times100\% \quad (2)$$

Where,

α = Number of Handover Failures

δ = Number of Handover Attempt

D. Evaluation of Performance Values

Performance analysis for the evaluation of the QoS results obtained from the study of this paper was conducted using Call Setup Failure Rate (CSFR) and Handover Failure Rate (HOFR) as indices. The performance index values were plotted on graphs for Kaduna South, Kaduna North, Kafanchan, and Zaria for the month of January, February, and March, 2016 and measured against the NCC benchmark.

The NCC is empowered to establish minimum QoS standards in service delivery for the telecommunications industry. These QoS standards ensure that consumers continue to have access to high quality telecommunications services because basic minimum quality levels are set for all operators to satisfy in order to remain in operation (NCC technical standard, 2016). Table 1 gives the NCC KPI benchmarks.

Table 1: NCC KPI Benchmarks (NCC, 2016)

Quality Parameter	**Target Value**
Call Setup Failure Rate (CSFR)	≤ 2(%)
Handover Failure Rate (HOFR)	≤ 2(%)

From Table 1, the QoS was measured based on the NCC standard values for each of the key performance indicators.

III. RESULTS ANALYSIS AND DISCUSSION

Analysis of QoS results using CSFR and HOFR as performance indices is done in this Section. Graphical representations were used for the plots of daily data values for Kaduna South, Kaduna North, Kafanchan, and Zaria for the months of January, February, and March, 2016. This is the ability of a subscriber to initiate a call and granted access. Technically, during a GSM call setup, a speech call is assigned from a SDCCH to a TCH. If the TCH selected suffers from interference, then the mission fails. In addition, the mission's failure message is sent to the Mobile

Switching Center (MSC). The call is then re-established back **[8].**

A. CSFR Results for January, 2016

Figure 1 shows the analysis of January call failure rate. Zaria was completely out of the threshold restricted limit of NCC throughout the month of January by its inability to establish calls successfully, except on 14th day of the month. This was an indication of poor QoS provided to subscribers during the month January, 2016.

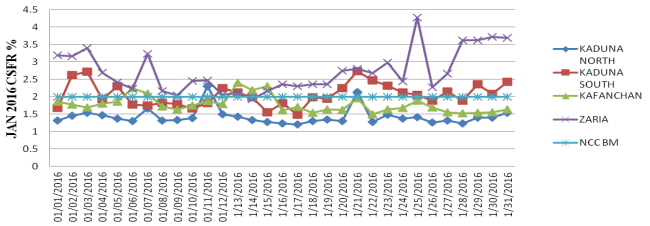


Figure 1: January, 2016 CSFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

From data analysis most often cells availability reached 100%, but the (PTTCH) performance was less which indicated that there was high congestion of SDCCH and the TCH. This made it difficult for any subscriber to have a successful ratio of calls unblocked to the calls attempted. Kaduna North had the best performance with average failure rate of 1.43% compared to the NCC benchmark of 2%. Kafanchan also had a good performance, except on the 6th, 7th, 13th, 14th, and 15th days of the month, which could have been due to power failure (generator fault, hybrid dc power failure, and hardware unit's faults, etc). As for Kaduna South, the failure rate was higher with an average record value of 2.06% when compared to the threshold level of ≤ 2% set by NCC. There was frequent fluctuation in the network due to lack of power supply to service some of the cell sites or Base Transceiver Stations (BTS) in their respective locations. These cell sites had power failure and a sector was down completely on the 13th, likewise sector E of another cell site registered zero performance for long hours. All of these resulted into low signal strength for effective coverage of the neighboring base station, thereby causing a rise in the failure rate.

B. CSFR Results for February, 2016

Figure 2 shows the CSFR of the four locations during the month of February based on the statistics obtained from Network Management Centre (NMC) and the graphical representations. Zaria had the worst performance, with monthly average value of 2.41% that was more than the NCC threshold of ≤ 2%. From the NMC, it showed that at the beginning of the month 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 10th and 11th, except on 9th, subscribers were unable to gain call access initiated to the BTS through to the MSC as a result of poor coverage and due to power outages recorded at the base stations and hardware faults on some of the cells. Kaduna South and Kafancha recorded average performances were 1.93% and 1.81%, respectively which were within the NCC benchmark of ≤ 2%. Between 8th to 18th of February and 18th through 30th, there were severe failure rates in the three locations of Kaduna south, Zaria, and Kafanchan. The causes of this rise could have been due to TCH congestion, interference, and poor coverage. On average, Kaduna North had the excellent best service performance of 1.45% CSFR.

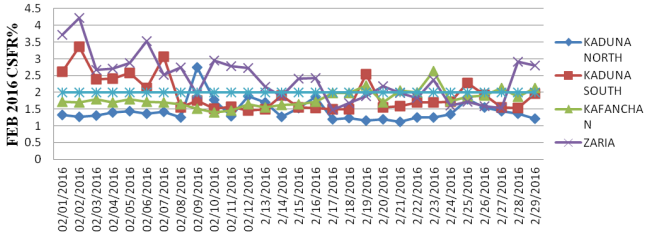


Figure 2: February, 2016 CSFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

C. CSFR Results for March, 2016

Kafanchan area of Figure 3 had a high rise in failure rate from the beginning of the month 1st to 10th, except for the 3rd that had a sharp improvement, but was unable to sustain that this availability from 19th to 31st of the month. Network performance was seriously degraded with the monthly average value of CSFR as high as 2.75%, greater than the recommended NCC threshold of ≤ 2%. This was as a result of TCH congestion, frequent power failure, and faulty hardware.

Also, in Zaria, the subscribers experienced bad network performance with the persistent failure rate throughout the month, except for 1st, 3rd, and 4th whose performances fall within the NCC threshold benchmark of ≤ 2%. From the analysis of the collated data, hardware unit faults and protracted power failures contributed to the low PTTCH or cell availability for most of the BTSs which led to poor coverage, thereby impacting negatively on the network. Zaria had an average CSFR value of 2.59% that was more that recommended NCC benchmark of ≤ 2%.

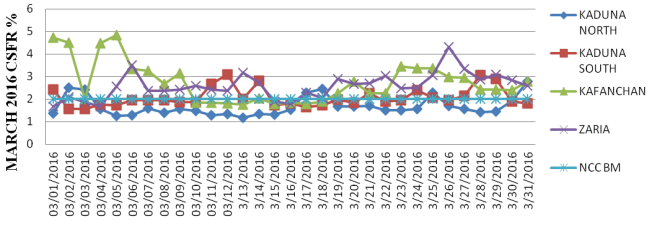


Figure 3: March, 2016 CSFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Considering the recommended CSFR value of ≤ 2% in the results for Kaduna North, Kaduna south, Kafanchan, and Zaria, Kaduna South had an overall monthly average performance of 2.07%. Kaduna North maintained its best performance with a low failure rate of 1.67%. Although, some days like 2nd, 3rd, 17th, 29th and 31st had failure rates more than the standard threshold value of NCC. The causes for the rise on these days were because of TCH congestion, interference, poor coverage due to power failures at the BTS sites and Faulty hardware units.

D. HOFR Results for January, 2016

Handover failure rate as a key performance indicator is the percentage of time that calls failed to be handed over to the

next neighboring BTS or BSC successfully. The NCC recommended benchmark Handover Failure Rate (HOFR) is set at 2%. This implies that HOFR value greater 2% indicates poor network quality and poor QoS. Figure. 4 shows the analysis of the HOFR for the four geographical areas studied in the month of January, 2016 on the assessment of their QoS performances. It was observed that Zaria had the worst performance on its HOFR, follow by Kaduna North and Kafanchan, while Kaduna South had the best HOFR record.

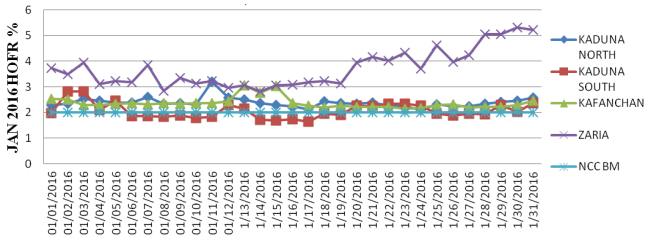


Figure 4: January, 2016 HOFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

From the graph, it is observed that the rise of Kaduna South's HOFR values above the 2% benchmark of NCC were recorded on the 2nd, 3rd, 4th, 5th, 20th through 24th, and 28th, to 30th of January. This rise of HOFR values are caused by traffic congestion, link connection failure, bad radio coverage, incorrect handover ratio, incorrect locating parameter setting, and high interference. Looking at the Zaria graph, there is sharp rise above the 2 % recommended value of NCC throughout the month of January due to same factors which affected HOFR of Kaduna South. Kafanchan and Kaduna North had few and mild sharp rise from the beginning to the end of the month .The causes for this rise were as a result of TCH congestion, interference, and poor coverage. These four locations studied showed an average HOFR performance for the month of January, 2016 as follows: Kaduna South recorded 2.07 %, Kafanchan had 2.37 %, Kaduna North registered 2.39 % and Zaria with worst performance of 3.721 %, which were all greater than the NCC recommended ≤ 2%.

E. HOFR Results for February, 2016

The assessment in Figure 5 shows that Kafanchan had the best service performance on the HOFR parameter during February, 2016 when compared to Kaduna South, Kaduna North, and Zaria. Kafanchan's performance was better than the rest of the locations, while Kaduna South was second best on the network's performance on HOFR.

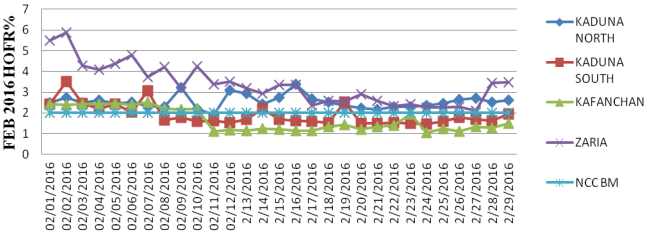


Figure 5: February, 2016 HOFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Zaria showed bad performance, where there were heavy fluctuation and sharp rises from the beginning right through 29th February, 2016. The performance was not in line with the benchmark standard of the NCC. The causes of these fluctuations were due to traffic congestion, link connection failure, bad radio coverage, incorrect handover ratio, incorrect location parameter setting, high interference, hardware failure, and the mobile station wrongly measuring signal strength of another co-or-adjacent cell. As can be seen from the graph Kaduna South had few rises on the 1st, 2nd, 3rd, 4th, 5th, 6th, and 7th while other days were 14th and 19th of February, 2016. The average monthly HOFR performances for other locations were as follows: Kafanchan had the average performance of 1.65 %, Kaduna South had 1.907140956%, that of Kaduna North was high at 2.54 %, and worst was Zaria with an average value of 3.36 %. The average performances of Zaria and Kaduna North were not in accordance with the NCC standard value of ≤ 2% for good QoS, which was an indication of poor network service in the two areas in February, 2016.

F. HOFR Results for March, 2016

Figure 6 shows the grahical assessment representations of HOFRs for Kaduna North, Kaduna south, Kafanchan, and Zaria for the month of March, 2016. There were great deviations of handover rate values for all locations studied which were above the standard ≤ 2% recommended threshold value by NCC.

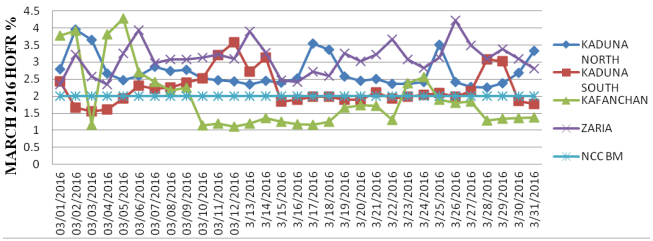


Figure 6: March, 2016 HOFR Results for Kaduna North, Kaduna south, Kafanchan, and Zaria

Only Kafanchan had a better QoS on handover rate from 10th to 22nd of March and between 25th-31st March,2016 with a total monthly average handover rate performance of 1.92 %, which was within the NCC benchmark. Looking at Kaduna North from the graphical analysis, there were high handover failure rates throughout the month, with an average performance HOFR value of 2.71 %, way above the threshold limit of ≤ 2% stipulated by NCC.

Also, Zaria's performace was very poor, with high rise in HOFR value of 3.09% well above the NCC benchmark recommended value of ≤ 2% which implies that MTN subcribers had a bitter experience on QoS in this Zaria location . Similarly, it was discovered that the causes for the rise in HOFR values and their fluctuations were due to traffic congestion, link connection failure, bad radio coverage, incorrect handover ratio, incorrect location parameter setting, and high interference, co-channel or adjacent-channel service, which resulted to improper handshaking during call establishment from one location to another. Kafanchan had the best performance, followed by

Kaduna South with 2.23 % and worst for Kaduna North and Zaria.

IV. RECOMMENDATIONS

In order to improve the network performance and the quality of service, it is recommended that the management of MTN Nigeria and the government NCC regulatory body should ensure the following:

5. More base stations are installed and carry-out Radio Frequency (RF) optimization often on antennae to ensure total coverage and successful handover of calls in the network.
6. The federal government, NCC, and other regulatory bodies should enforce a regular monitoring schedule on networks by operators and performance reports at very short intervals are filed for inspection by NCC so that Quality of Service (QoS) is established to meet up demands of subscribers.
7. Similar researches should be carried out in other parts of the country using the same or other QoS parameters to give a more detail picture of the network trend in the country as a whole.
8. MTN, in particular, should ensure that all traffic affected alarms on the BTS such as Voltage Standing Wave Ratio (VSWR), Receive Signal Strength Indicator (RSSI), etc. are cleared to reduce call setup failure rate.

V. CONCLUSION

In general, Call Setup Failure Rate (CSFR) is caused by the following reasons which also contribute to unpleasant experience subscribers subjected to when NCC benchmark values are not respected.

- Low Signal Strength.
- Interference.
- High Congestion on Standalone Dedicated Control Channel (SDCCH).
- High Congestion on Traffic Channel (TCH).
- Transmission Fault.
- Faulty Transceiver Unit (TRU)/Hardware.
- Central Processor Overload.
- Software File Congestion.
- Cell is not defined in Mobile Switching Center (MSC).

REFERENCE

[1] Eleni Rozaki (2015) "Design and Implementation for Automated Network Troubleshooting using Data Mining" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.3, pp 9-27

[2] Galadanci G. S. M and Abdullahi S. B., (2018) "Performance Analysis of GSM Networks in Kano Metropolis of Nigeria", American Journal of Engineering Research (AJER) e-ISSN: 2320- 0847 pp. 69-79, www.ajer.org.

[3] Gordon Ononiwu1, Bukola O. H. Akinwole2, Cosmos Agubor1, and James Onojo (2016). "Performance Evaluation of Major Mobile Network Operators in Owerri Metropolis of Nigeria" ISSN (Print): 2279-0047, ISSN (Online): 2279-0055, pp. 06-13 International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), www.iasir.net

[4] Idigo V. E., Azubogu A. C. O., Ohaneme C. O., and Akpado K. A. (2012). "Real-Time Assessments of Quality of service of Mobile Cellular Networks in Nigeria", International Journal of Engineering Inventions. Vol.1, No. 6, pp 64-68.

[5] Jatin, Karishan Kumar (2016) ''Study and Analysis of Call dropping and Handover Problem in cellular system'' International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), ISSN: 2278 – 1323 Volume 5, Issue 6, pp1776-1

[6] Lawal B. Y., Ukhurebor K. E., Adekoya M. A., Aigbe E.E (2016) ''Quality of Service and Performance Analysis of a GSM Network in Eagle Square, Abuja and Its Environs, Nigeria''
International Journal of Scientific & Engineering Research, ISSN 2229-5518 Volume 7, Issue 8, pp1992-2003

[7] Martin KOLLÁR, (2008). "EVALUATION OF REAL CALL SET UP SUCCESS RATE IN GSM", Acta Electrotechnica etc Informatica Vol. 8, No. 3, pp. 54–56. ISSN 1335-8243 © 2008 FEI TUKE

[8] Nnochiri Ifeoma.U. (July, 2015). "Evaluation of the Quality of Service of Global System for Mobile Telecommunication (GSM) Operators in Nigeria", Journal of Multidisciplinary Engineering Science and Technology (JMEST), ISSN: 3159-0040, Vol. 2, No. 7, nwaife2002@yahoo.com.

[9] Rajesh Kumar Upadhyay, Vijay Kumar Singh, Rajnish Kumar, (May, 2014) "PERFORMANCE ANALYSIS OF GSM NETWORK", International Journal of Advance Research in Science and Engineering (IJARSE), Vol. 3, No.5, ISSN-2319-8354(E) pp. 244 www.ijarse.com.

[10] Shoewu, O. and Edeko, F. O (2011) "Outgoing Call Quality Evaluation of GSM Network services in Epe, Lagos State" AMERICAN INTERBATIONAL JOURNAL OF SCIENTIFIC AND INDUSTRIAL RESEARCH" SSN 2153- 649XZ (Am j. sci Ind. Res.) pp409- 417

[11] Sheik Manzoor A. K. and Shahabudeen F., (July-Dec., 2014). "A Study on Key Performance Indicators and their Influence on Customer Satisfaction in Call Centers", International Journal for Engineering Research (IJER). Serials Publications 11(2), ISSN: 0972-9380, pp. 1-6

[12] Syed Foysol Islam, Fahmi Ahmed (2014) ''Simulation and Modeling of Handover Failure and Call Drop In GSM Network for Different Scenarios'' International Journal of Computer Science and Information Security (IJCSIS), Vol. 12, No. 4, pp 18-27

[13] Ukhurebor K. E, Awodu O. M, Abiodun I. C, Azi S. O., (Oct., 2015). "A Comparative Study of the Quality of Service of GSM Network during Crowd Upsurge in University of Benin Nigeria", International Journal of Scientific & Engineering Research, Vol. 6, No. 10, ISSN 2229-5518.

[14] Ukhurebor Kingsley Eghonghon, (Nov., 2017). "Evaluation of the quality of service of a cellular network using the network statistics", International Journal of Advanced Engineering and Technology. ISSN: 2456-7655, Vol. 1; No. 5; pp. 01-07 www.newengineeringjournal.com.

[15] Venkata-Sai Sireesha B., Dr. S. Varadarajan, Vivek and Naresh, (2015). "Increasing of Call Success Rate in GSM Service Area using RF Optimization", Engineering Research and Applications (IJERA). ISSN: 2248-9622. Vol. 1, No. 4, pp.1479-1485. www.ijera.com.

Evaluating the State of the Art Antivirus Evasion Tools on Windows and Android Platform

Faisal A. Garba
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria
alifa2try@gmail.com

Kabiru I. Kunya
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria

Shazali I. Ibrahim
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria

Abubakar B. Isa
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria

Khadija M. Muhammad
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria

Nasiru N. Wali
Department of Computer Science Education
Sa'adatu Rimi College of Education, Kano, Kano, Nigeria

***Abstract*— Hackers use malware to gain access to target computers. Malicious payloads are usually generated using tools such as Metasploit. As a means of defense, the target computers deploy anti-virus solution to detect these malicious payloads and protect the victim machines. In a reaction to this, the hackers created anti-virus evasion tools to evade detection by this antivirus solutions. But how effective are these antivirus evasion tools? This paper seeks to evaluate the effectiveness of some selected anti-virus evasion tools: Avet, Veil 3.0, The Fat Rat, PeCloak.py, Phantom-Evasion, Shellter, Unicorn and Hercules against current best Antivirus Solutions on Windows and Android platforms.**

***Keywords*— hackers, anti-virus, evasion tools, malware, payloads**

I. INTRODUCTION

The use of malware, is one way of getting access to a computer and the generation of malware is relatively simple given the given the number of freely available tools that available on the Internet. Malwares generated using tools such as Metasploit are easily detected nowadays due to the techniques used by antiviruses. In a reaction to this, tools that enables these malwares evade antivirus detection are also developed [1]. Three main techniques used in detecting malwares include signature based, behavior based and heuristic based techniques [2] as cited in [1]. Antivirus evasion tool creators also deploy a variety of techniques to avoid detection. These techniques are encryption, oligomorphism, polymorphism, metamorphism, obfuscation, and code reuse attacks. Antivirus is one of the first line of defense attackers faced when they try to hack a computer. This obstacle was minimal in the past when most of the AV products uses signature-based detection and the effort required to evade them was insignificant. The techniques used today to evade antivirus are multi-layered and include heuristics, behavioral, cloud-based detection and static scanning [3]. Antivirus evasion tools is used by both malicious attackers and penetration testers. The practice used by security professionals to assess the security strength of a system is known as penetration test. Penetration test involves attacking the system so as to uncover vulnerabilities that could be exploited by malicious hackers and inform the client of those vulnerabilities and ways of mitigating those vulnerabilities [5]. Penetration testers uses the same tools, techniques and procedures as does ethical hackers in order to attempt to compromise a computer network. There is therefore a need to evaluate the effectiveness of these antivirus evasion tools used by black hat hackers and penetration testers to compromise computer networks, by evaluating the effectiveness of the evasion tools on Windows and Android platform.

II. ANTIVIRUS EVASION TOOLS

A. Veil 3.0

The Veil Framework comprises of tools used during penetration testing. Veil-Evasion is the most commonly used tool, which can turn transform arbitrary script or piece of shellcode into a windows executable undetectable by antivirus [6]. Veil-3.0 is available for download on the official github repository page at https://github.com/Veil-Framework/Veil.

B. Avet

The According to the official github repository page "AVET is an AntiVirus Evasion Tool, which was developed for making life easier for pentesters and for experimenting with antivirus evasion techniques" [7].

C. The Fat Rat

The Fat Rat is a tool used in generating a backdoor to evade Antivirus and also used in post exploitation attack. The Fat Rat compiles a malware with a popular payload and then the compiled payload can be executed on Windows, Android and Mac systems.

D. PeCloak.py

PeCloak.py was a product of result of an experiment with AV evasion by Mike Czumak [8]. The original PeCloak.py has however not been maintained, so a fork from the original PeCloak.py was created as peCloakCapstone which is a Platform independent. PeCloak.py fork based on Capstone is

available at the github repository at https://github.com/v-p-b/peCloakCapstone.

E. Phantom-Evasion 2.0.1

With the capability of generating almost fully undetectable executable even with the most common 32-bit msfvenom payload, Phantom-Evasion is an interactive antivirus evasion tool written in Python. Through the use of modules focused on polymorphic code and antivirus sandbox detection techniques, the tool aimed to make antivirus evasion an easy task for pentesters [9].

F. Magic Unicorn

Through injecting shellcode straight into memory, Magic Unicorn is a simple tool for using a PowerShell downgrade attack. Magic Unicorn is based upon the works of Matthew Graeber's powershell attacks and the powershell techniques presented by David Kennedy and Josh Kelly presented at Defcon 2018. Magic Unicorn has support for Cobalt Strike beacon into the PowerShell evasion framework. Unicorn also supports user shellcode inserted into various attacks including PowerShell attack, HTA and Macro attack vectors [11].

G. Hercules

Hercules is an adaptable payload generator that can evade antivirus software [12]. Hercules is available for download from github.

III. Review of Free Antivirus Products

A review of antivirus products by PC Mag and written by [13], listed out Avast Free Antivirus, Kaspersky Antivirus, AVG Free Antivirus, Bitdefender Antivirus Free Edition, Check Point ZoneAlarm Free Antivirus+ 2017, Sophos Home Free, Avira Antivirus, Adaware Antivirus Free, Comodo Antivirus 10 and Panda Free Antivirus as the best 10 free antivirus products.

In another review by Safety Detective and written by [14] listed out Norton Security Antivirus, McAfee Free Antivirus, Total AV Free Antivirus, Avira Free Antivirus, Panda Free Antivirus, Intrusta Antivirus, CYLANCE Antivirus, Heimdal Antivirus Free, Webroot SecureAnywhere Free and Bitdefender Antivirus Free Edition.

A review by Lifewire and written by [15] listed out Avira Free Security Suite, Bitdefender Antivirus Free, Adaware Antivirus Free, Avast Free Antivirus, Panda Dome, AVG Antivirus Free, COMODO Antivirus Free, FortiClient, Immunet Antivirus, and Kaspersky Free as the best 10 free antivirus solutions.

In another review by Tom's Guide and written by [16], the free antivirus solutions that made the first 10 list are Kaspersky Free Antivirus, Bitdefender Free Antivirus, Avast Free Antivirus, Microsoft Windows Defender, AVG, Avira, Panda, Malwarebytes. Tom's Guide uses four evaluation labs AV-TEST in Germany, AV-Comparatives in Austria, SE Labs in England and their own labs in Utah for testing the antiviruses malware detection capability [16].

Mashable review listed out 8 best free antivirus products as Avast, Bitdefender, AVG, Sophos Home Free, Panda Free Antivirus, ZoneAlarm Free Antivirus, Comodo Antivirus and Avira Free Antivirus [17].

Orphanides [18] wrotes the results of Antivirus review conducted by Trusted Reviews where Kaspersky Free Antivirus, Microsoft Windows Defender, Bitdefender Free Antivirus, Avira Free Antivirus, Avast Free Antivirus, and AVG Free Antivirus made the list of the 6 best antivirus free solutions.

IV. Selected Free Antivirus Products

Based on the above reviews, we awarded scores to each antivirus that made an appearance in a review a score of 1 point. The antivirus products with the most points were selected for our review in this research.

Table 6: Scoring Antivirus Products

Antivirus	Rubenking (2019)	Zacks (2018)	Fisher (2019)	Wagenseil (2019)	Allen (2019)	Orphanides (2019)	Total Score
Avast Free	1	0	1	1	1	1	5
Kaspersky Free	1	0	1	1	0	1	4
AVG Free	1	0	1	1	1	1	5
Bitdefender	1	1	1	1	1	1	6
Check Point	1	0	0	0	0	0	1
Sophos	1	0	0	0	0	0	1
Avira	1	1	1	1	1	1	6
Adaware	1	0	1	0	0	0	2
Comodo	1	0	1	0	1	0	3
Panda	1	1	1	1	0	0	4
Total	0	1	0	0	0	0	1
Norton	0	1	0	0	0	0	1
McAfee	0	1	0	0	0	0	1
Intrusta	0	1	0	0	0	0	1
Cylance	0	1	0	0	0	0	1
Heimdal	0	1	0	0	0	0	1
Webroot	0	1	0	0	0	0	1
FortiClient	0	0	1	0	0	0	1
Immunet	0	0	1	0	0	0	1
Windows Defender	0	0	1	0	0	1	2
ZoneAlarm	0	0	0	0	1	0	1

V. Methodology

The experiment will be carried out in a lab set up with VirtualBox VMware on a Windows 8 host machine with 64 bit processor, 12GB of RAM, Core 5 intel processor and

500GB HDD. Two virtual machines the attacking machine - Kali Linux and the target machine Windows 8 are setup in the VirtualBox. A physical Android phone will also be required. A wireless router is used to set up the network as seen in Figure 1. All the evasion tools will be installed on the Kali Linux machine to generate the malware samples and deployed to the target machines (Windows 10 and Android Phone) via Apache server for the target machines to download and run.

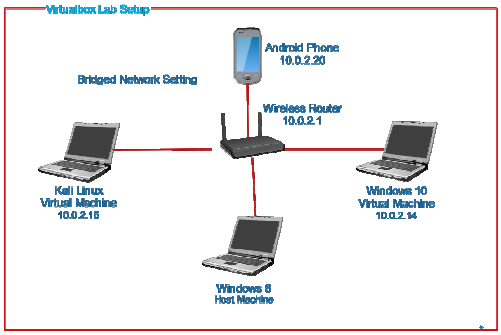


Figure 18: The Network Lab Setup

The antivirus solution will be installed one at a time, tested with a malware generated from one of the evasion tools and the test result recorded. If the antivirus software detects the malware, a score of 1 is awarded to the antivirus, otherwise it is awarded with a score of 0. The evasion tool is awarded a score of 1 point if it is able to evade the antivirus and a score of 0 otherwise. The antivirus software with the highest detection score will be awarded the most efficient and the evasion tool with the highest score will be awarded the best antivirus evasion tool. The tests results will be evaluated by comparing it with related works.

VI. Conclusion

The results of the research work will be the best antivirus evasion tool that will be recommended to penetration testers for use during engagements and the most effective antivirus solution that will be recommended to end users to use on their devices.

VII. Future Work

Future work will be the implementation of this research work. There is a need to extend this work to Linux and Mac platforms.

References

[1] C. Kalogranis, "AntiVirus Software Evasion: An Evaluation Of The AV Evasion Tools," University of Piraeus, 2018.

[2] J. Barriga and S. G. Yoo, "Malware Detection and Evasion with Machine Learning Techniques: A Survey," *International Journal of Applied Engineering Research* , vol. 12, pp. 7207-7214, 2017.

[3] W. Chen, "Encapsulating Antivirus (AV) Evasion Techniques in Metasploit Framework," Rapid 7, 2018.

[4] O. Sukwong and H. S. Kim, "Commercial Antivirus Software Effectiveness: An Empirical Study," *IEEE Computer Society,* pp. 63-70, 2011.

[5] N. Themelis, "A Tool for Antivirus Evasion: pyRAT," University of Piraeus, 2018.

[6] C. Truncer, "Veil 3.0 Release," 2017. [Online]. Available: https://www.veil-framework.com/veil-3-0-release/. [Accessed 12 July 2019].

[7] "AntiVirus Evasion Tool," 2019. [Online]. Available: https://github.com/govolution/avet. [Accessed 12 July 2019].

[8] M. Czumak, "peCloak.py – An Experiment in AV Evasion," 9 March 2015. [Online]. Available: https://www.securitysift.com/pecloak-py-an-experiment-in-av-evasion/.

[9] D. Cornacchini, "Phantom-Evasion," 8 December 2018. [Online]. Available: https://github.com/oddcod3/Phantom-Evasion.

[10] P. Ubuntu, "Venom," 25 February 2018. [Online]. Available: https://github.com/r00t-3xp10it/venom/tree/master/obfuscate/shellter.

[11] D. Kennedy, "Unicorn," 26 June 2019. [Online]. Available: https://github.com/trustedsec/unicorn.

[12] E. Balcı, "Hercules," 31 October 2017. [Online]. Available: https://github.com/EgeBalci/HERCULES.

[13] J. N. Rubenking, "The Best Free Antivirus Protection for 2019," 27 June 2019. [Online]. Available: https://www.pcmag.com/roundup/267984/the-best-free-antivirus-protection.

[14] A. Zacks, "10 Best Free Windows Antivirus Software for 2019," 12 December 2018. [Online]. Available: https://www.safetydetective.com/blog/best-really-free-antivirus-for-windows/.

[15] T. Fisher, " The 10 Best Free Antivirus Software of 2019," 8 July 2019. [Online]. Available: https://www.lifewire.com/best-free-antivirus-software-4151895.

[16] P. Wagenseil, "Best Free Antivirus Software 2019," 10 July 2019. [Online]. Available: https://www.tomsguide.com/us/best-free-antivirus,review-6003.html.

[17] J. Allen, "8 of the best free antivirus software options," 22 January 2019. [Online]. Available: https://mashable.com/roundup/best-free-antivirus/.

[18] K. G. Orphanides, "Best Free Antivirus 2019: 6 tried and tested ways to stay safe online," 25 June 2019. [Online]. Available: https://www.trustedreviews.com/best/best-free-antivirus-3633595.

Real Time Analysis of Telephone Conversation to Detect and Prevent Attacks in Nigeria

Aliyu Garba
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
garba.aliyu@aun.edu.ng

Ridwan Salahudeen
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
ridwan.salahudeen@aun.edu.ng

Sandip Rakshit
School of Information Technology & Computing
American University of Nigeria
Yola, Nigeria
sandip.rakshit.aun.edu.ng

***Abstract*— Nigeria is a nation with wide range of ethnic groups with a population of around 195 million. The nation has been having security difficulties, for example, Insurgency, Cattle rustling, Arm banditry, Political hooligans, Kidnapping, among others which prompts loss of lives and properties. The most happening and challenging are the kidnapping and insurgencies that spread into numerous parts (Borno, Adamawa, Yobe and so on) of the nation. It has been a very challenge to the security agents to find and continue to intercept any uprising before it happens. This proposal tends to address these difficulties by investigating telephone discussions through media transmission systems to anticipate the assault before it happens. The most commonly spoken languages (English, Hausa, Yoruba, and Igbo) are to be fed into a developed analyzer engine for processing furthermore, passes it to an inference engine for decision making. The inference engine passes the processed conversations as either threats or not. If found to be distrusted, a notification will be sent to the security agents for investigation and restraint**

Keywords*— *Security, Insurgency, Telecommunication, Conversation, Prevention

I. INTRODUCTION

In nowadays, the use of mobile phones have turned out to be so widespread, individuals use it consistently and night. How these telephones are being used stay significant worry in the general public. It is being used to carry out criminal offenses, for example, kidnapping, insurgency, arm banditry, and so on. The behaviour of human data grows exponentially in connection to the utilization of cell phones [1]. The enormous volume of data that emanate from telecom organizational records during each telephone call, ends up being amazingly hard to separate data that has to do with security threats. The tracking and analysing of telephone calls not in real time means playing the recorded transformations for which is practically infeasible. However, various methodologies can be applied to security issues because of the fast advancement and high capability of Natural Language Processing (NLP) [2], The tracking and analysing of phone call conversations can be accomplished in real-time with the use of Artificial Intelligence (AI) and its application, Machine Learning. Having achieved that, it helps security personnel to discover an attack and avert before it occurs

The NLP is a field in computing that guides the machine to derive what the client is attempting to state through his voice directions [3]. It is described as a computational procedure that analyses natural language units in particular degrees of phonetic analysis [4]. Most of the challenges in NLP are recorded as human language understanding, enabling computers to get significance from the voice directions given to it through human language, and others incorporate natural language communication among computers and people. By far most of the latest NLP algorithms depend on machine learning, especially statistical machine learning [3].

NLP gives the likelihood of human language to be broken down and subsequently extract commands or helpful information [5]. Uses of NLP are winding up increasingly extensive every day. Present cell phones are dispatched with some form of intelligent personal assistants like Google's Now and Apple's Siri [5]. Following that, a phone discussion can be processed, analysed, and induce important information from it without human mediation. As the amount of data streaming into and across over organizations becomes progressively tremendous, the issue is not just one of content dissemination, yet of the time it takes to exhaustively recognize and have experiences of what is important and consumable[6].

II. RELATED WORKS

NLP, Machine Learning Technique, and Artificial Intelligence (AI) become the most rise zone of research in the field of computing. The transcribed data extracted from a human spoken language is required to be processed. NLP is an algorithm developed primarily to handle human language for the computer to make sense and infer meaning out of it.

[7] stated the challenges in tracking and predicting the location of mobile phone call user due to irregular mobility patterns. In their research, experimental research reveals that the call patterns are greatly connected with co-locate patterns and the call patterns principally affect user short-time movement. Their research is mainly to approve the location prediction accuracy considering the social interplay unveiled in the mobile calls. This proposed a Critical Call Pattern (CCP) and Critical Call (CC) to know when the social interplay will influence user movement

[8] taken research where an analyst can trace and find the suspects by using call logs from suspects' phone numbers

and their contacts. All things considered, the suspects changed their phone numbers to keep away from tracking. The problem is that the investigators experience issues to trace these suspects from their call logs. Besides, their discoveries can improve human's recognition capacity to interface the connection between related telephone numbers regardless of the disguise by the suspects. This research concentrated on studying the conduct of the offender. If the telephone numbers are changed, the prediction model can foresee the conceivable new telephone number used. This has not considered the discovery of offender from the speeches made yet just that the offender must be known and on account of changing his/her telephone number, his/her conduct will be tracked. In this manner, the model can anticipate the likely telephone numbers changed for further investigation.

[1] filter and group client telephone calls into distinct patterns in real-time and analyses in an attempt to get significant information about the clients. This incorporates the online processing of data gathered from the telephone calls and essentially to identify callers' anomalies using an extensive algorithm of [9]. In [1] research, a basic distinguish of clients' calls was considered close by callers that are portrayed as exceptions (outliers). The finding can help shield from any coercion or spam calls that are not made by people, the unexpected dial of calls, among others.

The novel research by [5], attempt to gaze at in each business, regardless, its size and its last item seeks after a couple of techniques and plays out explicit activities with the ultimate objective to accomplish a result. Business Process Management (BPM) can be depicted as an organization order to accomplish such. The purpose of their research is to investigate how NLP can be associated in the BPM territory with the ultimate objective to subsequently make business procedure models from existing documentation inside an organization.

The research that is very close to this proposed work is the work of [8]. A client is track and finds to be suspect by utilizing call logs. If the client changes his telephone number, he/she can be traced considering his conduct from the past calls that are as of now put away as call logs. In any case, in this research, the discussion of at least two individuals is examined in real-time to identify if there is any security threat. On the off chance that there is a danger, our intelligence system ought to have the ability to identify the target area and when and at what time the attack is been talked about. Not only limited to that, as well as taking a gander at the sort of crime (arm banditry, kidnapping, cattle rustling, and so on.), and names referenced during the discussion.

III. METHOD

The speeches of at least two individuals having discussions through phone are translated into a digital form of data (text). The way that the voice is changed into text, we utilize the NLP and Artificial Intelligence to find from it whether there are security threats. The analyser engine is to be built which processes the text and passes it to deduction engine for decision making. The induction engine deduces based on the processed language as either the communicators are threats. Whenever observed to be suspected, a warning will be sent to the security agents to explore or potentially avoid. Figure 1 below demonstrates the well-ordered strategy for accomplishing every one of our targets in this research.

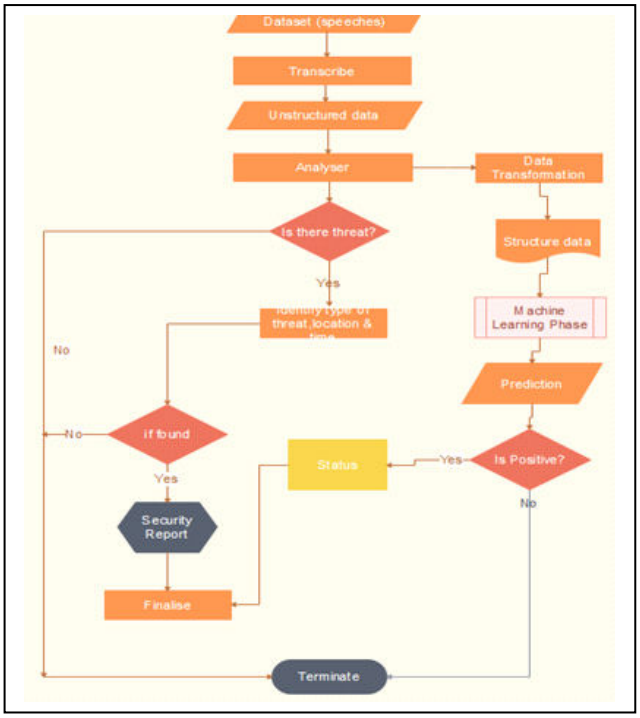


Figure 19: Process Flow

IV. PROPOSED SYSTEM MODEL

The proposed system model as delineated from figure 2 below comprises of six noteworthy parts, Citizens; where individuals do have conversation for different purposes, for example, education, business, criminal activities, among others, Analyser; which intercept the discussions, analyses, and translates into advanced unstructured data format, Extractor; this transform the unstructured data into structured data for processing, Machine learning; this is the place the training of the model with a structured data will be carried out and make possible prediction (either there is threat or not). Report; this is a segment that gets the consequence of the prediction alongside the subtleties (target area, time, kind of threat) if the security threat is found and Security agents; this is the place the warning of the subtleties will be sent to cell phones of the concerned security agents to take preventive measures before the event.

V. PROPOSED ALGORITHM/PSEUDOCODE

The efficiency of every computation is ability run under small amount space and small amount of time. Figure 3 shows a proposed step by step way of achieving said problem while figure 4 shows the proposed algorithm.

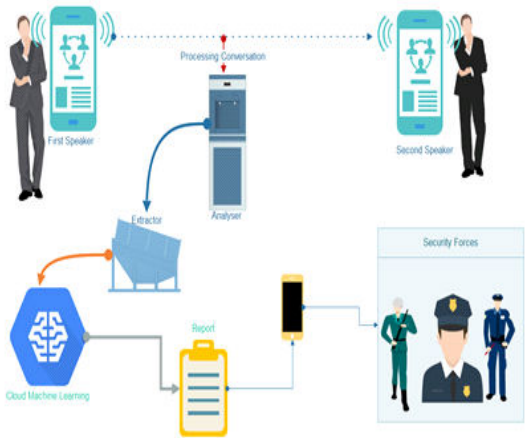


Figure 2: Proposed System Model

Step 1: track the conversation of the two speakers
Step 2: get the conversation
Step 3: while there is nothing said, to go step 4
    Transcribe the conversation and store into an array
    Get the next conversation
Step 4: pre-processed the unstructured transcribed data into structured data
Step 5: train the model with the structured data for prediction
Step 6: fit the model
Step 7: if predicted model is true then the there is a threat and to go step 8
Step 8: track the location, details of suspects, likely date and time
Step 9: send SMS notification to security forces concerned, along the details captured
Step 10: end of the tracking

Figure 3: Proposed Algorithm

```
// Speaker1 and speaker2 are the two communicators
track(speaker1, speaker2)
// output: criminal activities, target location, time, date, target individual, suspect identity,

UnstructuredData = ArrayList()

speech = getSpeech(speaker1, speaker2)

While(speech != null)
    unstructuredData.add(transcribe(speech))
    speech = getSpeech(speeker1, speeker2)
structuredData  = preprocess(unstructuredData)
model = train_model(structuredData) // using deep learning
predicted = model.fit()
if predicted = true // there is threet
    location = trackLocation()
    target = namesMentioned()// names of targeted people, if any?
    suspect = indentifySuspect() //track the suspected people using SIM registration
    likelyDate = dateMentioned()
    likelyTime = timeMentioned()
    sendSMSAlert(suspect, location, likelyDate, likeTime, target) // to security forces concerned
```

Figure 3: Proposed Pseudocode

VI. CONCLUSION

Every day, press reports instances of people losing their lives because of uncertainty which escalated virtually everywhere in the country. Even though each part has its exceptional issue of insecurity in Nigeria, in the north-eastern part, it is insurgency, while in the south-eastern part, it is pipeline vandalism and kidnapping, etc. Moreover, essentially wheresoever now in the country is bedevilled by kidnapping. Native and non-citizen when kidnapped and are forced to pay a ransom. Hoodlums plot through communication which for the most part through telecom, especially mobile phone before they penetrate a spot for an attack. Ability to detect and anticipate this attack, will help the advancement of the nation as far as the social prosperity of the people and consequently in the development of the economy of the country. Local and foreign investors have a dread of investing because of uncertainty in the country. This work tends to address these issues that escalate into the nation through cell phone discussions. Although, the residents are wherever howling of the absence of security, however, continue underlining of user privacy. The national security issues supplant individual interest and therefore, ought not to be undermined. Having many the most populous languages in Nigeria, and the tricky in understanding, this research focuses on one of the natively spoken dialects but, Hausa. However, this is not to detect or recognize the voice of the speakers but only to analyse security threats. This research is very crucial in leading to the realization of all the aforementioned vital development.

REFERENCES

J. Iglesias, A. Ledezma, A. Sanchis, and P. Angelov, “Real-Time Recognition of Calling Pattern and Behaviour of Mobile Phone Users through Anomaly Detection and Dynamically-Evolving Clustering,” *Appl. Sci.*, vol. 7, no. 8, p. 798, Aug. 2017.

K. Steve, “Artificial Intelligence Techniques for Security Vulnerability Prevention,” p. 8, Apr. 2018.

P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar, and S. Kumar, “Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT),” in *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)*, Chennai, India, 2017, pp. 368–373.

S. A. Al-Ghamdi, J. Khabti, and H. S. Al-Khalifa, “Exploring NLP web APIs for building Arabic systems,” in *2017 Twelfth International Conference on Digital Information Management (ICDIM)*, Fukuoka, 2017, pp. 175–178.

K. Sintoris and K. Vergidis, “Extracting Business Process Models Using Natural Language Processing (NLP) Techniques,” in *2017 IEEE 19th Conference on Business Informatics (CBI)*, Thessaloniki, Greece, 2017, pp. 135–139.

G. Jessica, “Artificial Intelligence Impacts Every Business Function | Tractica,” 13-Sep-2017. .

D. Zhang, A. V. Vasilakos, and H. Xiong, “Predicting location using mobile phone calls,” ACM, p. 2, Sep. 2012.

Y. Longtong and L. Narupiyakul, “Suspect tracking based on call logs analysis and visualization,” in *2016 International Computer Science and Engineering Conference (ICSEC)*, 2016, pp. 1–6.

P. P. Angelov and D. P. Filev, “An Approach to Online Identification of Takagi-Sugeno Fuzzy Models,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 34, no. 1, pp. 484–498, Feb. 2004

Design and Implementation of a Solar Powered Security/Garden Light

Iyere, S. F*
Department of Elect/ Elect Engineering
Ambrose Alli Univerisity
Ekpoma, Nigeria
sunday4success2007@gmail.com

Yeboah, J. K.
Department of Elect/ Elect Engineering
Ambrose Alli Univerisity
Ekpoma, Nigeria
josephyeboah1947@gmail.com

Erua, J. B.
Department of Elect/ Elect Engineering
Ambrose Alli Univerisity
Ekpoma, Nigeria
eruajb@yahoo.com

John-Otumu, A. M.
Training, Research & Innovations Unit, Directorate of ICT
Ambrose Alli Univerisity
Ekpoma, Nigeria
macgregor.otumu@gmail.com

***Abstract*— The Faculty of Engineering and Technology in Ambrose Alli university, Ekpoma has been facing the problem of lack of security lights. This has resulted into series of burglary attempts on the laboratories where sensitive equipment and other valuable items for teaching and research are kept, in addition to the inability of students to move freely about their studies and other activities at nights. This research work discusses the design and implementation of a security lighting system using renewable energy such as solar to illuminate all the courtyards in the faculty. A central solar power supply was installed to power a total of twenty 36W12/24V LED lamps which are positioned at various points in the courtyards. The power supply consists of 130W/17V solar panels, an 80A MPPT charge controller and a bank of 200AH/12V batteries. The solar arrays and the battery bank are both connected in hybrid form. A series of test were carried out including voltage drop across the loads charging tests at day time as well as load test at nights. A dusk-to-dawn D.C power supply was also installed to switch the loads (lamps) ON at about 6:45 pm in the night, and switch them OFF at about 6 30am in the morning. In order to cushion the effect of power losses particularly during the period of poor solar radiation, an AC to DC power supply was also designed and incorporated into the power supply unit. An automatic change-over switch was built for the purpose of selecting the required power supply to the loads. Results from various tests indicate that all the lamps can be effectively powered and controlled from a single source of power throughout the night.**

Keywords*— *Solar, Photovoltaic Energy, Power, Voltage Current, Lighting Battery

I. INTRODUCTION

This research work highlights the use of solar energy in providing a centrally controlled security lighting at the faculty of Engineering and Technology in Ambrose Alli University, Ekpoma, Edo state.

The idea behind the whole work is a deviation from the usual method of installation of solar street/security lighting, where each of the lighting points or lamps (as the case may be) carries its own power supply as a stand-alone unit comprising; solar module, battery bank and a dusk-to-dawn solar charge controller. Other methods of central solar street lighting may involve the use of an inverter to power A.C lamps.

In this research work, a central solar power supply unit was designed and assembled in conjunction with an A C-to-D C power supply built for inter-change of power to the loads whenever power from the public supply is available. The load comprises of twenty LED lamps rated 36W each. An automatic change-over switch was constructed to switch power between the two sources.

Solar powered security lighting is now an area of interest in research due to the increasing energy demand for the growing population of the world. Recent application shows that most of the common High Intensity Discharge (HID) lamps, often High Pressure Sodium (HPS) lamps are being replaced by more low powered Light Emitting Diode (LED) lamps [1] – [2].

According to Ramadhani, Bakar and Shafer [3], the basic components of a solar powered LED security lighting system are:

i. Solar energy
ii. Solar panel
iii. Lighting fixtures – LED lamp
iv. Rechargeable battery
v. Charge controller
vi. Pole

1.1 Photovoltaic Energy (Solar Energy)

Solar energy from the sun is harnessed and converted to electrical energy in two ways, namely;

i. Direct conversion into electricity that takes place in semiconductor devices called *solar cells.*
ii. Accumulation of heat in *solar collectors*.

Solar cells are different from solar collectors. The direct conversion of solar radiation into electricity is often described as *photovoltaic (PV)* energy conversion because it is based on the *photovoltaic effect.* In general, the photovoltaic effect means the generation of a potential difference at the junction of two different materials in response to visible or other radiation. The whole process of solar energy conversion into electricity is denoted as

"*photovoltaic"*. Photovoltaic literally means "light-electricity", because "photo" is a term from Greek word "phos" meaning light and "volt" is an abbreviation of Alessandro Volta's (1745-1827) name that pioneered the study of electricity [4].

The motives behind the development and application of the PV solar energy were in general the same as for all renewable energy sources. The motives were based on the protection of climate and environment and providing clean energy for all people. The current motives can be divided into three categories: energy, ecology and economy [5] – [6].

1.2 Energy

There is a growing need for energy in the world and since the conventional energy sources based on fossil fuels are exhaustible with time, PV energy is considered a prominent alternative energy source. Large-scale application of PV solar energy will also contribute immensely to the broad use of energy sources [7].

1.3 Ecology

Large-scale use of PV solar energy, which is considered an environmentally friendly source of energy, can lead to a substantial decrease in the emission of poisonous gases such as CO_2, SO_x and NO_x that pollute the atmosphere during burning of the fossil fuels. When we closely look at the contribution of the PV solar energy to the total energy production in the world we see that the PV solar energy contribution is only a tiny part of the total energy production. At present, the total energy production is estimated to be 1.6×10^{10} kW compared to 1.0×10^{6} kWp that can be delivered by all solar cells installed worldwide. By W_p (Watt peak) we understand a unit of power that is delivered by a solar cell under standard illumination. When PV starts to make a substantial contribution to the energy production and consequently to the decrease in gas emissions depends on the growth rate of the PV solar energy production. When the annual growth of PV solar energy production is 15% then in year 2050 solar cells will produce 2.0×10^{8} kW_p. The annual growth of 25% will result in the solar electricity power production of $2.4 \times 10^{10} kW_p$ in 2030. This demonstrates that there must be a steady growth in solar cells production so that PV solar energy becomes a significant energy source after a period of 30 years [8].

1.4 Economy

Solar cells and solar panels are readily available in the local market. An advantage of the PV solar energy is that the solar panels are modular and can be combined and connected together in such a way that they deliver exactly the required power in a systems referred to as "custom-made" energy.

The reliability and maintenance costs, as well as modularity and expandability, are some of the enormous advantages of PV solar energy in rural applications. There are two billion people in mostly rural parts of the world who have no access to electricity and solar electricity today is the most cost-effective solution. Bringing solar electricity to these people represents an enormous market. Some companies and people have realized that solar electricity can make money already and this fact is probably the real driving force to a wide spread development and deployment of PV solar energy. The advantages and drawbacks of the PV solar energy, as seen today and are summarized as follows [5].

Advantages

(a) Environmentally friendly
(b) No noise, no moving parts
(c) No emissions
(d) No use of fuels and water
(e) Minimal maintenance requirements
(f) Long lifetime, up to 30 years
(g) Electricity is generated wherever there is light, solar or artificial
(h) PV operates even in cloudy weather conditions
(i) Modular or "custom-made" energy can be designed for any application from small to multi-megawatt power plant

Drawbacks

(a) PV cannot operate without light
(b) High initial cost that overshadow the low maintenance costs and lack of fuel costs
(c) Large area needed for large scale applications
(d) PV generates direct current: special DC appliances or inverter are needed in off-grid application; also energy storage is needed such as batteries.

1.5 Photovoltaic (PV) System

The conversion of solar energy into electrical energy is done by a semiconductor device which forms the solar cell. A solar cell is a unit that delivers a certain amount of electrical power characterized by an output voltage and current. In order to use solar energy to power devices, which require a particular voltage or current for their operation, a number of solar cells are connected together to form a *solar panel,* also called a *PV module*. For large-scale generation of electrical power the solar panels are connected to form what is known as *solar array* [4].

The solar panels are part of a complete *PV solar system,* which, depending on the application, comprises batteries for electricity storage, DC/AC inverters that connect a PV solar system to the electrical grid, and other miscellaneous electrical components or mounting elements. These additional parts of the PV solar system form a second part of the system that is called *balance of system (BOS)*. Also, solar power system may contain loads such as household appliances; radio, TV set and lamp(s) that can only operate on D.C voltage [4].

We refer to these products as a load. In summary, the PV solar system consists of three parts namely;

i. Solar panel or solar array
ii. Balance of system
iii. Load.

1.6 Solar Panel

The solar panel provides electrical energy to charge the battery during the day time. The battery's charging is controlled by a charge controller. The operation of the LED

bulb is controlled by a control circuit either by using sensors such as Light Dependent Resistor (LDR). The maximum power point (MPP) is the spot near the knee of the V-I curve, and the voltage and current at the MPP are designated as V_m and I_m. For a particular load, the maximum point is varying following insolation, shading and temperature. It is important to operate panels at their maximum power condition [8].

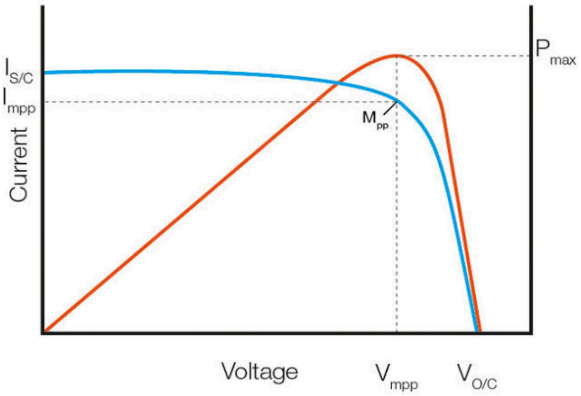


Fig. 1: V-I curve of a solar panel [9]

1.6.1 Selection of Solar Panel

The solar panel is the core part of the solar street light system which converts the sun's radiant energy to electrical energy, and then transmits through the controller to be stored in the battery.

1.6.2 Working Principle

The Photovoltaic (PV) cell is composed of at least two layers of the semiconductors, "doped" with different impurities. This makes an excess of free electrons (n-type) on one side of the junction, and a lack of free electrons (p-type) on another side.

When the photovoltaic cells are irradiated with sunlight, some photons are reflected and the others are absorbed by the solar cell. When the photovoltaic cells keep enough photons, the negative electrons are released from the semiconductor material. Due to the manufacturing process of the positive layer, these free electrons naturally migrate to the positive layer which creates voltage differential. When the solar cell is connected with the external load, there will be a current circulation in the circuit. Each single solar energy cell produces only 1-2 watts. In order to increase output power, these cells (from one to several thousands) are connected in series or in parallel with others, what is called a solar array.

1.6.3 V-I Characteristics of a Solar Cell

Fig. 1 shows the V-I characteristic curve and output power of a solar panel. The curve has two parts; one indicates the trend of current with respect to increasing voltage. The other curve is the power-voltage curve and is obtained by the equation P = V x I. If no load is connected with the solar panel which is working in sun light, an open-circuit voltage Voc will be produced but no current follows. If the terminals of the solar panel are shorted together, the short circuit current will flow but the output voltage will be zero.

In both cases, when a load is connected, we need to consider V-I curve of the panel and V-I curve of the load to find out how much power can be transmitted to the load.

In sunny areas, the polycrystalline silicon solar cell is more appropriate, because the price of the polycrystalline Silicon solar cells is lower than the Mono-crystalline cells. But in the more rainy days in which sunlight is relatively not very adequate. It is better to choose the Mono-crystalline silicon solar cells, because the optical conversion efficiency of the mono-crystalline silicon solar cells is higher [5] – [6].

1.6.4 Solar Panel

A Solar Panel is basically a module that converts light energy (photons) from the sun to generate electricity in DC form. There are two types of solar panels, mainly crystalline and thin-film types. There are basically two types of solar panels (See Fig. 2):

i. The Crystalline
ii. The Thin-film

The crystalline type of solar panels has two varieties; namely;

i. Poly-crystalline solar panel
ii. Mono-crystalline solar panel

While the thin-film type are about four varieties:

i. Amorphous silicon (a-Si)
ii. Cadmium Telluride (Cd-Te)
iii. Coppér Indium Gallium Selenide (CIGS)
iv. Dye-Sensitized Solar Cell (DSC).

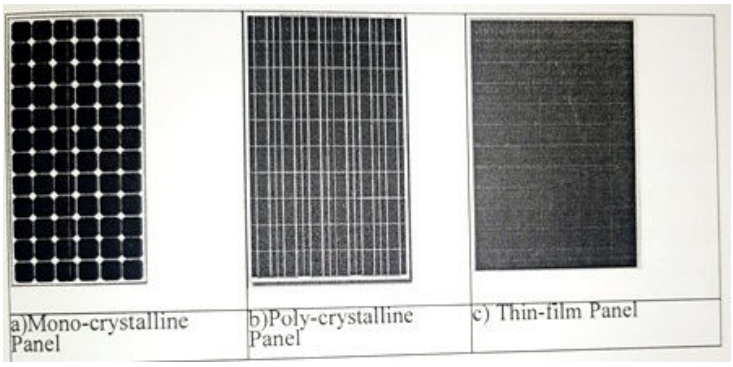


Fig. 2: types of solar panels [9]

The crystalline based solar panels are commonly used due to its maturity in terms of its technology and price. Even though poly-crystalline solar panels are cheaper compared to Mono-crystalline solar panels, Mono-crystalline panels are preferred in Street lights applications because it is smaller compared to poly-crystalline due to its higher efficiency, making the design for the pole to be easier and cheaper too.

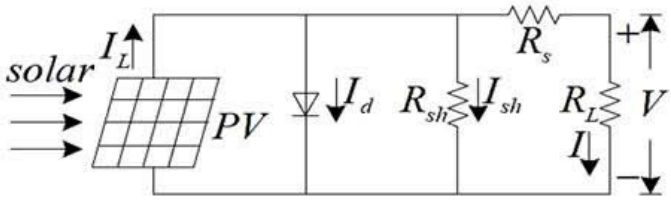


Fig. 3: Equivalent circuit of solar panel

Fig 3 shows the equivalent circuit of a solar panel, where it shows the diode and ground leakage currents.

1.7 Battery

Batteries are used to store the electricity generated by the solar panel, during the day, electricity generated by the solar panels is supplied to the battery and/or the load. When the load demand is higher than the energy received from the solar panels, these batteries will provide stable energy to the load [10].

There are a few types of rechargeable batteries, which are:

1.7.1 Lead-Acid (LA) Battery

These batteries are the most commonly used in solar powered systems due to its maturity in technology and low pricing. They can only be used with low Depth of Discharge (DOD) in order to extend its lifespan. Its DOD ranges from 25% to-50%. There are two types of Lead-Acid batteries, i.e. flooded and Valve Regulated Lead Acid (VRLA) batteries which are maintenance free batteries

1.8 Charge Controller

Charge controllers are used to control the charging of the batteries. Since the output from the solar panels are variable and needs adjustments, charge controller fetches the variable voltage/current from solar panels, condition it to suit the safety of the batteries. The main functions of charge controllers are to prevent over-charging of batteries from solar panels, over-discharging of batteries to the load and to control the functionalities of the load.

Charge controllers are basically DC-DC converters, where PWM technique (preferred scheme) is used to regulate the switches of the controller. There are three general types of charge controllers, mainly:

1. Simple ON/OFF controller
2. Pulse Width Modulated (PWM) controller
3. Maximum Power Point Tracking (MPPT) controller

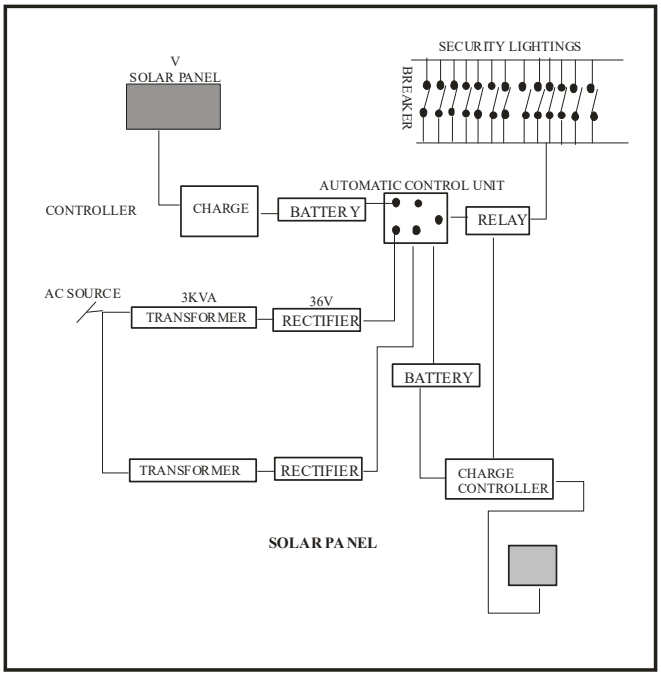


Fig. 4: Block diagram representation of the system

II. METHODOLOGY

In this section, the authors performed the following activities in order to achieve our desired goal.

We reviewed the solar security and garden lighting using existing information, articles and literatures. We identified strategic positions to install the lamps and determined the actual number of lighting points necessary to provide enough illumination in the entire faculty. The layout is shown in Fig. 2 and the satellite view in Fig. 3.

We estimated the total load power consumption by all the lamps and designed a suitable power supply that can effectively power the loads.

We also determined the open market specifications and prices of power supply and load components in view of the design.

Finally, we obtained and mounted / installed the power supply and load components.

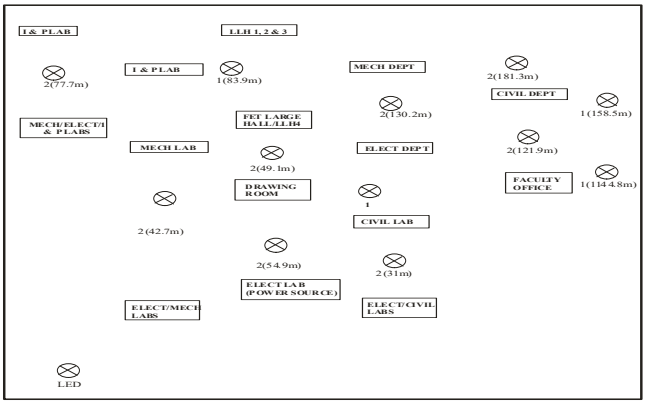


Fig. 5: Lighting Point Layout in the Faculty of Engineering and Technology

Fig. 6: Satellite map of Faculty of Engineering and Technology showing the lighting points

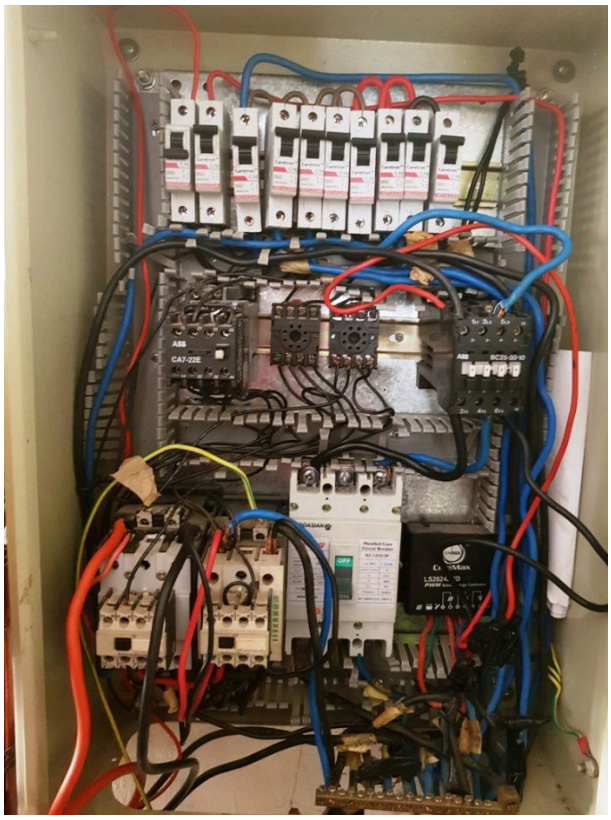


Fig 7: Interior view of the completed circuit arrangement of the control panel

Fig 9: Photo view of the 220V A.C to 24V D.C Circuit

Fig 8: Battery bank and A.C-to-D.C power supply

Fig. 10: Photo view of the Charge Controller Assembly

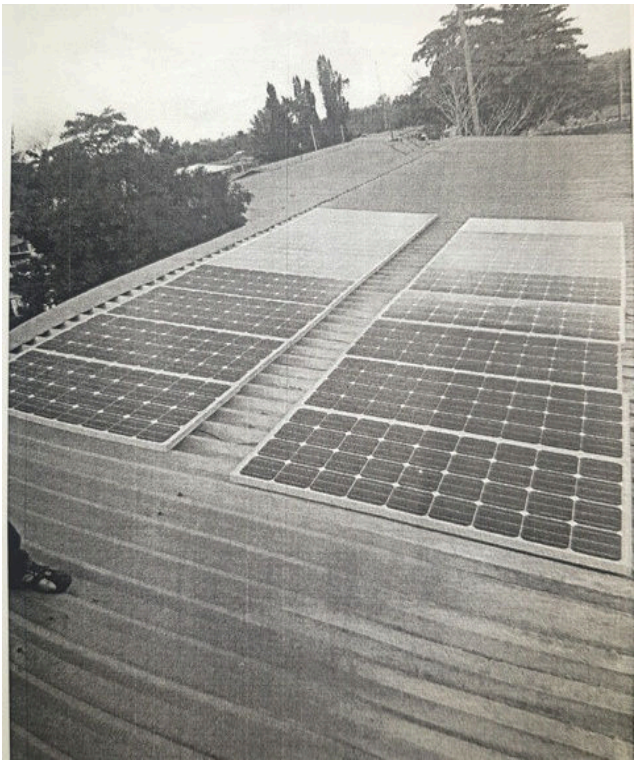


Fig.11: Photo view of the Solar Array

2.1 Design Specifications

The circuit design incorporates several parameters and circuits and their specifications are;

i. Rating of individual lamps ---------- 36W
ii. Number of lamps ---------- 20
iii. Battery bank voltage ---------- 24V
iv. Number of hours lamps would be ON ----- 12 hrs
v. Selected battery voltage ---------- 12V
vi. Selected battery capacity (AH) ---------- 200AH
vii. Rating of transformer for rectifier circuit -- 3KVA
viii. Voltage ratio of transformer for rectifier circuit- 240:24

2.2 Load List

Total load demand = 36 x 20 = load demand
= 720 Watts

Total load current =

$$\frac{\text{total lamps capacity}}{\text{battery bank voltage}} - \text{load demand}$$

$$= \frac{720\text{watts}}{24\text{v}} = 3\text{A}$$

2.3 Sizing of the Battery Bank

The total energy consumed by lamps per day = total lamps capacity x hours lamps would be ON
= 720 x 12 = 8640Wh/day

Ampere-hours consumed by the load In a day (LAH/day)

$$= \frac{\text{total energy consumed by lamps per day}}{\text{battery bank voltage}}$$

$$= \frac{9640\text{wh}}{24} = 360\text{AH/day}$$

Maximum depth of discharge (M_{DD}) = 50%
Days of autonomy (D_{AT}) = 1 day
Derate factor (DF) = 1 at 25°c

$$\frac{\text{LAH/day x DST}}{\text{MDD x DF}} = \frac{360 \times 1}{0.5 \times 1} = 720\text{AH}$$

Batteries in parallel (Bp) =

$$\frac{\text{required battery bank capacity}}{\text{selected batery capacity}} = \frac{720}{200} = 3.6 \approx 4$$

Batteries in series (Bs) = $\frac{\text{battery bank voltage}}{\text{selected battery voltage}} = \frac{24}{12} = 2$

Total number of batteries (N_B) = B_s x B_p = 2 x 4 = 8

2.4 Sizing of Photo-Voltaic System Array

i. Total Amp hours per day = 360 AH/day
ii. Average sum hours = 4.2 hours
iii. System efficiency = 70%
iv. Nominal module voltage = 17V
v. Nominal module capacity = 130W

Total array current required = $\frac{\text{total Amp hour per day}}{\text{average sun hour}} = \frac{360}{4.2}$

= 85.714A

Peak current produced by selected module =

$$\frac{\text{selected module size (W)}}{\text{norminal module voltage (V)}} = \frac{130}{17} = 7.647\text{A}$$

Total number of modules required =

$$= \frac{\text{total array of current}}{\text{peak current of module x efficiency}}$$

$$= \frac{85.714}{7.647 \times 0.7} = 16.02 \approx 16$$

Number of modules in each series string =

$$= \frac{\text{battery bank voltage}}{\text{normal module voltage}}$$

$$= \frac{24}{17} = 1.41 \approx 2$$

Number of parallel strings =

$$= \frac{\text{total number of modules}}{\text{number of modules per series strings}} = \frac{16}{2} = 8$$

2.5 Sizing of Charge Controller

Minimum current capacity of controller;

= Total number of parallel strings x Module short circuit current x Safety factor (1.25)

= 8 x 8 x 1.25

=80A

III. RESULTS AND DISCUSSION

This section reveals and discusses the various results obtained in tabular and graphical forms. Generally, performance test is to evaluate the performance of a system ranging from the services it can render, its usability, durability, life span, and other important parameters.

The performance tests carried out on the solar system security lighting were as follows:

(a) Solar battery performance test.
(b) Open circuit test.
(c) Load voltage of each lighting point i.e. load voltage test at the base of the lighting pole.
(d) Testing of each lamp for functionality before mounting.
(e) Rate discharge of battery on load over a period of approximately 12 hours i.e. at night.

3.1 Test Result Tables and Graphs

TABLE 1: open circuit test results:

Circuit breaker number	Location of lighting point	No load voltage (V)	Current (A)	Voltage on load (V)
	Open circuit voltage	21.56		
1	M & P Lab	19.8	2.64	11.0
2	Civil dept	21.15	1.75	19.47
3	Behind Civil	21.23	2.0	15.61
4	Car park	21.20	1.77	19.35
5	Electrical dept	21.20	3.45	18.10
6	Mechanical lab	21.21	3.03	19.3
7	Drawing room	21.18	2.18	16.31
8	Electrical lab	21.37	3.6	18.5
9	Civil lab	21.20	2.7	18.70

TABLE 2: Load characteristics

Circuit breaker number	Location and distance from battery bank (m)	Load voltage (V)	Load current (A)
1	P & P lab 77.7	19.8	6.24
2	Civil dept. 121.9	21.12	3.9
3	Behind civil 181.3 Car park 144.8 & 158.5	21.23 21.20	6.54 1.77
4	Elect dept. 130.2	21.20	3.45
5	Mech. Lab 42.7	21.21	6.03
6	Drawing room 49.1	21.18	6.83
7	Electrical lab 54.9	21.57	3.6
8	Civil lab 61.0	21.20	7.37

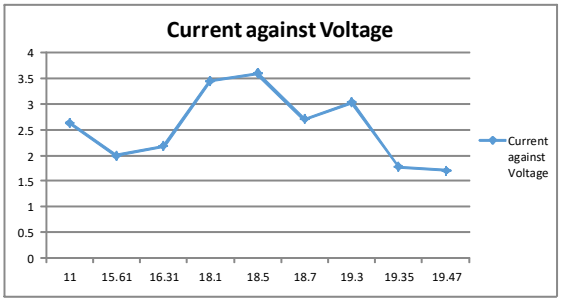


Fig. 12: Graph of load voltage versus load current

Table 2 shows the shape of the graph indicates the characteristics of load current variation with respect to load voltage.

Table 3 is the results of load voltage and load current test based on location and distance of the poles from the source of supply.

TABLE 3: Lighting point test of the circuit and load voltage results

Days	Voltage before switch on (V)	Voltage after switch on (V)	Morning voltage (V)
1	28.5	23.7	19.0
2	28.0	23.7	18.2
3	27.5	23.7	18.0
4	28.0	23.7	18.2
5	28.0	23.7	18.3
6	27.5	23.7	18.1
7	28.5	23.7	18.3
8	27.0	23.7	18.0
9	28.0	23.7	19.0
10	28.5	23.7	18.0
11	27.5	23.7	18.3
12	28.0	23.7	18.1
13	28.5	23.7	18.2
14	28.5	23.7	18.0

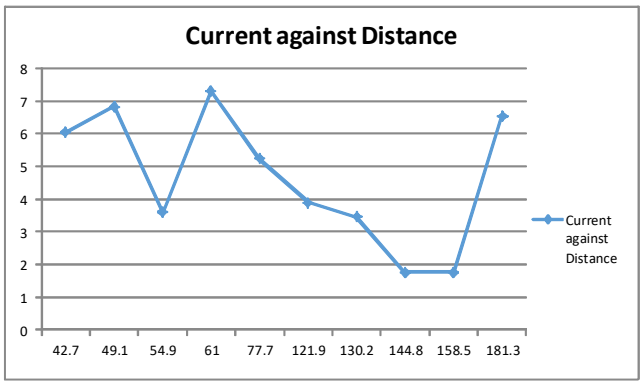


Fig.13: Graph of load current versus distance

The shape of the graph in Fig 13 indicates the characteristics of load current variation with respect to location and distance of pole from the source of supply.

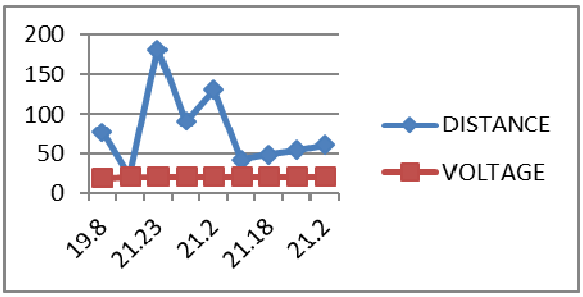


Fig 14: Graph of load voltage versus distance

The shape of the graph in Fig 14 indicates the characteristics of load voltage variation with respect to location and distance of pole from the source of supply.

IV. CONCLUSION

A central solar powered security lighting, different from the stand-alone method has been designed and implemented including the built-in 24VD.C power supply in this research work. Results from various tests indicate that LED-based security/street lighting can be energized and controlled from a central source. This method will help in reducing the installation cost for solar security/street lighting. And because of the change-over between solar power supply and the AC-to-DC power supply, the batteries will last longer with an improved state of charge characteristics.

.

REFERENCES

[1]. Costa, M. A. D., Costa, G. H, Santos, A., Schuch, L., and Pinheiro, J. R. (2009). "A High Efficiency Autonomous Street Lighting System Based on Solar Energy and LEDs". IEEE Transactions on Industrial Electronics. pp265-273.

[2]. Faith M., and Chikouche A. (2010), "LEDs Application to the Photovoltaic Street Lighting. pp. 101-104.

[3]. Ramadhani, F., Bakar, K. A., and Shafer, M. G. (2013), "Optimization of Standalone Street Light System with Consideration of Lighting.

[4]. Jonathan, G.,"A brief History of Solar Power."[0nline], available: http://ezinearticles.com, 2010.

[5]. Ogunleye, I. O., Awogbemi, O. (2010), "Constraints to the use of solar photovoltaic as a sustainable power source in Nigeria." American Journal of Scientific and Industrial Research, 2(1): 11-16.

[6]. Owolabi I. E. (2008), "Solar Photovoltaic power technology: Prospects and challenges for socioeconomic empowerment in Nigeria." Technical Paper presented to the Nigerian Society of Engineers (NSE) Ado-Ekiti branch. Pg 3-5.

[7]. IEA, 2012, World Energy Outlook 2012, International Energy Agency, Paris

[8]. Nunoo, S., Attachie, J. C., and Abraham, C. K. (2010), "Using Solar Power as an Alternative Source of Electrical Energy for Street Lighting in Ghana". IEEE Conference on Innovative Technologies for an Ejj'icient and Reliable Electricity Supply (CITRES). pp 467-471.

[9]. JHA, A.R. (2010), "Solar Cell Technology and Applications." U.S, CRC Press.

[10]. Kiehne, H. A. (2005). "Battery Technology Handbook", 2nd Edition, Germany, a. Merkel Dekker Inc.

Big Data Analytic: Strategic Tool For Counter Insurgency

Sagir Lawan*, Ashraf A. Ahmad, Fatai O. Adunola
Dept of Electrical/Electronic Engineering,
Faculty of Engineering and Technology,
Nigerian Defence Academy (NDA),
Kaduna, Nigeria.
*slawal@nda.edu.ng

***Abstract*— The problem of insurgency and terrorism in the world is a growing phenomenon especially in the face of emerging information and communication technology (ICT). This ICT phenomenon has brought the development of smart communication devices and cloud computing generation huge volume of data with devastating consequences. The enormous use of ICT by insurgents and terrorists caused a worldwide buzz for Big Data. This paper investigated the strategic employment of Big Data Analytic for not only profitable counter insurgency operation, but also for Internet-based information warfare surveillance, and sensor-based military monitoring. Additionally, review of the latest strategic technologies of gathering, sharing and storage of Big Data using the state-of-the-art analytical methods was studied. Finally, the paper discussed the future perspectives of counter insurgency in the era of Big Data.**

***Index Terms*— Big Data Analytic, information and communication technology (ICT), Strategic Counter Insurgency**

I. INTRODUCTION

In the present 'Information Age', computer networks, smart communication devices and cloud computing play an important role in all aspect of modern warfare. The emergency of information and Communication Technology (ICT) have provide military and terrorists with a vast array of new communication, surveillance and platform sensor capabilities [1]. For this reason, ICT is an acceptable veritable instrument for creation of large pool of data. The phenomenon of describing these very large datasets is known today as Big Data [2]. Growing Big Data has made combat more strategic thereby increasing the need for employing Memetic Warfare in counter insurgency operation.

Memetic Warfare is a modern type of warfare for which insurgents/terrorist spread information from person to person and uses psychological warfare to carry out their operations. It involves the propagation of memes on social media through platform weaponization. The 9/11 terrorist attack on the US highlighted the increase danger of Memetic warfare conducted by insurgents [3, 4, 5]. Insurgency can have many forms; it can be religious, economic or even cultural. Unfortunately, nowadays all the types of insurgency employ Internet based Big Data for their criminal activities. So commanders that use Big Data Analytic technology will doubtless prosper in the ongoing terrorist asymmetric warfare. Thus, military Big Data Analytics borders on the process of seeking insights and examining large data for crime prevention and enhancing security agencies capacity [4].

Irrespective of huge growth in storage capacity, Big Data is arguably one of the most important tools insurgents/terrorists use for operational plans, store institutional knowledge and develop strategy. Hence, the saying that, 'No terrorist activities can develop beyond its ICT Big Data system' [4]. Big Data strategy in counter insurgency operation has a clear connection with the proliferation of uncompromising information from Internet. Consequently, the product of Big Data analysis can transform modern warfare. The quality of these analysis and network centric infrastructure is the core function of find, fix strike the adversary in digital battle field environment. According to Burton, "The enemy will conceal himself to resist the effects of our weapons and conceal his plans, and therefore he must be found. He will resist destruction, and attempt to damage or defeat us, and therefore he must be found [6].

II. NATURE AND CHARACTER OF INSURGENTS/TERRORIST

The fundamental nature of insurgent's activities does not change; they use indiscriminate violence to spread fear in a bid to achieve political, religious, ideological or financial aims. It is adversarial, human and political [7, 8]. Yet each insurgent/terrorist has a different character. These show that the management of insurgents and terrorist activities is complex and problematic because insurgent hardly adapt conventional tactics.

Since 1970, there have been at least 4,908 terror-related deaths in Nigeria, most of these occurred by Boko Haram in

North East [9]. The global terrorist database shows that the numbers of terror related deaths have risen sharply worldwide. Therefore, the most pressing security challenge in countering terrorism and insurgency is gathering sufficient intelligence to combat the criminal activities. The military is trained to combat such situations but without proper combat intelligent data, there is not much the military can do in this asymmetric way of fighting.

While combating the insurgency, the military has suffered significant losses of personnel and platforms as a result of lack of real time data intelligence. Over time, the insecurity Africa face is a reflection of intelligence gather from Big Data by the criminal elements of the society [10]. For example, Nigeria has been fighting Boko Haram and Niger Delta terrorists for more than a decade and these insurgent are innovative, ruthless and persistent in using ICT modern communication and warfare [11].

In response to the evolving adversary, the Armed Force of Nigerian (AFN) has developed numerous counterinsurgency mechanisms to ensure the creation of a framework to view volume of Big Data generate from digital battle operational environment (OE). Big Data can provide new methods of applying all the capabilities at national defence disposal to dislodge the terrorist and insurgents. While counter insurgency operation provides a comprehensive strategic framework to use in defeating insurgents, most of the national defence of developing countries lacks any substantive analysis concerning the value of Big Data gather from cyberspace. Therefore, developing and under developed countries need to fully understand the modern network centric warfare that insurgent/terrorist employed in their operational and tactical levels. Moreover, today's military must be adaptable to constantly changing operational situation and condition brought by Big Data and Internet based network battlefield.

III. INTERNET BASED INFORMATION WARFARE AND BIG DATA

Over the last half-decade, with the emergence of Internet based information warfare, network centric warfare (NCW) battlefield rose to prominence [12]. Computer network technology has brought revolutionary changes to Modern Warfare. Computer network is the product of information age. The Internet has also become more familiar tool with insurgent and terrorist combat environment. For instant, the wars in Iraq and Afghanistan, witnessed a fundamental shift from voice to data transmission [13]. Data communication has long been an inseparable feature of counter insurgency operation and unavoidable experience for both military and insurgents.

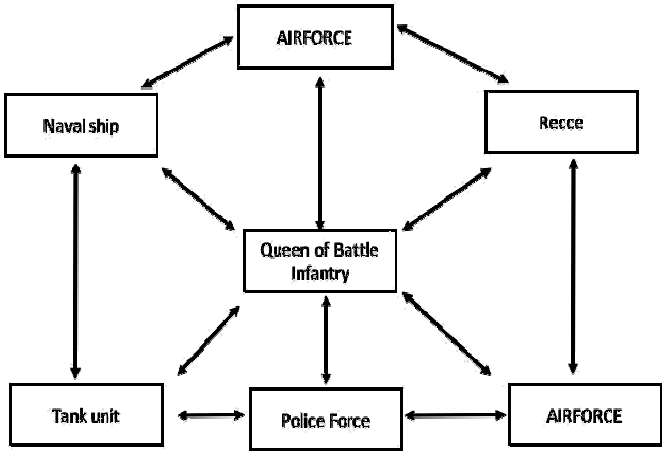
Insurgents/terrorists need to share their interests in an open, secure and stable cyberspace to maximize the economic, social and cultural aspects of the Internet. Computer networking at a stage must be purposeful, interesting, and it is the most effective mode of combat to bring its potential and effectiveness into full play. The tempo of acquiring intelligence from Internet must be carefully adjusted so that information ammunition, information warfare platform, and individual digital equipment are efficiently utilized [14]. The information source equipment have strong ability of information processing and exchange. Hence, obtaining information on rapid network response, accurate attack and high-efficiency integrated combat capability can be achieved.

Network centric warfare are endemic in digital information age whether in an undeveloped, developing or developed countries, insurgents and terrorist have embrace the use of network centric infrastructure. As a result, modern warfare strategy defends on network grid infrastructure and cyberspace technology. In spite of this development, the essence of traditional counter insurgency operation has never disappeared. Yet, amid the complexities of ICT technology, NCW and cyberspace have assumed an explosive dimension in insurgency operation. For example, in Nigeria, Boko Haram insurgents establish service by exploring growth of internet, intranet and extranets networks. These technologies combined with high value, high speed data access network led to the recent resurgence of the Boko Haram [15].

In as much as information technology provides the platform for NCW growth through globalization, there are few negative effects that are more prevalent in NCW and cyberspace technology. There are some disruptive technologies capabilities that were not available and unimaginable in the past decade. For example, the television broadcast is changing from traditional baseband based on serial interface to an internet protocol (IP)-based infrastructure. The private nature of the IP's physical layer as part of unmanned aerial vehicle (UAVs), driverless vehicle, cyber technology, wearable technologies and Internet of Things (IoT) has also become cybersecurity threat, therefore there is need to explore the use of Society of Motion Picture and Television Engineers (SMPTE) ST 2110 standard in an IP based data network to enhance the security of Big Data from video transmission [16, 17].

. Widespread adaptation of Internet led to arrival of NCW. Network centric warfare strategy is highly orchestrated, modern, dynamic and autonomous digital battle field. All military units need to be trained for employment of Network Centric Warfare and unconventional operation and especially counter-insurgency. It might now be argued that with general proliferation of worldwide terrorism all military forces and police forces should be trained for unconventional operation. The personal computers (PCs) on tactical vehicles and in satellite should be used to create process and distribute large volumes of data in order manage forces function. The PCs of units involved as shown in Fig. 1, can be integrated and interconnected within same networks environment. Consequently, exchange of information creates situational

awareness leading to fast and correct decision making and quick response action.



Integrated Network Centric Warfare.

The Integrated PCs collect and transmit large volume of data with various relationships shown in Fig. 1. Therefore, extraction of useful information and intelligence out of these data as well as dissemination, of the information for decision making process is paramount to counter insurgency operation. Appropriate advance measures on weaponry and platforms feedback should be the focus in NCW.

IV. TRENDS IN BIG DATA ANALYTIC

Mentioned below are some of the trends that are having huge impact on Big Data Analytic.

a. **Web Servers**: investment in web serve led to key technologies creating conditions for emergence of cyber-attack. The web application is vulnerable to cyber threat attack. Consequently, malicious codes are threat to data extraction or distribution. Nowadays, criminal use legitimate web to distribute malicious codes on the network. In the process, data may be damaged. Moreover, cybercriminals use to steal data from web server platforms. Therefore, there is need for greater emphases on data protection, data transmission and control/internet protocol (TCP/IP) monitoring in order not to fall a prey of cyber-attack.

b. **Real Time Collection and Feedback of Data**: Real time data manually collected to create final report analysis requires real time feedback. This is achievable through encryption of the code. Encryption is the process of encoding data and message information content in such a way that eavesdroppers or hackers cannot read it. There is variety of algorithm to encode data into unreadable cipher text. The use of enhanced encryption could douse the disparate ongoing argument in safety of data as well as enhance the integrity of real time data collection and protect data in transit.

c. **Internet of Thing and its Services**: The internet of Things (IoT) connects all manner of end point, unscrambling a treasure trove of data. Big data analytic can improve the performance of treasure trove information. The IoT device proliferation and ubiquitous networks enable access to massive growth of large amount of weapon and platform information. Both structure and unstructured data are difficult to process using traditional database and software techniques. In this regards, analytic and military intelligence tools required to be predictive than reactive in empowering decision making.

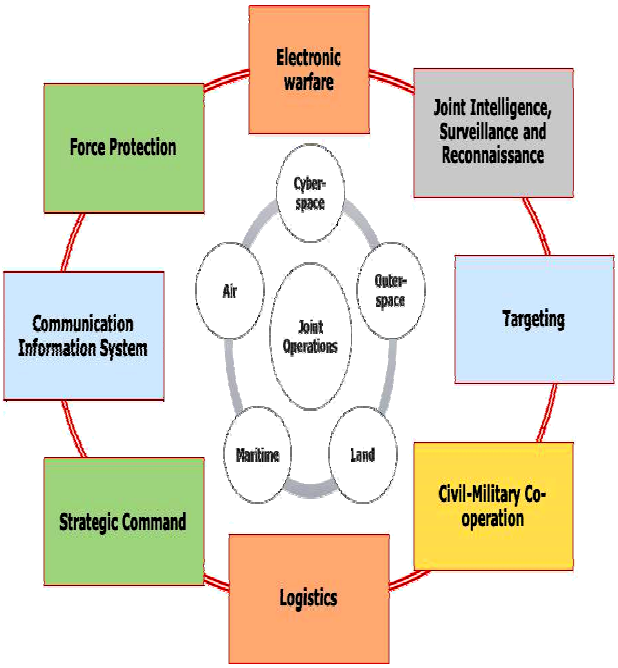
d. **Big Data from Mobile Network**. Mobile cellular network are generators of massive data. Mobile phone devices are themselves data collection tool. The Big data in mobile phones network can be gathered from either internal or external sources. The method of data collection can be through data sources and through auxiliary tools. According to [18 and 19], Apache Hadoop is an open source software framework which is used for Big Data analytic platform. This software usually scans all the files and documents present in the system for extraction of information. The full extent of insurgents' employment of this software for intelligent management is high. This is because, all the features of the software make processing and analyzing of data in mobile cellular network easy. These requirements are vital to war, emergency aid, aid to civil power, maintenance of law and order, the management of hijacking, terrorism, sanctions, coercive diplomacy, alliance warfare, localized war, peace enforcement, limited intervention, protection of allied interests, humanitarian assistance, peacekeeping, peace support, regional war, general war, war of regime survival or war of national survival. Reduced budgetary allocations and advances in technology make it imperative that defence and security planners plan jointly and efficiently.

V. BIG DATA SENSOR BASED MONITORING IN 5G TECHNOLOGY

The concept of the nature and of using sensors-based monitoring technology particularly with 5G technology would provide unlimited access to information and services anywhere and anytime. This calls for skillful application of Big Data Analysis for military counter insurgency operations. To draw on the right balance of using Big data from 5G, network centric battlefield and real time combat surveillance would come into play. This is possible because military personnel are equipped with advanced electronic devices that help them to operate. As Big Data and sensor dynamics lie at the heart of all insurgent's activities, it follows that the nature of counterinsurgency operation will continue to be influenced by and represent the entire spectrum of sensor capability.

The drivers of sensor capabilities include combat identification, adverse weather report, and countermeasure; predict/maintain target tracking and component coordination. This is provided by space communications infrastructure and Unmanned Aerial Vehicle (UAV) imagery. With the 5G, road

sensors can be buried in the road and report status like traffic volumes, congestion, temperature, road damage and wildlife presence [20]. Due to hazardous environments for humans or for increased efficiency and cost, controlling sensor remotely is believed to become an important use of efficient 5G. This led to generation of large amount of data due to integration with other operational domains and supporting disciplines as shown in Fig. 2. Predictive battle analysis of these data is requires for combat identification, and precision engagement. Precision strict is necessary in counterinsurgency operation in order to find, fix, strike target. Francis rightly explain data analysis by saying that "the nation that will insist on drawing a broad line between the data analytic and the non-data analytic is liable to find its fighting done by commanders and its non-data analytic done by cowards" [21].



Counterinsurgency Operational Domains and Supporting Disciplines

Sensor use for social media plays a significant role in generating Big Data in terrorism as well as in the fight against terrorism. Information shared in the field of terrorism, propaganda, and their combats in social media need to be analyzed [22]. For example, it is the responsibility of Big Data analytic to use the Twitter to explored intelligence. In addition, the whole spectrum of the phenomena on Twitter needs to be considered. Measure need to be taken on Twitter by employing clarification, parody and satire, and hacking techniques.

In Big Data acquisition and condition monitoring, military has to make an extension work on existing on-line distributed information system for monitoring on line the condition of insurgent/terrorists communication equipment that are geographically at variously located. There is need to replaced an earlier system of one-to one communication via a modem with an Internet-based client server. The monitoring of the sub-station by remote vision could as well be integrated with the Internet-based system.

VI. CONCLUSION

Insurgents and terrorist criminal activities have globally impacted negatively on the socioeconomic development over the years. As a result, life and property of citizens are threatened continually. In recent times, ICT and Big Data technologies are being leveraged to overcome this challenge. Properly applied network centric warfare and Big Data would secure a fear-free counterinsurgency operational environment. In particular, the advent of IoT, social media and smart sensor equipment shows great potential in this regard. So, this paper proposes a framework of military Big Data technologies to track and monitor crime and criminalities real-time online in order to reduce crime rate in Nigeria.

REFERENCES

[1] Osabiya, B.J.: Ethnic militancy and internal terrorism on Nigeria's national security. Int. J. Dev. Confl. **5**, 59–75 (2015)

[2] P. O'Donovan, K. Leahy, K. Bruton, and D. O'Sullivan, "An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities," Journal of Big Data, vol. 2, no. 1, p. 25, 2015

[3] C. Archetti, "Terrorism, Communication and New Media: Explaining Radicalization in the Digital Age", Perspectives on Terrorism 9(1) (2015): 49-59.

[4] James D. Kiras, "Irregular Warfare" in Understanding Modern Warfare, eds. David Jordan et al. (Cambridge: Cambridge University Press, 2008),229–232.

[5] Sebastian L.V. Gorka and David Kilcullen, "An Actor-Centric Theory of War: Understanding the Difference between COIN and Counterinsurgency", Joint Force Quarterly 60, no. 2 (2011): 17

[6] Burton JW, Conflict: Resolution, Prevention and Prevention. St Martin's Press, New York, 1997.

[7] Monde Muyangwa and M.A. Vogt, An Assessment of the OAU Mechanism for Conflict Prevention, Management and Resolution 1993-2000 p.11.

[8] Rohde, Markus; Aal, Konstantin; Misaki, Kaoru; Randall, Dave; Weibert, Anne; Wulf, Volker: Out of Syria: Mobile Media in Use at the Time of Civil War. In: International Journal of Human-Computer Interaction Bd. 32 (2016), Nr. 7, S. 515–531.

[9] MO Opeloye, "The Boko Haram Insurgency in Nigeria: A Crucial Study of the Movement's Ideology Posture and Implications", Department of Religious Studies, Obafemi Awolowo University, Nigeria, 10 Nov 13

[10] PK Mbote, "Gender, Conflict and Regional Security", in M Mwagiru (ed), Africa Regional Security in the Age of Globalization, (Nairobi: Heinrich Boll Foundation, 2004) p.83

[11] L Pettiford and D Harding, Terrorism The New World War, (London: Accturus Publishing Ltd 26/27 Bickels Yard 151 – 153 Bermondsey Street SE/3HA).

[12] VAdm Arthur K Cebrowski and John J. Grstike "Network Centric Warfare: Its Origin and Future" Proceedings of the Naval Institute 124:1 (January, 1998) 28-35.

[13] David S Alberts, The Unintended Consequences of Information age Technologies (Washinton DC National Defence Unversity Press, 1996.

[14] Major General Jasper A. Welch, Jr., "Command and Control Simulation – A Common Thread", Key Note Address, AGARD Conference Proceedings No. 268 ("Modeling and Simulation of Avionics Systems and C3 Systems"), Paris, France, 15 – 19 October 1979.

[15] Hamza Idris. "Boko Haram: Why we Attacked Telecoms Facilities, "Daily Trust, September 7, 2012, <http://allafrica.com/stories/201209070716.html>, accessed June 9, 2013.

[16] Elder, Lt Gen (ret.) Robert, review of draft of Redefining Information Warfare Boundaries for an Army in a Wireless World, October 13, 2010.

[17] Shancang Li & Li Da Xu & Shanshan Zhao, "The internet of things: a survey", Published online: 26 April 2014,Springer Science+Business Media New York 2014

[18] S.Vikram Phaneendra & E.Madhusudhan Reddy "Big Data-solutions for RDBMS problems-A Survey" In 12th IE

[19] Mrigank Mridul, Akashdeep Khajuria, Snehasish Dutta, Kumar N "Analysis of Bidgata using Apache Hadoop and Map Reduce" Volume 4, Issue 5, May 2014" 27

[20] T.Venkat Narayana Rao,‖5g technologies – an anecdote of network service for the future‖, Journal of Global Research in Computer Science Volume 2 No (7), July 2011 164-170.

[21] Carter, Ashton B., and David B. Omand, "Countering the Proliferation Risks: Adapting the Alliance to the New Security Environment," NATO Review, Vol. 44, No. 5, September 1996, pp. 10–15

[22] Geers, Kenneth, and Nadiya Kostyuk, "Hackers Are Using Malware to Find Vulnerabilities in U.S. Swing States. Expect Cyberattacks," Washington Post, November 5, 2018.

[16]

Combined Cycle Power Plant Technology: Prospects in Boosting Nigeria's Electricity Production

Fatima Sule Alhassan
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
fatimaxarah@gmail.com

Sadiq Thomas
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
sadiqthomas@nileuniversity.edu.ng

Rabiu Aminu
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
abbarabiu09@gmail.com

Zainab Jibril Barau
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
xaynabjb@gmail.com

Oluseun Oyeleke
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
oluseun.oyeleke@nileuniversity.edu.ng

Rukaiya Abdullahi
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
rukiea@yahoo.com

Abstract— The Combined Cycle Power Plant or Combined Cycle Gas Turbine is described as a gas turbine generator in which electricity is generated and the resulting waste heat is used to produce steam to generate additional electricity by means of a steam turbine. The gas turbine is considered one of the most resourceful for conversion of gas fuel to mechanical power or in this case electricity. The electricity sector in Nigeria's has suffered a systemic failure over the years. In the earlier years, considering the 70s and 80s, electric power production and supply was considered ample and barely sufficient to meet the National demand. And recently, it is still experiencing gradual deterioration and is still struggling to meet the increase in the demand for electricity as a result of increase in population. Electricity production in Nigeria over the last 35 years constitutes various methods of generation of which are gas-fired, oil-fired, hydro-electric power stations, and coal-fired station with hydroelectric power system. And with infrastructural developments, oil and gas-fired power stations make up about 85% of her power plants. This paper discusses the prospects of combined cycle power plants (CCPPs) for boosting generation of electricity in Nigeria, the associated technology, and the possibilities to be exploited, hence, meeting up with the rise in the demand for energy for sustainable national development. The paper also suggests the potentials of improving present gas-fired power plants into combined cycle power plants for enhanced electricity generation and supply.

Keywords—Combined Cycle Power Plant, Nigeria, electricity, gas, steam, generation.

I. INTRODUCTION

Nigeria is blessed with sufficient energy resources to meet its present and future developmental needs. The country is ranked the sixth biggest reserve of crude oil in the world and is progressively a domineering gas region with established reserves of almost 5,000 billion cubic meters. Coal and lignite are assessed to be 2.7 billion tons, while tar sand reserves signify an equivalent of 31 billion barrels of oil. Acknowledged hydroelectric generation sites have an expected capacity of about 14,250MW of electric power. Nigeria also has substantial biomass resources for both old and modern energy uses, as well as electricity generation.
The total capacity of power self-generation units in Nigeria is estimated at about 4,000MW. Constant power supply at the required quality still remains a pressing challenge for Nigeria in spite of her abundant energy resources. The installed capacity is about 8,000MW, but only about 7,000MW is available to generate electricity. This paper focuses on the application of the Gas Turbine (GT) – Steam Turbine (ST) type of CCPP for improved electricity production in Nigeria.

The objectives of this paper are as stated below;

- Examine the combined cycle power plant technology as it relates to generation and supply of electricity in Nigeria.
- Maximize energy utilization and increase in the efficiency of plants with the application of the Combined Cycle technology.
- Review the benefits of successful incorporation of this new technology in the current power system in Nigeria and its effect on the National grid.
- Compare emissions and fuel cost of running a CCPP as related to other conventional fossil power plants.

II. LITERATURE REVIEW

In the study of Ertesvag et al., the energy analysis of a gas-turbine combined-cycle power plant with pre-combustion CO_2 capture is presented. A conception for natural-gas (NG) fired power plants with CO_2 apprehension was studied using exergy analysis. NG was restructured in an auto-thermal reformer (ATR), and the CO_2 was separated before the use of a hydrogen-rich fuel for a

conventional combined-cycle (CC) process. The key purpose of the study was to investigate the incorporation of the restructuring process and the combined cycle. A corresponding conventional CC power plant with no CO_2 capture was simulated for comparison[1].

Consequently, Khaliq and Choudhary have studied the combined first and second-law analysis of cogeneration system of gas turbine including inlet air cooling and evaporative after-cooling of the compressor discharge. They implemented Computational analysis to study overall pressure ratio, turbine inlet temperature, and ambient relative humidity as related to the exergy destruction available in each component, first-law efficiency, power-to-heat ratio, and second-law efficiency of the cycle[2].

In the research work of Sanjay et al., energy and exergy analysis of steam cooled reheat gas steam combined cycle was investigated. This particular research paper deals with energy and exergy analysis of repeated gas-steam cycle with the use of steam cooling in a closed loop manner. Compared to the another technique for instance blade cooling, it is shown that the use of the closed loop steam cooling technique is far superior [3].

In the recent work of Sengupta et al., the study of the analysis of exergy is also done with regards to a coal-based thermal power plant with reference to a data provided by the 210MW thermal power station in India.[4]

Consequently, Butcher and Reddy have researched on the Second law analysis of a waste heat recovery-based power generation system. The research lays emphasis on the performance a power generation system with the integration of a waste heat generator in varying conditions of operation[5].

III. Overview of Combined Cycle Power Plants

A combined cycle gas turbine power plant is basically an electrical power plant in which a gas turbine and a steam turbine are used in combination to attain greater efficiency than would be possible if operated independently. In a combined cycle power plant (CCPP) or combined cycle gas turbine (CCGT) plant, as gas turbine generator generates electricity and waste heat is used to make steam to generate additional electricity via a steam turbine. Enhancement in the efficiency of electricity generation is achieved in this last step. In a bid to achieve high efficiency, it follows a rule where the temperature difference between the input and output heat levels be as high as possible. This can be attained by combination of the Brayton (gas) and Rankine (steam) thermodynamics cycle.

Combined Cycle Power Plant fundamentally uses gas and steam technologies in combination which results in higher thermal efficiency compared to other steam plants. In a combined cycle plant, the thermal efficiency is stretched to approximately 50-60% which is as a result of the exhaust gas being piped from the gas turbine to a machine called a heat recovery steam generator. Nevertheless, the heat recovered in this process is suitable enough to startup a steam turbine attached to a single generator[6].

The 'open cycle' operation involves the gas turbine as a standalone unit and a hydraulic clutch suitable for disengaging the steam turbine. As an overall investment, a single-shaft system is usually about 5% lesser in cost, with a higher reliability as shown in the simplicity of its operation. [6].

The main components that comprise a combined cycle are compressor, combustor, gas turbine, Heat Recovery Steam Generator (HRSG), and steam turbine as seen in Fig 1.

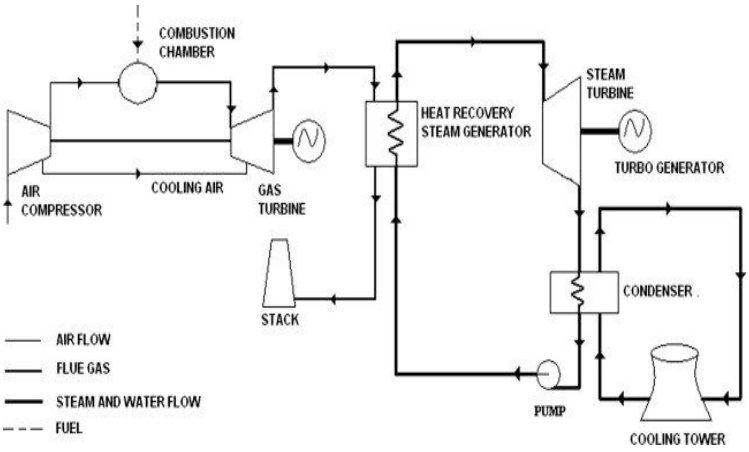


Fig. 1: Schematic flow diagram of combined cycle power plant[7]

a) Compressor: A compressor is a device, which pressurizes a working fluid. Almost all GT plants with generating capacities of over 5 MW have axial flow compressors.

b) Combustor: A combustor typically increases the temperature of the resulting high-pressure gas. In the combustion procedure, the gas turbine combustor air usage is about 10%. The remaining air can be useful in the cooling and mixing stages.

c) Gas turbine: There are two groups of turbines used in a GT plant;
   - Axial flow turbine
   - Radial flow turbine

d) Heat recovery steam generator (HRSG): The subsequent exhaust make its way into the Heat Recovery Steam Generator (HRSG), where transfer of energy which occurs in water produces steam. Usually, the HRSG has different divisions; an Economizer, and Feed-water, an inbuilt Pre-heater and finally a Super-heater. Therefore, the steam leaving the HRSG is consequently superheated before entrance into the steam turbine.

e) Steam turbine: The steam turbines in many large power plants are grouped into two main segments: the High Pressure section (HP) and the Low Pressure section (LP). In certain plants, the HP section is further separated into a High Pressure section and an Intermediate Pressure section (IP).

The division of the HRSG is in such a way that it corresponds to the division of the steam turbine. The action of the back pressure of the condenser defines the overall performance of the LP steam turbine, which is a role of the cooling and the fouling.

f) Generators: the use of air cooled generators are distinctive for small combined cycle systems using gas turbines rated at approximately 100 MW and less. They can either be open ventilated or completely enclosed and water-to-air cooled. Hydrogen-cooled generators are considered more adequate for bigger systems. The cooling of these specific generators can be achieved via plant-cooling or alternatively with water-to-air heat exchangers using ambient air.

IV. Working Principle of Combined Cycle Power Plant

Commonly used combined cycle power plant employs the use of gas turbines known as a combined cycle gas turbine (CCGT) plant.

In regards to the low efficiency present in gas turbine operation in an open cycle, the product of the steam turbine process compensates for approximately half of the CCGT plant output. Fig. 2 shows the basic principles of operation of a typical CCGT.

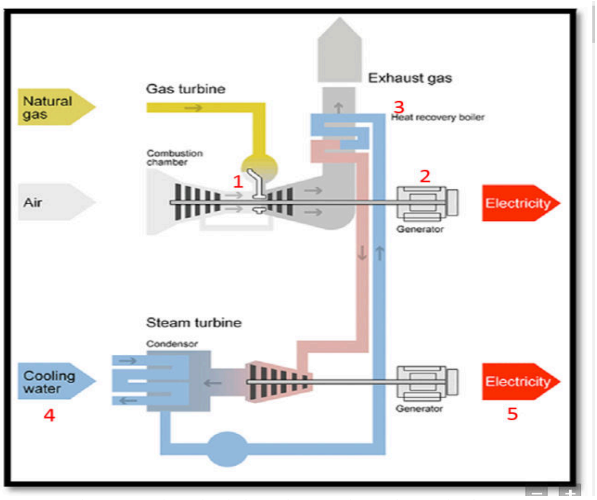


Fig. 2: Basic principles of operation of a typical CCGT[8]

The Air Inlet Stage

In the large air inlet division, air is drawn and it is hereafter cleaned, cooled and controlled. There is a success in operation in various climates and environs as a result of air inlet filtration systems in heavy duty gas turbines. These air filter systems are designed specifically to have compatibility with the plant location. In general, the incoming air has various contaminants[6].

- In gaseous state, contaminants present are Ammonia, Chlorine, Hydrocarbon gases etc.
- In liquid state, contaminants present are Chloride salts dissolved in water, Nitrates, Sulphates etc.
- In solid state, contaminants present are Sand, alumina and silica, Rust, Road dust etc.

The Turbine Cycle Stage

The air is purified, compressed and made into a mixture with natural gas and ignited, which results in expansion of air. The pressure formed from the expansion turns the turbine blades, which are attached to a shaft and a coupled generator, creating electricity. In the next step, steam is generated by the heat produced in the gas turbine by passing it through a HRSG with a live steam temperature of about 420 to 600^0C[6].

Heat Recovery Steam Generator

In a HRSG, highly purified water flows in tubes and the hot gases are circulated, hence, producing steam. The steam then rotates the steam turbine and a coupled generator to produce electricity[6].

V. Thermodynamical Efficiency

The literature survey shows that one-third of the power is being produced by the steam turbine and two-thirds of the power output is produced by the gas turbine. By combination of both gas and steam cycles, an achievement in high input temperatures and low output temperatures can be realized. An addition in the both cycle efficiency is possible because they are characterized by the same fuel source usage. In a bid to increase the efficiency of the system, it is important that the heat recovery steam generator is fully optimized, because it serves as a vital link between the gas turbine and steam turbine cycles, thereby increasing the output of the steam turbine. The performance of the HRSG is considered very important as it reflects directly on the overall performance of the plant. The efficiency of electricity of this plant can increase to approximately 60% when operated first hand without interruption in the output[6].

VI. Electricity Generating Plants in Nigeria

In 2009, in spite of a total grid capacity of 6036 MW, a capacity of 4732MW only was made available. Hence, it shows that approximately 22% of the installed capacity at that period was unavailable. This was due to inconsistencies in operation and failure of generating units to operate at full capacities[9].

Nonetheless, with the current initiative of the federal government via its power plan road map, an improvement has been made on the input of generation as shown in Table 1. The table illustrates that Nigeria presently has twenty-three (23) thermal power plants which are typically gas-fired and three operational hydropower plants; with three other hydropower plants under construction.

TABLE I. Current and Planned Power Plants in Nigeria and their Locations[10]

Power station	Location	Type	Installed capacity (MW)	Year completed
AES Barge	Egbin	SCGT	270	2001
Aba	Aba, Abia State	SCGT	140	2012
Afam IV-V	Afam, Rivers State	SCGT	726	1982
Afam VI	Afam, Rivers State	CCGT	624	2009
Alaoji (NIPP)	Abia State	CCGT	1074	2013
Calabar (NIPP)	Cross River State	SCGT	561	2014
Egbema (NIPP)	Imo State	SCGT	338	2013
Egbin	Egbin	Gas-fired steam Turbine	1320	1986
Geregu I	Geregu, Kogi State	SCGT	414	2007
Geregu II (NIPP)	Geregu, Kogi State	SCGT	434	2013
Ibom (IPP)	Ikot Abasi	SCGT	190	2009
Ihorbor (NIPP)	Benin City	SCGT	450	2013
Okpai	Okpai	CCGT	480	2005
Olorunsogo I	Olorunsogo	CCGT	336	2007
Olorunsogo II	Olorunsogo	CCGT	675	2012
Omoku I	Omoku	SCGT	150	2005
Omoku II (NIPP)	Omoku	SCGT	225	2013
Omotosho I	Omotosho	SCGT	336	2005
Omotosho II (NIPP)	Omotosho	SCGT	450	2012
Sapele	Sapele	Gas-fired steam Turbine	1020	1981
Sapele (NIPP)	Sapele	SCGT	450	2012
Ughelli	Delta State	SCGT	900	1990
Itobe	Kogi State	CFB Technology	1200	2015-2018
Kainji	Niger State	Hydro	800	1968
Jebba	Niger State	Hydro	540	1985
Shiroro	Kaduna State	Hydro	600	1990
Zamfara (Planned)	Zamfara State	Hydro	100	2012
Kano (Planned)	Kano State	Hydro	100	2015
Kiri (Planned)	Benue State	Hydro	35	2016
Mambilla (Planned)	Taraba State	Hydro	3050	2018

VII. Potentials of Combined Cycle Power Plants in Nigeria

Utilization of energy can be capitalized on and the efficiency of plants be increased with the application of the Combined Cycle technology. Nevertheless, this technology is not fully active in Nigeria among its generation options in spite of its prospective of producing larger megawatts (MW) of electricity.

In October 2008, the Afam VI power plant commenced generating electricity which was served by natural gas from the Okoloma gas plant. The power plant commenced generation of over 400 MW to the national grid through the open cycle phase and has reached a capacity of 624 MW through the complete combined-cycle phase contributing 14% to 20% of Nigeria's present power supply. The Afam VI power plant is a sample of the incorporation of modern technology in local generation of power. By exploiting this new combined cycle technology, the usage of gas has decreased by 40% as compared to when it was operated in the open cycle phase. The Afam VI plant conveys about 10 trillion Watt-hour electricity to the grid[10].

Some other power plants in Nigeria exploiting the combined cycle technology although not as efficiently as the Afam VI plant include[11]:

- Alaoji Power station which is partially operational with an installed capacity of 225MW
- Okpai power station with installed capacities of 480MW
- Olorunsogo II power station which is also partially operational with an installed capacity of 675 MW

VIII. RECOMMENDATION

In a bid to attain a more sustainable system in the power sector of Nigeria, the following recommendations are suggested:

1. Construction of more modern combined cycle power stations with greater generating capacities should be encouraged with the help of private investors so as to boost the country's electricity generation.
2. The integration of combined cycle technology, where possible, to the present thermal plants in Nigeria can constitute in more power generation at a reduced cost of installation and fuel consumption
3. Adequate and efficient measures of governance, monitoring and control of such projects will help in curbing corruption and timely deterioration of the proposed power plants.
4. The possibilities of the use of coal in a more efficient and advanced way should be explored such as Integrated Gasified Combined Cycle Power Plants to expand the possibilities and utilization of coal deposits in Nigeria

IX. CONCLUSION

Approximately 45% of Nigeria's populations have access to electricity, and only nearly 30% of their demand for electricity is being met. The power sector is overwhelmed by persistent power failures to the extent that most industrial consumers and a substantial number of residential and other non-residential customers make provision for their own power at a massive cost to both them and the Nigerian economy. Although Nigeria is characterized by abundance of energy resources, these energy potentials have not been fully exploited for production of electricity to its demand nationally. Generation sources are mostly hydro and gas-fired power plants, but there is need for more modern methods of generating electricity through the use of a more efficient, less fuel consuming and inexpensive cost of installation. A new method of generation has to be considered to help in solving these problems and the combined cycle technology offers a sustainable alternative. Integrating this technology to the current power system will open the way for generation of much larger megawatts of electricity to the national grid to meet the demand-supply gap. In conclusion to this, it is easily seen that the generation of electricity at sufficient and adequate amounts is the driving force to viable national growth.

REFERENCES

[1] I. Ertesvag, Kvamsdal, H. M., and Bolland, O., "Exergy Analysis of A Gas-Turbine CombinedCycle Power Plant with Precombustion CO2 Capture," *Energy,* vol. 30, pp. 5-39, 2005.

[2] A. Khaliq, and Choudhary, K., "Combined First and Second-Law Analysis of Gas Turbine Cogeneration System with Inlet Air Cooling and Evaporative Aftercooling of the Compressor Discharge," *ASME Journal of Engineering for Gas Turbine and Power,* vol. 129, pp. 1004-1011, 2007.

[3] Y. Sanjay, Singh, O., and Prasad, B.N., "Energy and Exergy Analysis of Steam Cooled Reheat Gas Steam Combined Cycle," *Applied Thermal Engineering,* vol. 27, pp. 2779-2790, 2007.

[4] S. Sengupta, Datta, A., and Duttagupta, S., "Exergy Analysis of a Coal-Based 210 MW Thermal Power Plant," *International Journal of Energy Research,* vol. 31, pp. 14-28, 2007.

[5] C. J. Butcher, and Reddy, B.V., "Second law analysis of a waste heat recovery based power generation system," *International Journal of Heat and Mass Transfer,* vol. 50, pp. 2355-2363, 2007.

[6] N. R. Jitendra, "Study of Combined Cycle Gas Turbine Power Plant- An Overview," *ieee jornal,* pp. 1-7, 2014.

[7] S. K. Surendra, "GTA Modelling of Combined Cycle Power Plant Efficiency Analysis," *Ain Shams Engineering Journal* vol. 8, pp. 1-22, 2014.

[8] A. Jeff, "Combined Cycle Power Plants " presented at the IMIA Annual Conference 2015, Merida (Yucatán), Mexico 2015.

[9] G. Okoro I. O., P., and Chikuni, E. Power Sector Reforms in Nigeria: Opportunities andChallenges [Online].

[10] A. Mohammed, Jacob, E., John, N., "Combined Cycle Power Plant (CCPPs): Prospects for Gas Turbine - Steam Turbine Power Plants in Nigeria's Electricity Production," *International Digital Organization for Scientific Research* vol. 2, pp. 1-10, 2018.

[11] A. S. Sambo, Garba, B., Zarma, I. H., & Gaji, M. M. , "Electricity Generation and the Present Challenges in the Nigerian Power Sector," pp. 8-12, 2013.

Energy Harvesting: Battery Less Life

Abdullahi O. Adegoke
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
adegoke147abdullahi@gmail.com

Armagan Dursun
Dept. of Elect/Elect Engineering
Nile University of Nigeria
Abuja, Nigeria
Armagan08@yahoo.com

Oluseun Oyeleke
Dept. of Computer Engineering
Nile University of Nigeria
Abuja, Nigeria
ahidbadamasi02@gmail.com

Sadiq Thomas
Dept. of Computer Engineering
Nile University of Nigeria
Abuja, Nigeria
sadiqthomas@nileuniversity.edu.ng

Dominic Nyitamen
Dept. of Elect/Elect Engineering
Nigerian Defence Academy
Kaduna, Nigeria
dsnyitamen@nda.edu.ng

Abstract-Energy harvesting can be defined as a method of capturing and accumulating by-product power when it is easily accessible and can be transformed into usable electrical energy–such as operating within its boundaries a microprocessor. The energy can be obtained from many sources that for any practical purpose would otherwise be wasted or stay unusable. This method, also known as energy scavenging, is about capturing residual energy as a by-product of a natural environmental or industrial phenomenon and can also be regarded as "free power".

Keywords—Energy Harvest, Battery Less, Solar Energy

I. INTRODUCTION

Twelve years have passed since the release of the first smartphone, and in these twelve years the globe has witnessed a revolutionary shift in the age of smartphones, now from 7 to 77 we all use smartphones not only for communication, but also for education, maps, sports etc. The intensive use of smartphones for different reasons has led in the battery shrinking at a fast pace over the previous pair of years. There have been many scientists working to discover a alternative to this but none of them have been willing to do so. Research nowadays focuses not on finding a alternative to the lengthy lives of the battery, but on operating it without batteries.

Residual power is emitted as waste into the atmosphere. For example, mechanical energy is generated through vibration, heat-borne thermal energy, stress and strain, combustion motors and other heating sources. Others are significant biological, solar energy from all kinds of light sources, coils or transformers, electromagnetic energy captured by inductors, air-and liquid fluid energy, chemicals from naturally recurrent or biological procedures, and the mammoth quantity of RF energy in the atmosphere by omnipresent radio and television broadcasting systems. Many materials, well-known devices or sensors are traditionally used to convert waste energy into electrical voltages and currents that can then be harvested, stored and conditioned for many kinds of wearable low-voltage electronics and wireless sensor apps that earlier required AC power supplies or batteries.

II. *LITERATURE REVIEW*

The idea of operating without battery was expressed by a group of researchers from “The University of Washington”, they have invented a cellphone that requires no batteries—a major leap forward in moving beyond chargers, cords and dying phones. Instead, the phone harvests the few microwatts of power it requires from either ambient radio signals or light.1

The entire process of energy harvesting can be classified in various forms depending on their source, amount, and type of energy being converted to electrical energy. In its simplest form, the energy harvesting system needs a source of energy such as heat, light, or vibration, and these three key components.

- Transducer/harvester: This is the energy harvester that collects and converts the energy from the source into electrical energy. Typical transducers consist of light photovoltaic, magnetic inductive, heat thermoelectric, radio frequency RF and kinetic energy piezoelectric.
- Energy storage: Such as a super-capacitor or battery.
- Power management: This makes the electrical energy an appropriate quantity for the implementation. Common conditioners are valves and complicated control circuits which, depending on energy requirements and accessible energy, can operate the engine.

III. OVERVIEW OF ENERGY HARVESTING

Materials such as solar photovoltaic cells, piezoelectric (PZT) crystals or fiber composites, thermoelectric generators (TEGs) and electric inductor coils are examples of energy generators. They generate a broad variety of output voltage and currents. But none of them can be used immediately as energy sources for low-energy electronics

operation without energy harvesting devices intended to catch, handle, and transmit handshake directions to compliant wireless sensor systems. These sources provide energy in many circumstances as bogus, random and otherwise uneven peaks or exceptionally low-level quantities. Recent innovations in MOSFET's "zero-threshold" transistor models have enabled energy-harvesting electronics to achieve fresh heights, help collect, store (in a condenser, super-capacitor or battery) and maintain high retention effectiveness. As we all understand, batteries power most low-power electronics, like integrated devices and distance detectors. Even though they have a restricted life when it comes to long-lasting batteries, they need to be substituted every few years. Replacements can be highly expensive because there are hundreds of detectors in distant places. Whereas energy-harvesting techniques provide endless operating life for low-power machinery and even eliminate the need to substitute batteries where they are expensive, unfeasible or dangerous. Most apps for energy conservation are produced to be self-sustaining, cost-effective, and involve for many years minimal or no service. Additionally, the power is used closest to the source, which completely eliminates transmission losses and long cables. When the energy is enough to power the device directly, the application or device powered by the energy can operate without batteries.

IV. *OVERVIEW OF BATTERY LESS LIFE*

In order to harvest energy from the ambient sources, a base station is required and the base station is placed as per the range specified(31feet-50feet), RF energy can be transmitted in unlicensed bands or grids ranging from 868MHz to 5.4GHz from their respective base stations. It is then, that this RF energy is transfigured into DC power through some energy harnessing devices such as “Powercast’s Powerharvester Receivers”, these receivers also contain criterion or conventional antennas having an average resistance of 40-50 ohms.

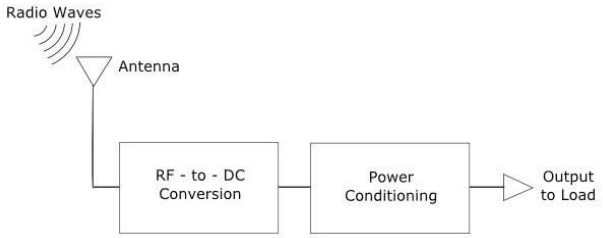


Figure1: Radio waves to DC Conversion 2

The primary characteristic of these harvesters is to preserve efficiency while preserving the disparities between RF and DC transfiguration in order to stabilize production, although proceeding with the latest battery-free telephone technology, the method was quite distinct from the one outlined. To transform ambient light into static or DC energy, a "picture diode" was used in the device. These picture diodes can also generate energy in the lack of light because they contain optical filters and built-in lenses for greater performance in dark circumstances.

Solar cell is a P-N junction diode made of silicon and germanium. Using vapor deposition, P-type coating is diffused over N-type semiconductor with some electrodes having p-type. This method creates open room for light to drop on the P layer and therefore the P-N junction underlying it. The N layer floor consists of the present assembly electrode. They exhilarate electrons from valence band to conductive band when a light photon comes at the intersection, discarding behind an equivalent amount of holes in the valence band.

Due to the barrier area, this produced electron hole in the depletion region accelerates in the reverse direction. The electrons produced by photographs move in the direction of the form n and the electrons to the side p. P side, a favorable electrode, and N side, a harmful electrode is formulated by the hoard of these requests holders. Photo voltage is therefore laid across the intersection. The current passes through the solar cell when the load resistance is attached in the loop. This current from the solar cell is larger than that generated by the RF signals and therefore a reason, the backscatter capability rises to deliver digital products back to the base station, and lastly the range improves. 2

On the other side, MIT scientists have lately taken a move towards achieving this objective in practice. They have created the first fully-flexible machine capable of converting energy from Wi-Fi transmissions into usable electricity with direct current.

The antenna collects the electromagnetic radiation and transforms it into an alternating current. It then goes through a diode that transforms it to direct current for use in electrical circuits that we can call rectenna, while they have been around since the 1960s, they could not function at optical wavelengths. The challenge in achieving this objective was to make the antenna portion of the devices small enough and to make some sort of rectifier in them.

As yet, new optical rectennas are low in efficiency, around 0.1 percent, and therefore can’t compete with the rising efficiency of photovoltaic solar panels. But the theoretical limit for rectenna-based solar cells is likely higher than the Shockley-Quiesser limit for solar cells, and can approach 100 percent if illuminated by radiation of a specific frequency. This makes efficient wireless power transmission feasible.

The fresh rectennas are produced of a molybdenum disulfide (MoS2) 2D material, which is only three atoms dense. Among its many intriguing characteristics is the decrease of parasite capacity— the inclination of metals in electrical systems to behave as condensers, carrying a certain quantity of charge. This can restrict the velocity of signal converters and appliances' capacity to react to elevated frequencies in AC electronics. The fresh molybdenum disulphide rectennas have an order of magnitude of parasitic capacity below those established so far, allowing the unit to catch signals up to 10GHz, including in the spectrum of typical Wi-Fi appliances.

Such a system would have fewer problems associated with a battery: its life cycle could be much longer, the electrical devices would be charged with ambient radiation, and there is no need to dispose of the components as batteries are required.

V. CONCLUSION

With the help of thermoelectric generators, piezoelectric transducers, and solar cells, it is always a challenge when it comes to harvesting electrical power from non-traditional power sources. All of these require a type of power conversion circuit to collect, manage and convert energy effectively from these sources into useful electrical energy for wireless systems, microcontrollers, detectors, and other low-power circuits.

REFERENCES

[1]. https://phys.org/news/2017-07-battery-free-cellphone-harvesting-ambient-power.html/

[2]. https://www.allaboutcircuits.com/technical-articles/wireless-rf-energy-harvesting-rf-to-dc-conversion-powercast-hardware.

[3]. https://singularityhub.com/2019/02/08/how-new-2d-materials-convert-wi-fi-signals-to-electricity/#sm.0001jcjxx1jxneiwsvk2om4cj6ih u

[4]. https://www.seminarsonly.com/electronics/batte ry-less-phones.php

Challenges to Blue Screen Composition Detection in Digital Videos.

[*]Kasim Shafii
Department of Computer Science
Ahmadu Bello University,
Zaria, Nigeria.
kshafii@abu.edu.ng

Mustapha Aminu Bagiwa
Department of Computer Science
Ahmadu Bello University,
Zaria, Nigeria.
mstphaminu@yahoo.com

Nura Sulaiman
Department of Computer Science
Ahmadu Bello University,
Zaria, Nigeria.
nuwair68@gmail.com

Abubakar Usman Mohammed
Department of Computer Science
Ahmadu Bello University,
Zaria, Nigeria.
absadiq_2@yahoo.com

Corresponding author: [*]

***Abstract*— The ability to mate different digital videos into one video using the technology of blue screen composition has become easily achievable due to the increasing availability of low-cost video editing soft wares. This is now creating a great deal of problem to forensic investigators and the general public. Thus, it is becoming important to analyze the genuineness of a digital video before it spreads fast through the Internet and deceive the public. However, only few algorithms have been proposed for detecting and localizing Blue Screen Composition in digital video. This paper presents a discussion on these algorithms highlighting and justifying the motivation as to why these algorithms need to be improved. The potency, limitations and challenges that surrounds these algorithms are also presented.**

Keywords— ***Blue Screen Composition, Video Forgery, Multimedia Forensics, Digital Video.***

I. INTRODUCTION

Tampered digital videos convey false data, spread in all respects rapidly by means of the system and confound the overall population which undoubtedly has certain impact on social stability [5]. These digital videos are easily captured via cameras in handheld devices like a smart phones and tablets which can easily be edited with various video altering software. For instance, digital criminals can create a fake digital video evidence by tampering a video using blue screen composition and present to a court or jurist. This fake digital video when used as admissible evidence might lead to a poor judgment. Hence, blue screen detection has become so relevant research area in digital forensics.

Tampering in digital videos are categorized into two classes as shown in fig. 1 below. Inter-frame and intra-frame tampering. The entire frame undergoes tampering process in the inter-frame. Example of inter-frame altering includes frame addition, frame cancellation, and edge duplicate move. The intra-frame altering, frames are manipulated partially. Examples of intra-frame tampering includes logo removal, tangling synthesis, target copy-paste etc. [1]

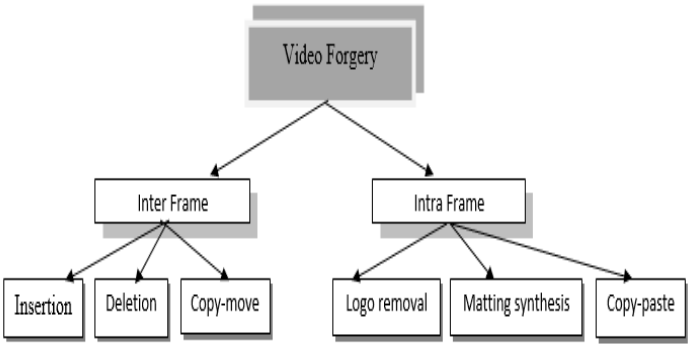


Fig. 1: Video Forgery Categories.

Video tampering detection techniques examine some digital footprints in order to show the difference between original and forged digital videos. The techniques for detecting tampering in digital videos are categorized into two groups namely; the active and the passive-blind detection techniques. The active detection technique requires the pre-inserting of data like watermark, unique finger impression or advanced marks and to identify them through trustworthiness location of the pre-implanted data. The passive-blind detection technique does not require the pre-inserting of additional data. [2]. The altering detection techniques is illustrated in fig. 2.

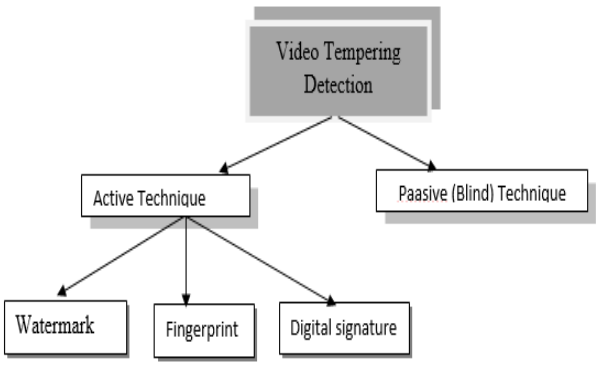


Fig. 2: Video Tampering Techniques

Blue screen composition is a post-production technique or visual effect for layering two pictures or two video streams together dependent on shading [8]. This technique is likewise alluded to as Chroma keying, green screen, or color separation overlay. It is mostly used in news casting, motion picture and video game industries. Blue screen tangling and compositing involves extracting elements of a foreground from a video frame with a steady background color of mostly green or blue and pre-inserting the element of the foreground into another video sequence. Blue or green colors are usually used because of their prominent difference in hue from most human skin. Blue screen tangling and compositing example is illustrated in fig. 3.

Fig. 3: Blue Screen tangling and composition example.

The rest of the paper is divided into 5 sections. Section 2 gives the current state of the art related works on blue screen detection techniques from the literature. Open challenges that surround blue screen detection are highlighted in section 3. In section 4, we provide future research directions and the paper is concluded in section 5.

II. RELATED WORKS

Detecting and localizing blue screen composition in digital video is challenging. Thus, to a great extent, few algorithms have been proposed to recognize and localize this sort of video forgery. In [5], the authors proposed a detection technique for blue screen tampering in digital video which is based on foreground analysis and tracking. This technique contains three main tiers namely; extraction of the foreground block, the detection of the forged block and the tracking of the forged block. In the foreground block extraction, foreground elements are been removed by a multi-pass foreground area finding technique. Luminance comparison and contrast comparison are utilized to gauge the likeness between the foreground component and the background component to pass judgment on whether the objective video has been altered in the subsequent stage. A quick objective search algorithm which is dependent on compressive tracking is utilized to follow the altered region of next frames in the last stage. The technique can rule out the diversion of clamor and other moving foreground in digital video and also can be connected to any video position, bit rate and encoding mechanism. The technique has a genuine positive identification rate of 97.3% and a bogus negative identification rate of7.8%. The technique executes very well in terms of speed. The limitation of their work is that the technique cannot locate forged regions of small sizes. Furthermore, the fast velocity movement of the background element will cause non-perfect test effect.
In addition, another detection technique of blue screen extraordinary effect in videos was proposed in [7]. The variance of statistical features of quantized DCT coefficient between foreground element and background element were exploited. However, the limitation of their work is that the accuracy of the technique decreases with different bit rate encoding and the technique is fit for MPEG video type only.

(Su et al, 2011) [6] Proposed a blue screen composition detection technique on edge features. The author using the Prewitt algorithm, exploited the changes of the relationship between the shading signals on the edge of each article in the altered video to find edges of each item and thus, calculating precise variables to discover questionable region in a altered video. Unfortunately, the accuracy of the technique can easily be affected by noise and other moving foreground.

Furthermore, digital video tampering using Chroma key background technology was proposed in [8]. Statistical correlation of blurring artifact was extracted from digital video to detect this type of tampering The technique can adequately detect Chroma key forgery in a manipulated digital multimedia video with a genuine positive identification rate of 91.12% and a bogus positive identification rate of 1.95%. The result in the technique shows that it cannot adequately satisfy accuracy and the reliability of the system simultaneously.

III. CHALLENGES

In this section, we discuss the research challenges affecting the detection techniques for blue screen composition in digital video from the literature.

A. The Size of Forged Block:

Most of the proposed techniques in the literature cannot detect a forged region or block when it is too small. This is because they have very low resolution and are greatly influenced by foreground elements. Thus, addressing this problem still remains a challenge.

B. Velocity Movement.

Fast velocity movement or severe shaking of the background element in a tampered video during shooting can cause non-ideal experimental effect to state of 'the art blue screen detection technique. The movements literally affects the accuracy of these algorithms.

A. Evaluation Method.

Developing an accepted method for evaluation poses a great challenge in digital video tampering. This is because the existing methods for evaluation only rely solemnly on qualitative evaluation

IV. FUTURE WORK

In future research direction, there is need for an accepted method of evaluation in order to uphold certainty in the genuineness of digital video content. Existing methods for evaluating tampered blue screen composition rely on qualitative evaluation, existing digital video quality algorithms adaptation and no reference quality measurements will accelerate improvement [4]. Detecting a forged region of small size and the fast velocity movement of background elements can be resolved by combining different techniques together for blue screen detection. Until then, tampering in blue screen composition remains the gold standard.

V. CONCLUSION

The detection of blue screen composition is becoming more crucial in digital forensic investigators due to the advance in digitalization today in our society. However, only very few algorithms have been developed that can adequately detect the presence of manipulation in digital video through blue screen composition. Furthermore, in order to keep up certainty in the genuineness of digital video content in the future it is very important to create a technique which can detect and localize manipulation through blue screen composition more reliable and accurately.

REFERENCES

[1] Sitara K, B.M. Mehtre (2018), Detection of inter-frame forgeries in digital videos. Forensic Science International.

[2] Omar Al-Sanjary, GhazaliSulong,(2015), Detection of video forgery: A review of literature. Journal of theoretical and applied information technology.

[3] RohiniSawant, ManajSabnis. A Review of Video Forgery and Its Detection (2018). Journal of Coputer Engineering. doi:https://doi.org/10.9790/0661-2002030104

[4] Johnston P, Elyan E, A review of digital video tampering: From simple editing to full synthesis, Digital Investigation (2019), doi:https://doi.org/10.1016/j.diin.2019.03.006.

[5] Yuqing Liu, Tianqiang Huang, Yanfang Liu, A vovel video forgery detection algorithm for blue screen compositing based on 3-stage foreground analysis and tracking, Multimed Tools Appl (2017), doi:https://doi.org/10.1007/s11042-017-4652-7

[6] Su Y, Han Y, Zhang C (2011) Detection of blue screen based on Edge Features, Information Technology an Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International. 469–472

[7] Xu J, Yu Y, Su Y, Dong B, You X, Detection of Blue Screen Special Effects in Videos (2012) International Conference on medical physics and biomedical engineering. PhysProcedia 33:1316–1322

[8] Bagiwa MA, Wahab AWA, Idris MYI, Khan S, Choo K-KR, Chroma key background detection for digital video using statistical correlation of blurring artifact, Digital Investigation (2016), doi: 10.1016/j.diin.2016.09.001.

A Secured Automated Electronic Bimodal Biometric Voting System

Kennedy Okokpujie
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
kennnedy.okokpujie@covenantuniversity.edu.ng

John Abubakar Amanesi
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
john.abubakar@stu.cu.edu.ng

Etinosa Noma-Osaghae
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
etinosa.noma-osaghae@covenantuniversity.edu.ng

Imhade Princess Okokpujie
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
imhade.okokpujie@covenantuniversity.edu.ng

Samuel Ndueso John
Department of Electrical and Electronics Engineering
Nigerian Defence Academy
Kaduna, Nigeria
samuel.john@nda.edu.ng

Charles U. Ndujiuba
Department of Telecommunication Engineering
Air Force Institute of Technology
Kaduna, Nigeria
ndujiuba@yahoo.com

***Abstract*—Insecurity, rigging and violence continue to mar electoral processes in developing nations. It has been difficult to enforce security and transparency in the voting process. This paper proposes a secure and automated bimodal voting system. The system uses three security layers, namely, a unique ID code, a token passcode that expires every five minutes and biometrics (iris and fingerprint). A scanner captures the fingerprint and iris of eligible voters. The fingerprint and iris images are stored along with the corresponding particulars in a database. The software implemented is a .net managed code in C#. The result of this system shows the system is transparent, fast and fraud-free. The proposed system had a Failure to Enroll (FTE) and a Failure to Capture (FTC) of zero.**

Keywords—biometric, iris, fingerprint, voting, secure, database, software, hardware, election.

I. INTRODUCTION

Voting is the most indispensable asset in any democratic country. It is the process of selecting a suitable candidate to lead the people. A democratic country is the people's country. This can only be true when there is provision for a trustworthy and secured electoral process [1]. E-voting is an emerging technology that has improved the traditional method of voting [2].

E-voting with the use of biometric has provided a more secure way of voting in a democratic country compared to the traditional voting where papers are used and voting is insecure [3]. Biometric is a physical and biological quality of an individual which is different for every person [4]. There are different types of biometric traits among which are facial recognition, fingerprint, iris recognition, palm print etc. [5].

This paper proposes a web-based and secured automatic bimodal biometric electronic voting system. The biometrics was used to identify individuals that are eligible to vote. The proposed system provided eligible voters with a unique ID and a token code.

The proposed electronic voting system also provided a dependable, transparent and secured electronic voting system that eliminated the possibility of impersonation by using two biometric traits and automating the voting process to save time. The resulting system had the propensity to improve the integrity factor of the voting process by making it fast, transparent and robust.

The paper is organized thus, a brief review of existing literature on adaptive trait age invariant face recognition systems is given in the section two (Literature Review). A thorough expository on the methodology used for the study is given in section three (Methodology). The results of the study are elucidated in section four (Results and Discussion). A conclusion is made and all parties involved in the success of the study are acknowledged. A list of references is given to close the study.

II. REVIEW OF RELATED WORKS

The [6], developed a web-based voting system using Fingerprint Recognition. The design proposed was used for a university's presidential election. Four candidates and 40 voters registered for the election. Each voter's particulars, biometric and regular were collected and stored in a database. During the Election, the registered voters were able to cast votes over the internet. The software used to implement the e-voting system was written in C#.

In [7], a framework was proposed for a low cost secured electronic voting system based on facial recognition using Local Binary Pattern (LBP) for extracting facial feature characterization in texture format and chi-square for image classification. A two level security using a passcode and biometric (face) was implemented. The proposed system was web-based. The system eliminated the need to wait for the vote to be counted by providing a page that shows the live count of the election every second, thereby minimizing vote counting time.

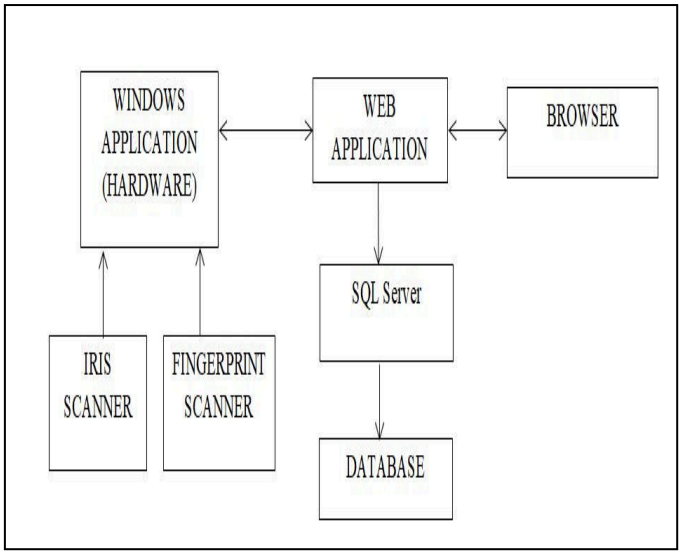
[8] proposed a framework to ensure secured identification and authentication processes for voters using Fingerprint biometric. The main aim was to eliminate fake voters, vote repetition and provide more transparency. The Fingerprint was used for the identification of the individual. In the proposed system, a network connection and electoral

officers were not needed. The electronic voting machine was designed to direct eligible voters on how to cast votes without the need of a network connection or an attending electoral officer.

M. Olaniyi et al, in [9] designed a secure electronic voting system using fingerprint biometric and the crypto-watermarking approach. The fingerprint biometric was used for the identification of the individual. The system also used an encryption standard (AES) cryptographic algorithm to improve the integrity of the proposed system.

III. METHODOLOGY

The proposed electronic voting system was centred on two trait biometrics, the fingerprint and the iris. The system was designed to improve the overall security and credibility of electoral processes. The proposed system was designed using an Irishield-UART MO2120, a DigitalPersona U&U 4500 fingerprint scanner and a Personal Computer. The biometric sensors were used to acquire the wanted biometric trait and the personal computer was used as a development environment to create the database and software used to cast votes. Figure 1 shows the function block diagram of the proposed voting system.

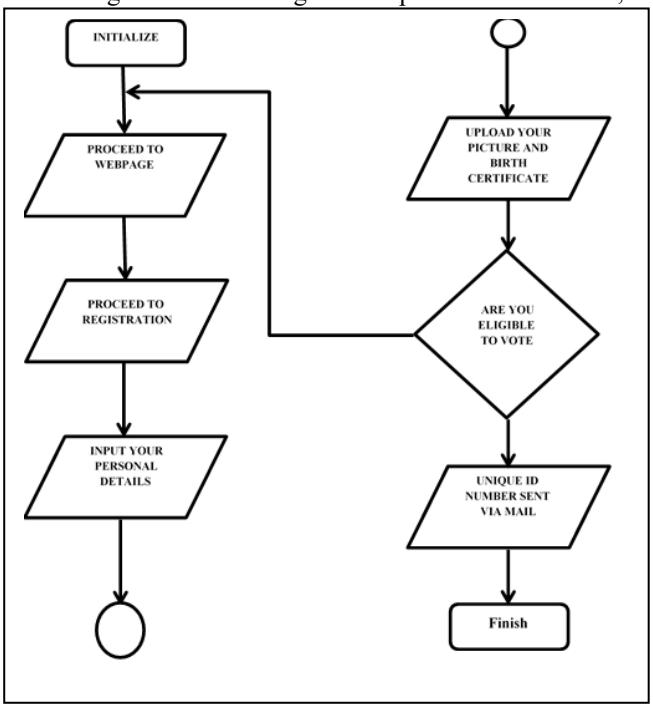


Functional block diagram of the secured biometric electronic voting system

The proposed voting system was programmed using C# language. SQL server was used to create the electronic voting database. The database created had three tables; the candidate table for the candidate data, the voter table for the voter (The eligible voters) information with their biometric details and the vote table for vote counting. Visual studio 2017 was the Integrated Development Environment (IDE) used to implement the voting system. Visual studio has a lot of unique that made it easy to integrate various plugins. The codes used in C# were long. Visual studio made it possible to find a code in the midst of lots of other codes.

During the registration phase, various information was required from the individual or user such as username, state of origin, password, also validation of birth certificate etc. The registration was used to create an online account for each eligible voter. Immediately after registration, individuals are given a unique code. The flow chart in Figure 2 shows how the registration process was carried out. Figure 3 shows the online registration page of the proposed voting system.

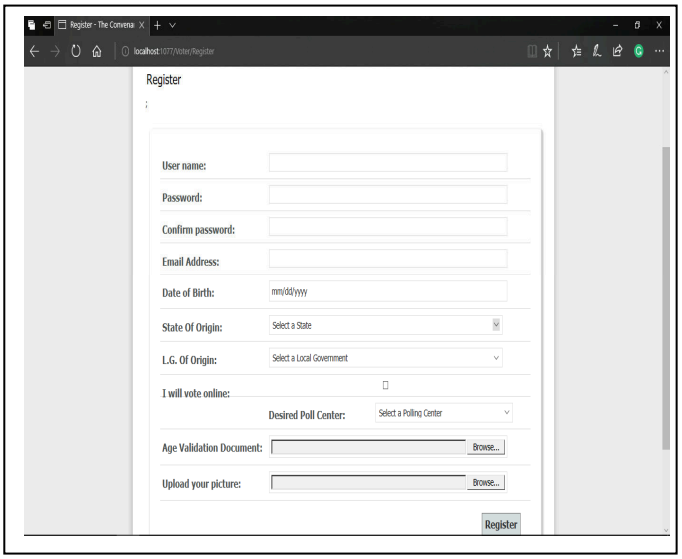
The algorithm for the registration process is as follows;



Step 1: Initialize
Step 2: Proceed to webpage
Step 3: Proceed to registration
Step 4: Input your personal details

Flow chart for the registration process

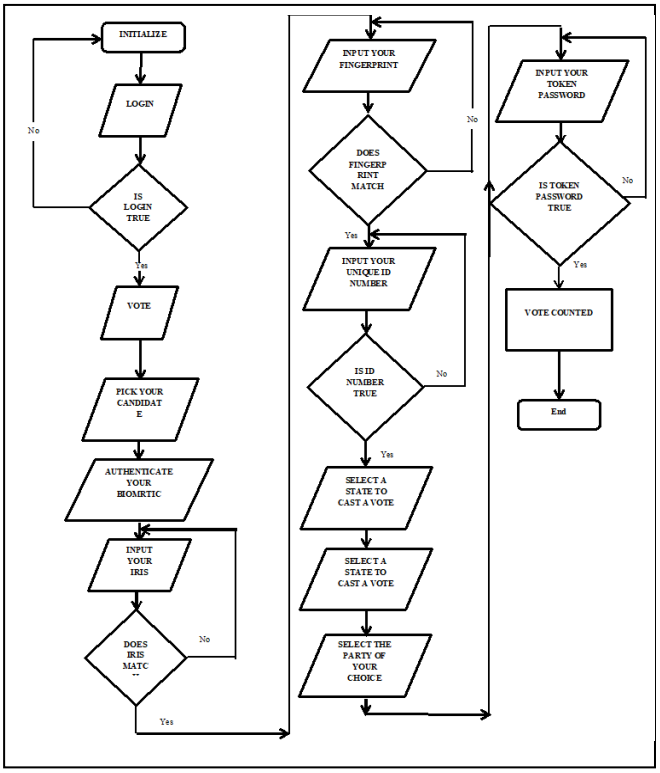
Step 5: Upload your picture and birth certificate
Step 6: Are you eligible to vote
Go to step 7
Else
Go to step 1
Step 7: Unique ID sent via mail
Step 8: Press finish to redirect to home page



Flow chart for the registration process

The next phase after registration was the biometric enrolment. Iris and fingerprint samples were acquired from eligible voters using the Irishield-UART MO2120 and the Digital Persona U&U 4500 fingerprint scanner respectively. The information (biometric images) got from each eligible voter were stored along with the voter's particulars in the database.

Once an individual has been enrolled into the system, the user has the privilege to vote. Authentication, also known as identification implies a one-to-one match. During the authentication stage, the biometric sample of the user is compared to the previously stored information. Figure 4



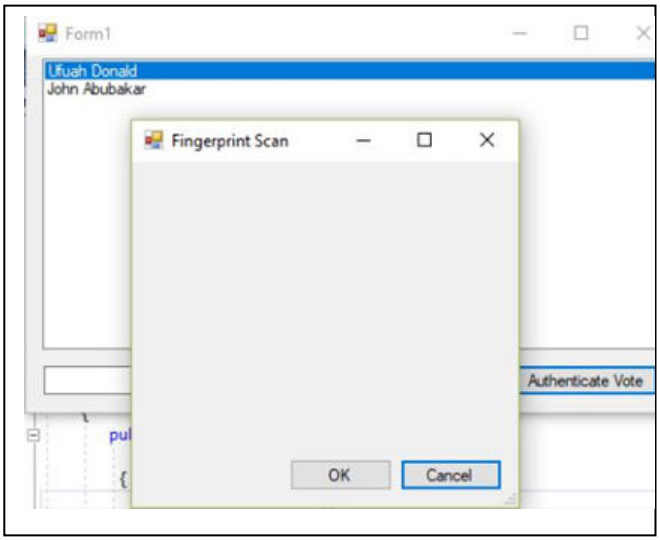
shows the flowchart of the voting process. The user is first prompted to provide a unique ID and token password.

Flow Chart of the Proposed Electronic Voting System

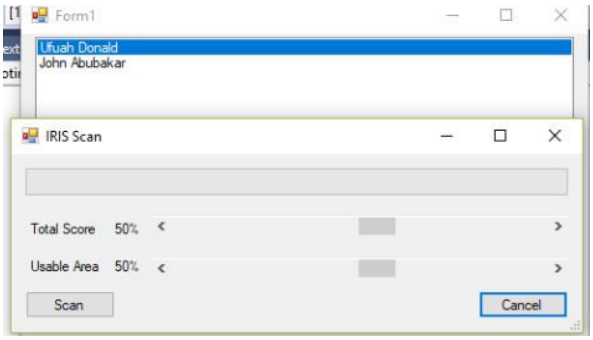
The algorithm of this process described by Figure 4 is as follows;

Step 1: Initialize
Step 2: Login
Step 3: Is login true
Go to step 4
Else
Go to step 1
Step 4: Vote
Step 5: Pick your candidate
Step 6: Authenticate your biometric
Step 7: Input your iris
Step 8: Does iris match
Go to step 9
Else
Go to step 7
Step 9: Input your fingerprint
Step 10: Does fingerprint match
Go to step 11
Else
Go to step 9
Step 11: Input you unique ID code
Step 12: Is ID true
Go to step 13
Else
Go to step 11
Step 13: Select a state to cast your vote
Step 14: Select the party of your choice
Step 15: Input your token password
Step 16: Is token password true
Go to step 17
Else
Go to step 15
Step 17: Vote counted
Step 18: End

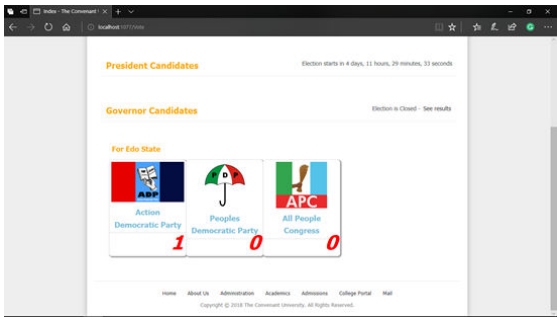
The biometric authentication stage requires the voter to provide biometric input (iris and fingerprint). The biometric input is compared to the biometric information saved on the database (i.e. the system performs a one-to-many authentication) as shown in Figure 5 and Figure 6.



Fingerprint Authentication



Iris Authentication

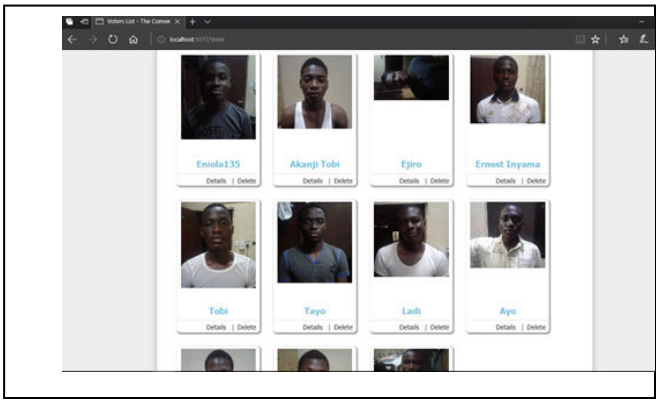


Vote counted immediately after voting

A correct unique ID, password token and a match for the biometric information provided qualifies a user to cast a vote. An administrator can login to the web-based platform to monitor the election and also view the total result as shown in Figure 7.

IV. Result and Discussion

After testing the functionality of the system, registration began and various students registered. A database of various users was developed as shown in Figure 8. The detail of users can be seen from the administrator's end.



Enrolled Users

It was discovered that the performance of the biometric system could be influenced by environmental factors during the image acquisition stage and the performance factors of image quality algorithm used. In order to measure the accuracy and performance of the biometric system, the following performance metrics were used;

1. Time of Enrolment (TOF): This was the time it took for an individual to enrol The enrolment time during voting was 5msec.

2. Failure to Enrol (FTE): This occurs when the iris scanner and the fingerprint sensor consider a data invalid during enrolment. During the measurement analysis, the system enrolled everybody. The FTE is zero (0)

3. Failure to Capture (FTC): This occurs when the sensors fail to capture the data presented by the individual (the fingerprint and the iris). The FTC is zero (0).

Conclusion

The major objectives of this study were achieved as the electronic voting system using bimodal biometric eliminated fraud and the possibility of voting more than once. The implementation of a three level security bimodal biometric e-voting system was successful. It was web-based and allowed only one time voting for each eligible user.

Further Work

The algorithms for classification and recognition needs to be fine-tuned or improved upon to achieve more accurate results.

Acknowledgement

This paper is sponsored by Covenant University, Ota, Ogun State, Nigeria.

References

K. O. Okokpujie, E. Noma-Osaghae, O. J. Okesola, S. N. John, and O. Robert, "Design and implementation of a student attendance system using iris biometric recognition," in 2017 International Conference on Computational Science and Computational Intelligence (CSCI), 2017, pp. 563-567: IEEE.

K. O. Okokpujie, S. N. John, E. Noma-Osaghae, C. Ndujiuba, and I. P. Okokpujie, "AN ENHANCED VOTERS REGISTRATION AND AUTHENTICATION APPLICATION USING IRIS RECOGNITION TECHNOLOGY," International Journal of Civil Engineering and Technology (IJCIET), vol. 10, no. 2, pp. 57-68, 2019.

C. Atuegwu, S. Daramola, K. O. Okokpujie, and E. Noma-Osaghae, "Development of an Improved Fingerprint Feature Extraction Algorithm for Personal Verification," International Journal of Applied Engineering Research, vol. 13, no. 9, pp. 6608-6612, 2018.

K. Okokpujie et al., "Integration of Iris Biometrics in Automated Teller Machines for Enhanced User Authentication," in International Conference on Information Science and Applications, 2018, pp. 219-228: Springer.

E. Noma-Osaghae, O. Robert, C. Okereke, O. J. Okesola, and K. Okokpujie, "Design and Implementation of an Iris Biometric Door Access Control System," in 2017 International Conference on Computational Science and Computational Intelligence (CSCI), 2017, pp. 590-593: IEEE.

R. F. L. Chavez, Y. Iano, and V. B. Sablon, "Process of recognition of human iris: Fast segmentation of iris," ed: IEEE, 2006

Neha Kak, Rishi Gupta, Sanchit Mahajan, "Iris Recognition System", (IJACSA)International Journal of Advanced Computer Science and Applications, Vol. 1, No. 1, 2010

F. I. Hazzaa, S. Kadry, and O. K. Zein, "Web-Based Voting System Using Fingerprint: Design and Implementation," International Journal of Computer Applications in Engineering Sciences, vol. 2, 2012

O. M. Olaniyi, T. A. Folorunso, A. Aliyu, and J. Olugbenga, "Design of Secure Electronic Voting System Using Fingerprint Biometrics and Crypto-Watermarking Approach," International Journal of Information Engineering and Electronic Business, vol. 8, p. 9, 2016.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

Development of Facial Trait Recognition System

Kennedy Okokpujie
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
kennnedy.okokpujie@covenantuniversity.edu.ng

Samuel Ndueso John
Department of Electrical and Electronics Engineering
Nigerian Defence Academy
Kaduna, Nigeria
samuel.john@nda.edu.ng

Ufuah Donald Ehikioya
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
donald.ufuah@stu.cu.edu.ng

Martins Nwagu
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
martins.nwagu@stu.cu.edu.ng

Etinosa Noma-Osaghae
Department of Electrical and Information Engineering
Covenant University
Ota, Ogun State, Nigeria
etinosa.noma-osaghae@covenantuniversity.edu.ng

Charles U. Ndujiuba
Department of Telecommunication Engineering
Air Force Institute of Technology
Kaduna, Nigeria
ndujiuba@yahoo.com

***Abstract*—Face recognition systems have gained much attention for applications in surveillance, access control, forensics, border control etc. Face recognition systems encounter challenges due to variation in illumination, pose, expression, occlusion and most importantly, aging. Face recognition systems are unable to adjust to the changes that occur in the human face as a result of trait aging. In this study, an age invariant face recognition system is developed using a 4-layered Convolutional Neural Network (CNN). The proposed system was able to recognize/identify face trait of individuals across age groups. The variations caused by trait aging was modelled as a form of time varying noise and it was validated by computing its error statistics and comparing its performance with existing models found in literature. The result of the study showed that an adaptive and robust face recognition system that is trait aging invariant can be achieved with CNN. The recognition accuracy achieved by the study was 99.22% with less than six (6) epochs.**

Keywords—adaptive, facial recognition, convoluted neural network, epochs, artificial intelligence, layers, biometrics, trait, aging

I. INTRODUCTION

The main objective of a biometric system is to accurately identify individuals. This implies that biometric systems must have low recognition error rates. Face recognition over the last decade has become an area of interest for many researchers because of its wide range of applications [1, 2, 3]

Face recognition systems are adversely affected by trait aging. Trait aging is a genetic factor that compromises recognition accuracy in face recognition systems. Trait aging, as it affects facial recognition system has become a hotbed for researchers in the area of image processing and computer vision [4, 5]. A robust Facial recognition system capable of adapting to variations in the face is necessary to curb the adverse effects of trait aging on facial recognition systems. This will reduce the need for re-enrolling individuals and the cost of infrastructure procurement in the long run. In this study, a robust facial recognition system capable of adapting to variations due to trait aging in individuals using Convolutional Neural Networks is considered.

The proposed system was able to recognize/identify face trait of individuals across age groups. The variations caused by trait aging was modelled as a form of time varying noise and it was validated by computing its error statistics and comparing its performance with existing models found in literature. The system was able to adapt to changes in facial features as individuals age. Face detection, pre-processing and feature extraction was done using a 4-layer convolution network of size 3×3, each having 5, 10, and 50 kernels respectively. Dimensionality reduction was done using the pooling layers of the network while face recognition was done using the fully connected layer of the CNN with a Softmax classification function. Training was done using Stochastic Gradient Decent with Momentum (SGDM). The system accepted input images of size 180×200×3 obtained from a moderate trait aging dataset that belonged to the university Essex, United Kingdom. Simulation was done using the MatConvNet toolbox in MATLAB R2017b.

The paper is organized thus, a brief review of existing literature on adaptive trait age invariant face recognition systems is given in the section two (Literature Review). A thorough expository on the methodology used for the study is given in section three (Methodology). The results of the study are elucidated in section four (Results and Discussion). A conclusion is made and all parties involved in the success of the study are acknowledged. A list of references is given to close the study.

II. REVIEW OF RELATED WORKS

The human face is made up of layers of muscles, skin and tissue that reside above the bones of the face. The face keeps changing after birth and its growth is widely affected by environmental factors such as body weight, lifestyle, sunrays, smoke and the type of food being consumed by an individual. The facial muscles may differ in their control, position, form and presence. Facial characteristics of individuals changes with increase in age.

Aging causes a reduction in the elasticity of the skin. However, the face is mapped with representations that are rapid and remain permanent. This is the main reason why identical twins can be easily identified when they are older than when they are younger. Geometric features and wrinkles that become more prominent are used for age estimation. According to [6], face wrinkle analysis aids in the categorization of the age group of humans. The facial aging features can be represented using Active Appearance Models (AAMs). AAMs extract the global features of the face using statistical and appearance models. They are limited in functionality because they do not consider the local features of the face embedded in wrinkles.

The use of coordinate transformations for modifying the structure of biological organisms so as to synthesize a structure of similar but different organisms was proposed by [7]. As a result of The investigation of the application of coordinate transformations was employed by some scientist [7]. This was a trial to imbibe age associated alterations to the face. These scientists came up with two of such transformations which comprise Cardiodal Strain and shear transformations. The former was effective for changing the detected face age.

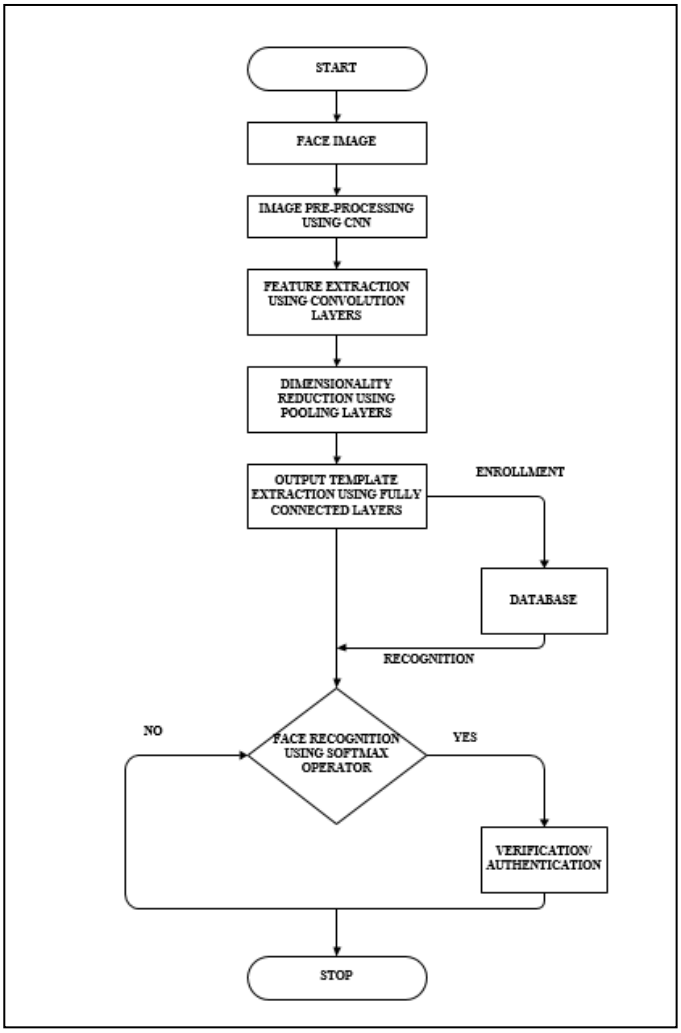
Further work was done by [8] showing outcomes. Here, the fundamental idea by Thompson was applied to 3D facial data. However, the method only created room for the manipulation of the shape of the face. It was oblivious to how variations in facial colour could be solved. The simulation of age effects on the basis of shape and colour detail was first attempted by Perrett and Burt. Various face composites obtained from several age groups and caricature algorithms were utilized to investigate the aging process. The authors concluded that the subjects' perceived age (blended images) utilized for producing respective composites does not change with respect to the real age employed in producing respective composites. This means that age data for each group is reserved through the blending process [9].

III. METHODOLOGY

The proposed system used Convolutional Neural Network (CNN) for pre-processing, feature extraction and face recognition. It used Softmax operator as the classifier. The proposed system used a moderate aging dataset set obtained from the University of Essex, United Kingdom to train the network. The dataset contained 395 individuals with 20 images per person making a total of 7,900 images. It contained both female and male subjects with ages ranging from 18 to 20 years. It also contained images of some older individuals.

The system was implemented using the MatConvNet toolbox which comprises a library of MATLAB functions implementing convolutional neural network architectures for computer vision applications. Figure 1 shows the flow chart of the implemented system.

Face recognition and classification of the input image was done using the fully connected layers of a convolutional neural network. The fully-connected layers of the network were utilized for learning the classification function. The output was classified using SVM (support Vector Machine), a probability distribution function which gives the top three



probabilities of the input from which the output with the greatest probability is identified as the input image.

The proposed system uses a CNN architecture shown in Figure 2. It has an input layer size of 180x200x3, three groups of convolutional layers, an activation layer (ReLU) and a pooling layer, followed by a fully connected layer of size 152 and a Softmax classification layer. The Input to the CNN is an image from the dataset. The Convolutional layers are responsible for optimum feature extraction while the pooling layers perform dimensionality reduction. The three

Flow chart of Proposed Adaptive Trait Age Face Recognition System

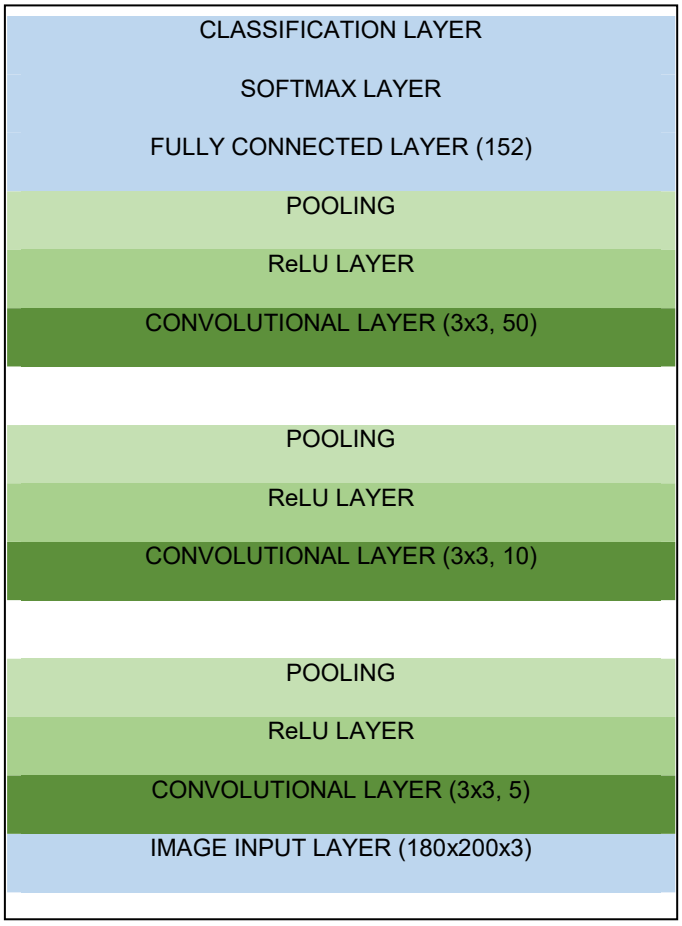
Convolutional layers use 5, 10 and 50 filters each of size 3x3 respectively. Each pooling layer uses a filter of size 2x2. The fully connected nodes represent the total image classes in the dataset.

The images in the database used for training and testing were face photographs obtained from undergraduate students of the University of Essex taken in a controlled environment of different illumination, pose, expression and age. Some of the samples are shown in Figure 3.

Three hundred and ninety-five (395) individual faces were obtained with a digital camera. Each individual has twenty (20) images with different face expressions, illumination, pose and age. The images obtained are in JPEG format. In each class, the first five images were utilized for

training while the last fifteen images were used for testing the system to ascertain its efficiency.

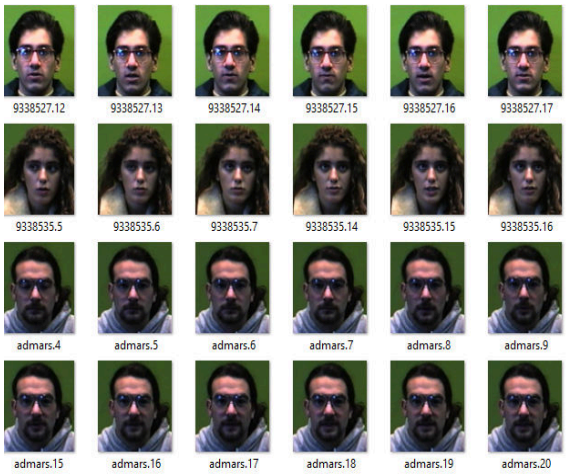
The test images were loaded from the test database and compare with the trained database. The image is classified by the neural network and the output was the user ID of the



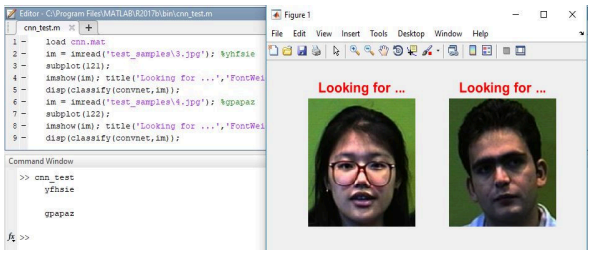
class containing all images of the test subjects. Figure 4 and Figure 5 show the user ID for the class of the test subjects. The proposed facial recognition system was executed on a personal computer with a 2.5GHz Intel i5-7200U Quad core processor and a 12GB DDR4 SDRAM (2DIMM).

IV. Result and Discussion

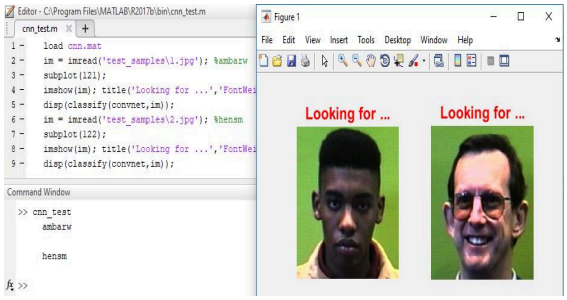
The face recognition system is developed using MATLAB programming language. The output of the training process is shown in the Figure 6. During pre-processing, each subject was scaled to 180×200. The images were partitioned in the ratio, 15:5 for training and testing respectively. When the system executes, the subjects in the training database were trained in min-batches. Training was done on the mini-batches and stopped when a new accuracy was obtained. Finally, the system selects an image to be tested which displays the user ID identical to the ID tag for each subject. The system was tested with multiple numbers of new subjects. The analysis of the number of epoch and average training time is shown Table 1.The accuracy report showed that only a few epochs are needed to train the images in the database. Less than 6 epochs were used. The system was able to classify images as either known or unknown with a 0.01learning rate. The average training time increased as the network grew. This implied that an architecture with more hidden layers would drastically increase the training time. The result showed that the proposed adaptive trait aging face recognition system had a testing accuracy of 99.22%.



Sample Faces from Database



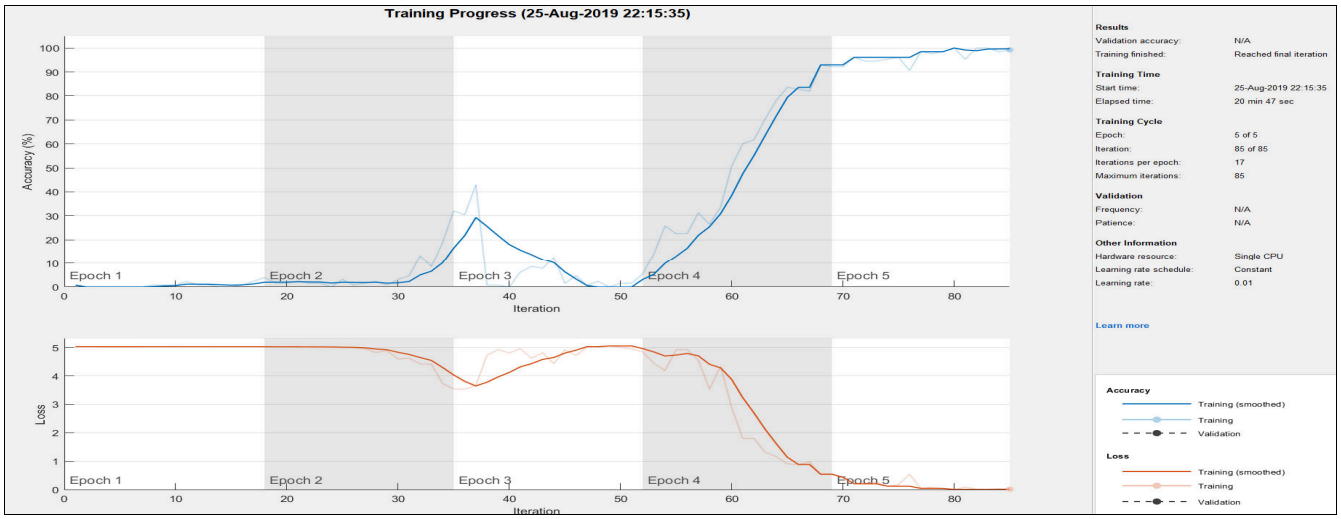
Test Result 1



Test Result 2

Conclusion

An adaptive age facial recognition system using convolutional neural networks was developed. The facial recognition system was invariant to changes in the face as a result of time lapse. The system was built using a four-layered CNN.



Training Progress of Proposed Adaptive Trait Age Face Recognition System

Table 1. Accuracy Report of Proposed Adaptive Trait Age Face Recognition System

```
Training on single CPU.
Initializing image normalization.
|=========================================================================================|
|     Epoch    |   Iteration  | Time Elapsed |  Mini-batch  |  Mini-batch  | Base Learning|
|              |              |  (seconds)   |     Loss     |   Accuracy   |     Rate     |
|=========================================================================================|
|            1 |            1 |        18.79 |       5.0253 |        0.78% |       0.0100 |
|            3 |           50 |       742.73 |       5.0062 |        1.56% |       0.0100 |
|            5 |           85 |      1247.57 |       0.0149 |       99.22% |       0.0100 |
|=========================================================================================|
CNN successfully trained with accuracy = 98.29
CNN successfully testing with accuracy = 99.22
```

FURTHER WORK

For the aging variation to be totally curbed, the system should be trained with a much larger database such as FG-NET and MORPH with an increase in the number of CNN layers. However, to reduce training time and number of layers required for this process, a combination of the subsampling and convolution layer can be done.

ACKNOWLEDGMENT

This paper is sponsored by Covenant University, Ota, Ogun State, Nigeria.

REFERENCES

A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters,* vol. 79, pp. 80-105, 2016/08/01/ 2016.

S. Ravishankar, T. Narayan and D. Deshpande, "Face Detection and Recognition using Viola-Jones algorithm and fusion of LDA and ANN," *IOSR Journal of Computer Engineering (IOSR-JCE),* vol. 18, no. 6, pp. 2-3, 2016.

S. Tolba, A.A. El-Harby, "Face Recognition: A Literature Review," *International Journal of Signal Processing,* vol. 2, 2015.

K. Okokpujie, E. Noma-Osaghae, S. John, K.-A. Grace, and I. Okokpujie, "A face recognition attendance system with GSM notification," in Electro-Technology for National Development (NIGERCON), 2017 IEEE 3rd International Conference on, 2017, pp. 239-244.

K. Okokpujie, E. Noma-Osaghae, S. John, and R. Oputa, "Development of a facial recognition system with email identification message relay mechanism," in Computing Networking and Informatics (ICCNI), 2017 International Conference on, 2017, pp. 1-6.

E. Z. Haoqiang Fan, "Approaching human level facial landmark localization by deep learning," *Image and Vision Computing,* vol. 47, 2015.

J.B. Pittenger, "Aging Faces and Visual-Elastic Events:Implications for Theory of Nonrigid Shape Perception," *Human Perception and Performance,* vol. 1, 1975.

Andreas Lanitis, "Toward Automatic Simulation of Aging Effects on Face Images," IEEE TRANSSCATIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 24, 2002.

Jayant Jagtap, Manesh Kokare, Human age classification using facial skin aging features and artificial neural network, Cognitive Systems Research, Volume 40, 2016, Pages 116-128.

A High Gain Patch Antenna Array for 5G Communication

Sikiru. O. Zakariyya
Department of Electrical and Electronics Engineering
University of Ilorin,
Ilorin, Kwara State, Nigeria
zakariyya.os@unilorin.edu.ng

Abdulazeez. F. Salami
Department of Computer Engineering
University of Ilorin,
Ilorin, Kwara State, Nigeria
salami.af@unilorin.edu.ng

Bashir. O. Sadiq
Department of Computer Engineering
Ahmadu Bello University,
Zaria, Kaduna State, Nigeria
bosadiq@abu.edu.ng

Ayinde. M. Usman
Department of Electrical and Electronics Engineering
University of Ilorin,
Ilorin, Kwara State, Nigeria
usman.am@unilorin.edu.ng

Muzzammil. A. Adebayo
Department of Electrical and Electronics Engineering
University of Ilorin,
Ilorin, Kwara State, Nigeria
muzzammiladebayo@gmail.com

Mubarak. A. Afolayan
Department of Electrical and Electronics Engineering
University of Ilorin,
Ilorin, Kwara State, Nigeria
afolayan.ma@unilorin.edu.ng

***Abstract*—This research proposed a 1X4 patch antenna array for 5G Communication. The proposed antenna was designed to resonate at 28GHz frequency band. CST Microwave studio was used for the simulation of the antenna by employing Roger RT 5880 LZ with thickness and permittivity 0.762mm and 1.96 respectively as the substrate. The antenna was fed using a quarter-wave transformer to match its impedance to 50 ohms. The single patch resonates at 27.9GHz frequency with; a bandwidth of 1.87GHz, and a gain of 7.79dB. The addition of another element brought about an increment in the above-mentioned values. Furthermore, the implementation of the four element array configuration brought about the significant enhancement of values for bandwidth and antenna gain with maximum achievable gain of 13.5 dB. The design procedure, structure and simulation of the antenna are discussed in this paper.**

Keywords—Array, 5G communication, microstrip patch antenna, quarter-wave transformer

I. INTRODUCTION

In recent years, the new generation of devices (smartphones, tablets), are preinstalled with data hungry and high bandwidth demanding applications which results in an exponential proliferation of wireless data traffic. Owing to the rapid growth of the wireless system, the cellular network is migrating from the conventional voice centric to a data network. It is expected that by the end of the year 2020, the amount of connected devices that will use the cellular network will be above 50 billion [1]. The 5G network is the potential candidate to satisfy the modern day data traffic requirement. The 5G wireless communication network is generally characterized by three distinct features: (i) connectivity is ever present (ii) Low latency (iii) High speed of data transfer [2]. The 5G cellular system utilizes the higher frequency in the millimeter wave which can provide wider bandwidth when compared with the 3G or 4G bands. Among the existing millimeter wave spectrum, 28- GHz is broadly underlined for 5G as a result of lesser atmospheric absorptions and relatively lower attenuations. [3,4]. To fully utilize the 5G communication technology, antenna design is one of the essential issues to consider as the antenna plays a vital role in any wireless communication. The benefit offered by microstrip patch antennae which includes; lightweight, less expensive and ease of incorporating with circuitry devices brands them an appropriate option for the current generation wireless communication device requirements [5, 6]. However, the conventional patch antenna is characterized with low gain and a small bandwidth. The patch antenna performance can, however, be improved by configuring the antenna in an array fashion. Hence this work proposes a 1X4 patch antenna array for 5G communication.

Recently, the 28 GHz has been receiving a lot of attention as several researchers are focusing on developing array antenna for 5G applications. A dual feed mesh type antenna was proposed to operate at 28 GHz [7]. However, the antenna bandwidth is too small. A steerable millimeter wave antenna array was designed for 5G smartphone application [9]. This antenna works within the 26-32 GHz frequency range and has a bandwidth of 6GHz and achieves a maximum directivity of 13dBi. A compact notch antenna which is based on aperture couple slot and microstrip feeding was designed in [10]. Though the antenna was compact, the achieved bandwidth is low. A compact dual-band antenna resonating at 28 GHz for 5G operation was designed in [11]. The antenna has narrow bandwidth and low gain with a maximum achievable gain of 6.6 dB. In [14], an EBG (Electromagnetic Band Gap) supersubstrate, 2 element array was design for future 5G. The antenna resonates at three frequency bands (25 GHz, 28 GHz and 38 GHz) with a maximum achievable gain of 12.5 dB and a bandwidth of 1.02 GHz. A 2 dimensional compact slotted antenna

consisting of 8 elements array was proposed for 5G applications. The antenna has a maximum achievable gain of 9 dB and a bandwidth of 1.38 GHz [15]. An elliptical shape slotted circular patch antenna array resonating at dual frequency band was proposed in [16]. The antenna has a bandwidth of 1.2 GHz and achieves highest gain of 13.5 dB. A millimeter wave slotted antenna array working at 28GHz was proposed for 5G applications [8]. The element array achieves a highest bandwidth of 1.38 GHz and maximum gain of 9 dB. In [12], a 1X2 rectangular patch antenna operating at 28GHz was designed for 5G application. The antenna attains a maximum achievable gain of 10.7dB. The performance of this antenna can, however, be improved by adding more elements to the array configuration as presented in this paper. Hence, the objective of this paper is to improve the work of [12] in terms of achievable gain appropriate for 5G communication.

In this work, a 1X4 microstrip patch antenna operating at 28 GHz is proposed. The four elements array gain is increased as compared to one and two elements with a maximum gain of 13.5 dB. The remaining part of the document is arranged accordingly: in section II, the design and analysis of the proposed antenna is discussed, the simulated result of the one element, two elements and four elements arrays are discussed in section III while section IV concludes the paper.

II. ANTENNA DESIGN

In the design of a patch antenna, it is essential to first select the substrate to be used since its parameters such as the thickness and dielectric constant have a great influence on the antenna characteristics. This design employed roger RT 5880LZ with h= 0.762mm thickness and a relative dielectric constant ε_r = 1.96. Copper material is used as the patch conductor and ground plane. Popular models such as transmission line model, cavity model, and the full wave model are employed in analysing microstrip patch antenna [5, 13]. Among the above-mentioned models, it is said that the transmission line model is the easiest [13]. The dimensions of single patch element are calculated using transmission line model accordingly:

$$W_p = \frac{c}{2f_o\sqrt{\frac{2}{\varepsilon_{rl}+1}}} \quad (1)$$

The patch width is w_p, ε_{rl} is the relative dielectric constant and f_o represents the frequency at which the antenna resonates.

The effective dielectric constant is expressed as:

$$\varepsilon_{reff} = \frac{\varepsilon_{rl}+1}{2} + \frac{\varepsilon_{rl}-1}{2}\left(1+\frac{12h_s}{w_p}\right)^{\frac{-1}{2}} \quad (2)$$

As a result of fringing, the length of the patch is extended by a distance equal to:

$$\Delta L_p = 0.412\,h_s \frac{(\varepsilon_{reff}+0.3)(^{W_p}/_{h_s}+0.264)}{(\varepsilon_{reff}-0.258)(^{W_p}/_{h_s}+0.8)} \quad (3)$$

h_s is the substrate height.

The length is determined by:

$$L_p = L_{eff} - 2\Delta L_p \quad (4)$$

$$L_{eff} = \frac{c}{2f_o\sqrt{\varepsilon_{reff}}} \quad (5)$$

The edge impedance $Z_{in} = 90\frac{\varepsilon_{rl}^2}{\varepsilon_{rl}-1}\left(\frac{L_p}{w_p}\right)^2$ (6)

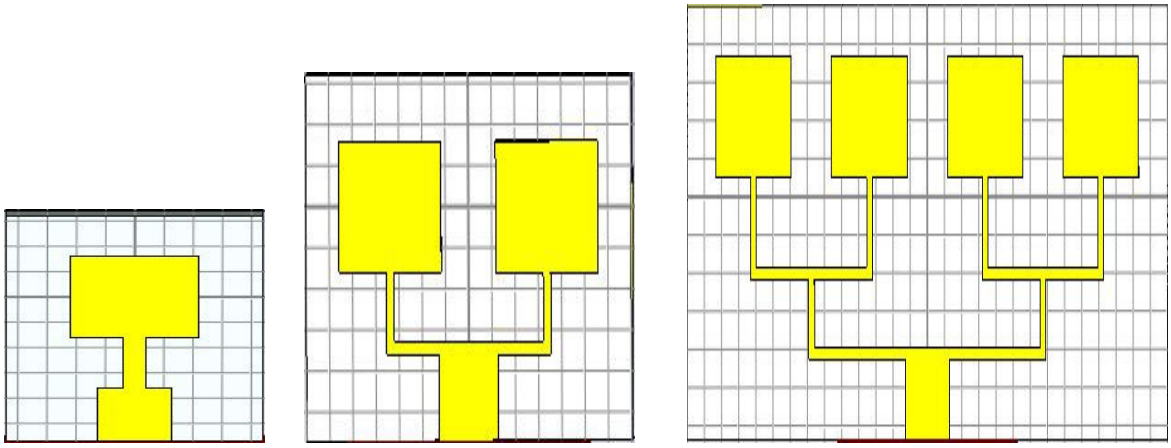
The line impedance of the quarter wave is given by:

$$Z_T = \sqrt{Z_0 Z_{in}} \quad (7)$$

The length (L_p) and width (w_p) of the radiating patch for the design was calculated to be 3.2 mm and 4.4 mm using the mathematical equations described above. The load impedance at the edge of the patch is calculated to be approximately 190.5 ohms using Eq. (6). A simulation model for the single element antenna implemented in CST-MWS is shown in Fig. 1(a).

The 1 **X** 2 array arrangement is obtained by replicating the structure of the single element patch in order to have an antenna with two patches joined together by a 100 ohms line. This arrangement enables to improve the antenna performance. The spacing between each element is 0.63λ [12]. A simulation model for the two elements antenna implemented in CST-MWS is shown in Fig 1(b).

For further enhancement of antenna performance, an arrangement of four elements array is implemented. The 1 **X** 4 array arrangement is obtained by following the same procedure used in the design of the 1 **X** 2 with the same spacing between each element. A simulation model for the four elements arrays implemented in CST-MWS is shown in Fig. 1(c).



III. (b) (c)

Figure 1: Schematics of the designed antenna (a) Single patch (b) 2 elements array (c) 4 elements array

III. SIMULATION RESULT AND DISCUSSIONS

The antenna proposed was simulated using CST Microwave Studio. The result of the simulated coefficient of reflection for the single element patch is shown in Fig. 2(a). The antenna resonates with a return loss of approximately –23dB with a bandwidth of 1.787GHz at 27.9GHz. It has been noted from the radiation pattern plot in Fig. 2(b), that the antenna has an achievable gain of 7.9 dB which can be increased by configuring the antenna in an array fashion. Fig. 3(a), shows the return loss plot of the two element arrays. It was noted that the antenna resonates at 2 frequencies (28GHz and 32GHz) with an estimated bandwidth of 8.2GHz within the 26.5GHz-34.7GHz frequency range. The gain of the antenna was found to be 10.8 dB as observed from the radiation pattern chart shown in Fig. 3(a), which is greater than the gain of a single element patch. The result of the simulated coefficient of reflection for the four elements patch is shown in Fig. 4(a). It was noted that the antenna resonates at 3 frequencies (28GHz, 32GHz and 34GHz) with a maximum bandwidth of 5.2 GHz in the range of 27 GHz - 32.2 GHz. The radiation pattern plot is shown in Fig. 4(b), which indicates that the realized gain has increased to 13.5 dB which is suitable for the 5G application [16].

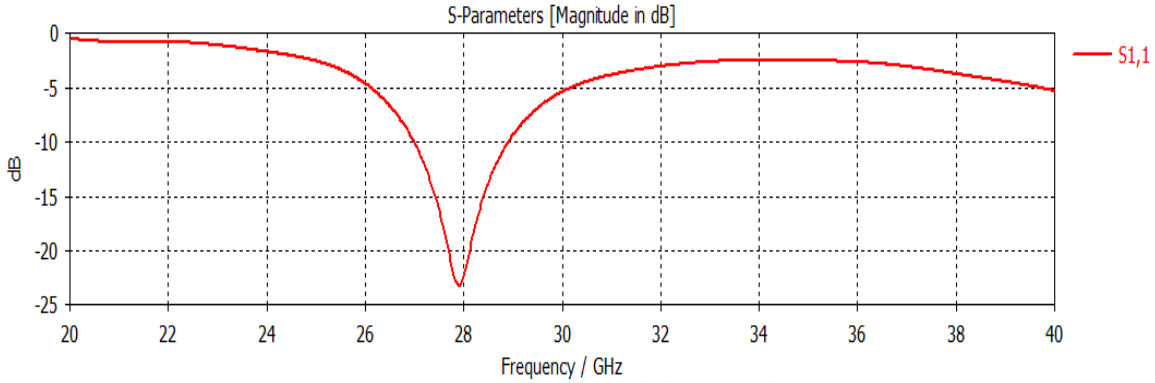


Fig. 2(a): Return loss plot of a single element patch antenna

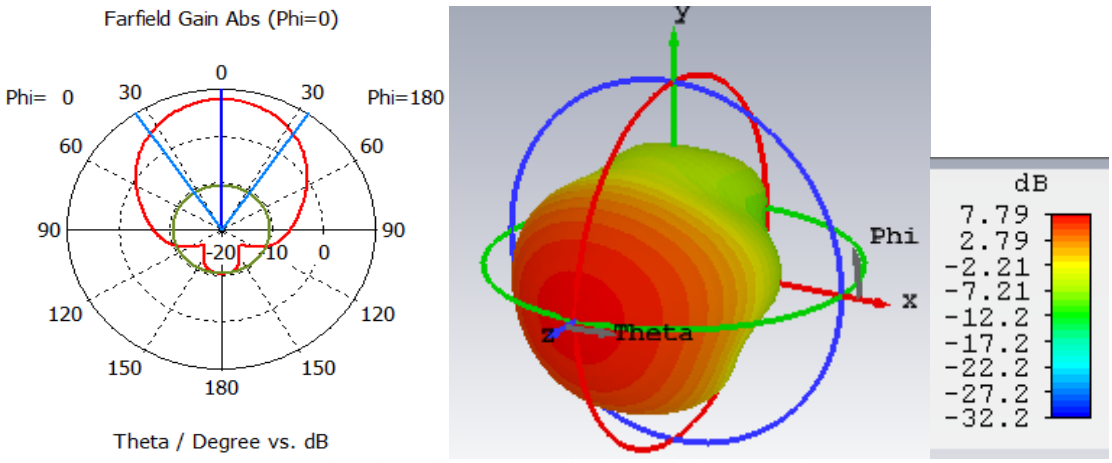


Fig. 2(b): Radiation pattern plot of a single element patch antenna

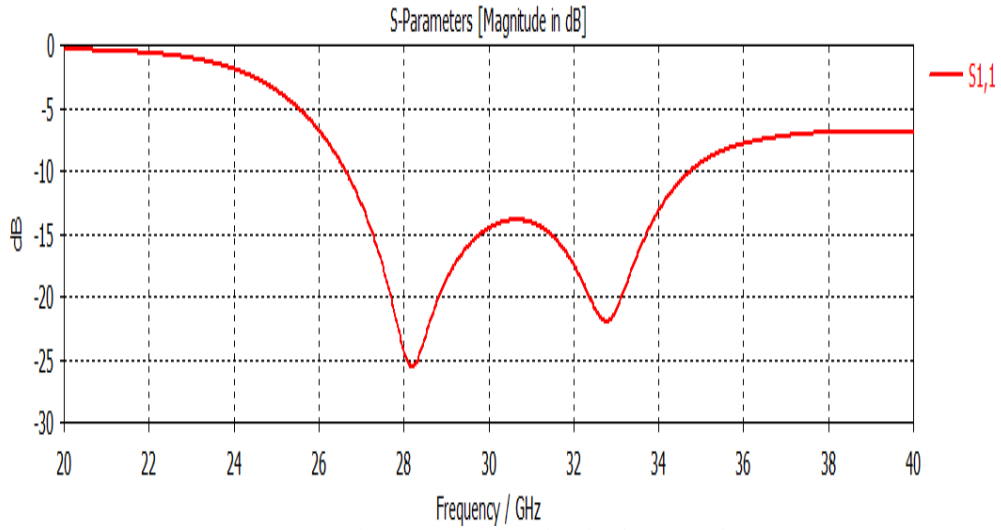


Fig. 3(a): Return loss plot of 2 element patch antenna array

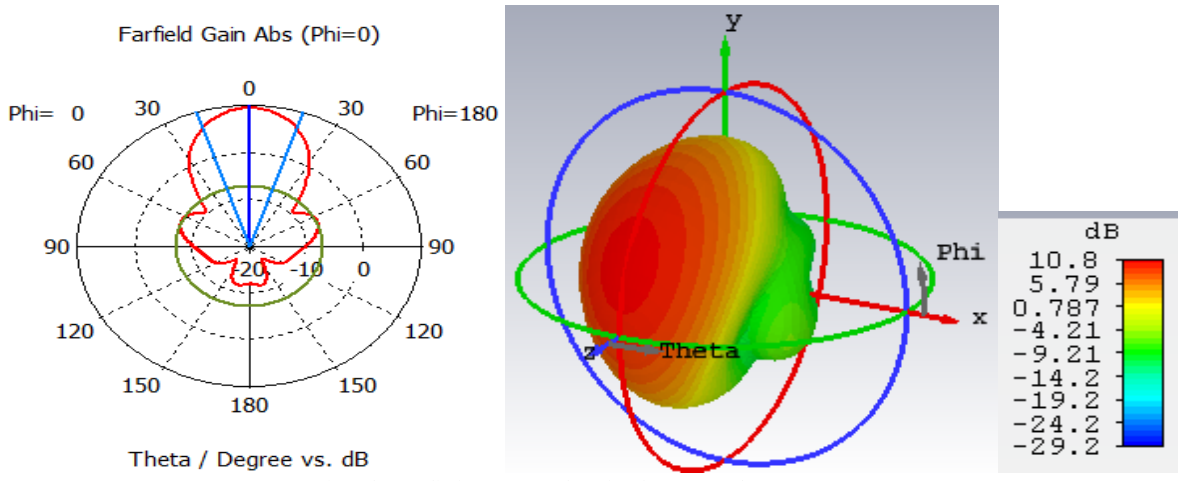


Fig. 3(b): Radiation pattern plot of 2 element patch antenna array

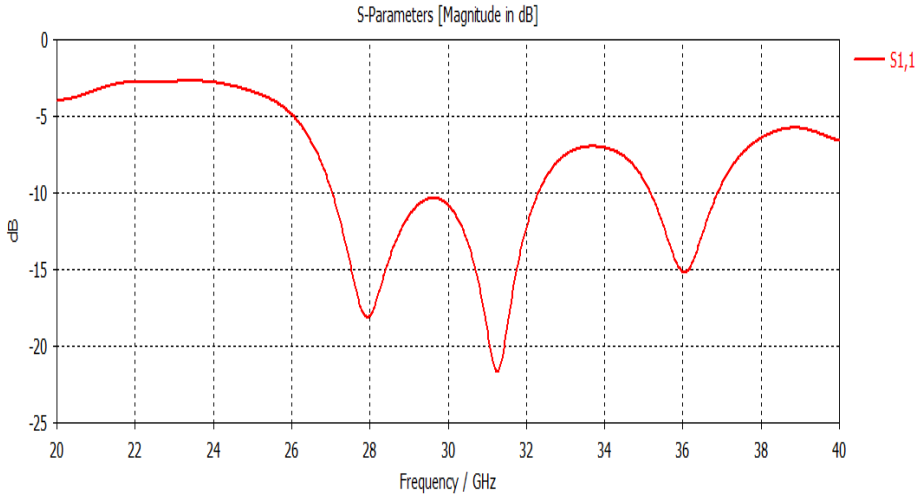


Fig. 4(a): Return loss plot of 4 element patch antenna array

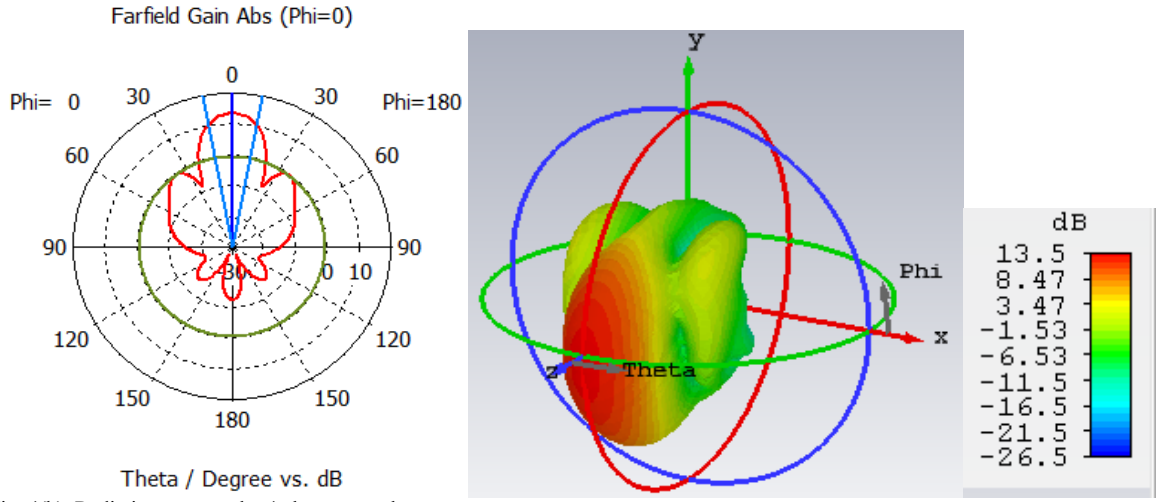


Fig. 4(b): Radiation pattern plot 4 element patch antenna array

TABLE I: COMPARISON BETWEEN THE PROPOSED DESIGN AND OTHER REPORTED ANTENNAS

Antennas	Number of array elements	Antenna array size (mm)	Bandwidth (GHz)	Gain (dB)
Proposed design	4	28x11x0.762	5.2	13.5
Reference [12]	2	----	8.9	10.7
Reference [14]	2	25x27x0.508	1.02	12.5
Reference [15]	8	29.9x28.7x0.13	1.38	9
Reference [16]	4	31x7x0.578	1.2	13.5

Table I summarize the comparison between the proposed design and other reported 5G array antennas. From the table, the proposed 5G antenna array antenna is found to perform better than the majority of the counterparts in the aspect of size, bandwidth and gain.

IV. CONCLUSION

Microstrip patch antenna has been designed for 5G application in mobile communication. The antenna was designed to resonate at 28GHz. The single element patch antenna has a bandwidth and gain of 1.87GHZ and 7.9dB which is not sufficient for the future 5G communication requirements. The antenna was configured in an array fashion to improve the gain and bandwidth suitable for the 5G communication. With the implementation of two elements array, antenna performance was improved as the gain increased to 10.8 dB while the bandwidth increased to 8.2 GHz. These values are seen to be better, compared to those obtained when a single element was used. In order to further improve the antenna gain, a configuration for the arrangement of four elements was implemented. This yielded improved values for gain with an achievable gain of 13.5 dB.

REFERENCES

More than 50 billion connected devices, Ericsson, white paper, 2011. Available at: https://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf

N. Panwar, S. Sharma, and A. K. Singh, "A survey on 5G: The next generation of mobile communication," Physical Communication, vol. 18, pp. 64-84, 2016..

T. S. Rappaport, S. Shu, R. Mayzus, Z. Hang, Y. Azar, K. Wang, and F. Gutierrez "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," IEEE Access, vol. 1, pp. 335–349, 2013.

Q. Zhao and J. Li, "Rain attenuation in millimeter wave ranges," in 7th Int. Symp. Antennas, Propag. & EM Theory, Oct. 2006, pp. 1–4.

O. S. Zakariyya, B. O. Sadiq, O. A. Abdulrahman and A. F. Salami, "Modified edge fed Sierpinski carpet miniaturized microstrip Patch antenna," Nigeria Journal of Technology, vol 35, no. 3, pp. 637-641, 2016.

O. S. Zakariyya, B. O. Sadiq, A. A. Olaniyan and A. F. Salami, "Dual Band Fractal Antenna Design for Wireless Application" Computer Engineering and Applications Journal, vol. 5, no. 3, pp. 101-108, 2016.

W. Hong, K.H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices,"'IEEE Commun. Mag, vol. 52, no. 9, pp. 63-69, Sep. 2014.

M. M. M. Ali, O. Haraz, S. Alshebeili, and A. R. Sebak, "Broadband printed slot antenna for the fifth generation (5G) mobile and wireless communications," 17th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM), 1–2, Montreal, QC, 2016.

M. Stanley, Y. Huang, T. Loh, Q. Xu, H. Wang, and H. Zhou, "A high gain steerable millimeter-wave antenna array for 5G smartphone applications," 11th European Conference on Antennas and Propagation (EUCAP), 2017

J. Helander, K. Zhao, Z. Ying, and D. Sjöberg, "Performance analysis of millimeter-wave phased array antennas in cellular handsets," IEEE Antennas Wireless Propag. Lett, vol. 15, pp. 504- 507, Jul. 2015.

Y. Jandi, F. Gharnati, and A. Oulad Said, "Design of a compact dual bands patch antenna for 5G applications," 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2017

S. U. Rahman, Q. Cao, I. Hussain, H. Khalil, M. Zeeshan, and W. Nazar, "Design of rectangular patch antenna array for 5G wireless communication," 2017 Progress In Electromagnetics Research Symposium - Spring (PIERS), 2017

C. A. Balanis, Antenna Theory: Analysis and Design, 3rd ed. New York: Wiley-Interscience, 2005.

Dellaoui, A. Kaabal, M. E. Halaoui, and A. Asselman, "Patch array antenna with high gain using EBG superstrate for future 5G cellular networks," Procedia Manufacturing, pp. 463–467, 2018.

A. T. Alreshaid, O. Hammi, M. S. Sharawi, and K. Sarabandi, "A compact millimeter-wave slot antenna array for 5G standards," 2015 IEEE 4th Asia-Pacific Conference on Antennas and Propagation, 2015.

M. I. Khattak, A. Sohail, U. Khan, Z. Ullah, and G. Witjaksono, "Elliptical Slot Circular Patch Antenna Array with Dual Band Behaviour for Future 5G Mobile Communication Networks" Progress In Electromagnetics Research C, vol. 89, pp. 133–147, 2019.

Crone Controller Based Speed Control of Permanent Magnet Direct Current Motor

Idris Garba
Department of Agricultural and Bio-environmental Engineering Technology,
Samaru College of Agriculture, Division of Agricultural Colleges,
Ahmadu Bello University,
Zaria, Kaduna State Nigeria.
(gidris21@gmail.com)

Yusuf Abubakar Sha'aban, Mohammed Bashir Mu'azu and Zaraddeen Haruna
Department of Computer Engineering, Ahmadu Bello University,
Zaria, Kaduna State Nigeria.
(yusufshaaban@gmail.com, muazumb1@yahoo.com, elzet2007@gmail.com)

Abstract - **This work aimed at developing a CRONE Controller based Speed Control of Permanent Magnet Direct Current Motors in order to achieve a robust and more effective speed control. Torque is one of the fundamental factors that affect PMDC motor's speed control, which generate uncertainty known as disturbance torque on shaft of the motor due to increase in temperature and decrease in frequency of the system. Proportional Integral Derivative (PID) controllers are commonly used in the speed control of PMDC motors, but PID controllers often do not fully reject external disturbances in the system. As such, the CRONE controller-based speed control of PMDC motors was developed to address the limitations of PID control. To achieve this, a PMDC motor model with the following system response characteristics was adopted; settling time of 6.2s, rise time of 0.79s and overshot of 30.4%. A second generation CRONE controller was developed and then applied for speed control of the PMDC motor. The performance of the developed controller was then compared with that of a Proportional Integral Derivative – Particle Swamp Algorithm (PIDPSO) controller for speed control of PMDC motor. The result of the second generation CRONE controller for the speed control of PMDC motor obtained a settling time of 0.018s and overshoot of 0.0001%. The PID-PSO controller-based speed control of the PMDC motor obtained a settling time and overshoot of 0.2s and 1.87% respectively. The effectiveness of CRONE controller was visualized in frequency domain, with the stability margin of 34.5dB and 90^0. Simulations were carried out using MATLAB 2016a.**

Keywords: pmdc motor, disturbance torque, speed control, crone controller, pid-pso

I. INTRODUCTION

The PMDC motors are generally utilized in different mechanical applications and industrial areas in fields going from toys to rocket because of their adaptability, high unwavering quality, and moderately ease. A standout amongst the most widely recognized methods used to control the position and speed of a PMDC engine is the PID controller, since its straightforward structure and control calculation (Charles *et al* 2015).

DC motors are specifically well known in high power and exact servo applications because of their reason-capable expense and simplicity of control Permanent Magnet DC motors (PMDC) are employed in a number of applications, from battery fueled gadgets like wheelchairs and power devices, to transports and entryway openers, welding gear, X-beam and tomographic frameworks, and siphoning hardware. They have the capacity to create high torque at low speed which make them reasonable substitutes for gear motors in numerous applications. (Grignion, 2012). PMDC uses permanent magnet to produce field flux. It has a great torque which aid in effective speed control (Gerada *et al*., 2014). It is a strong, dependable, economical and easy to use electromechanical machine. Its tractability has permitted PMDC motors to be conventional in today's society through numerous applications such as automation, modern supermarket and lift. Modern vehicles comprise several PMDC drives which control numerous automobiles moving parts act, and offer additional comfort besides relaxation of motorist or customers (Geest, *et al*., 2015). Considering the applications of PMDC motors, it shows clearly that research on the design and control of the motor is of great importance. Permanent magnet motors emerge over other DC motors since they don't require an outside field circuit, subsequently, dispensing with misfortunes in the field copper circuit (Bon, *et al*., 2016).

Motor speed control is achieved using armature current or field current. Numerous algorithms for position and speed control of the PMDC drive exist. Considering its remarkable speed control features, speed control of PMDC attracts substantial research which prompted the development diverse methodologies. A standout amongst the most

broadly utilized strategy for position and speed control of PMDC motor is Proportional Integral Derivative (PID) controllers (Aravin *et al*.,2017).

A lot of factors may disturb PMDC motor's performance, notable amongst them are the environmental loading conditions and working temperature, which plays an important role. For instance, a variation in the motor's loading conditions produces a disturbance torque which causes equal variation in armature current and speed respectively. Information about the disturbance torque is usually used to achieve control of PMDC motors (Zwysig *et al*., 2009). Similarly, disturbance torque is generated as a result of the motor temperature increase producing a variation in resistive circuit of the motor. (Grignion, 2012).
PID control has been broadly used to solve the PMDC motor speed control issue. In any case, PID controller may not be adequate in taking care of the disturbance torque. It frequently results in long settling time and extensive system overshoot. Consequently, there is the need for improvement of the performance achieved with PID controllers and a further developed control plan might be required. Consequently, a CRONE based controller speed control of PMDC Motor is produced to address the restrictions of PID control. This makes it more suitable for control of PMDC Motor.

II. CRONE CONTROL TECHNIQUES

The CRONE controller is a fractional order robust control using fractional order derivatives as high-level design parameters. This controller is frequency domain method based on unity feedback arrangement which allows liner robust control system to be simple rationally designed. CRONE is the French acronym of 'Commande Robuste d'Ordre Non Entier' were given in the second half of 1970s, when Oustaloup synthesized fractional order controllers for strong control of continuous colorant lasers (Oustaloup, 1975; Oustaloup & Coiffet, 1983; Sabatier *et al.*, 2002).
The CRONE controller is structured dependent on the normal solidarity input design (Figure 2.4), which the vigorous controller or open-circle exchange work is characterized utilizing vital separation with non-whole number (or fragmentary) arrange (Lanusse et al., 2008). The required strength is that of both dependability edges and execution and especially the heartiness of the pinnacle esteem (called full crest) of the corresponding affectability function (Lanusse et al., 2013).

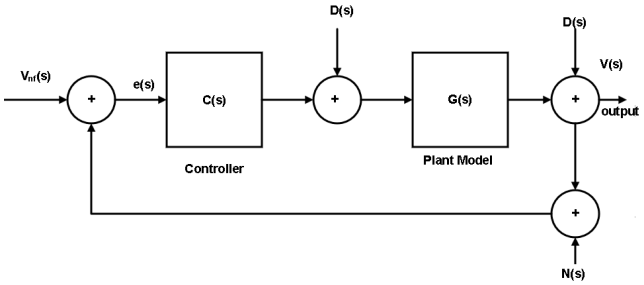


Figure 2.5: Common CRONE Control System Diagram (Lanusse *et al.*, 2008)

From Figure 2.4, the output response, error signal and the controller effect are represented by the following equations (Morand *et al.*, 2015):

Output:

$$V_s(s) = -T(s)N_m + GS(s)D_u(s) + T(s)V_{ref}(s) \quad (2.10)$$

error:

$$\varepsilon(s) = -S(s)N_m(s) - GS(s)D_u(s) + S(s)V_{ref}(s) \quad (2.11)$$

control effort:

$$U(s) = -CS(s)N_m(s) - T(s)D_u(s) + CS(s)V_{ref}(s) \quad (2.12)$$

The stability margins are the phase and gain margins, but also the resonant peak of the output equation (2.37) indicate the closed loop response $V_s(s)$ relate to the loop input reference signal $V_{ref}(s)$ and in the regulation mode the closed-loop response $V_s(s)$ to a disturbance $D_T(s)$ on the plant output as depict on the Figure 2.4 (Lanusse *et al.*, 2011). The peak value of the magnitude of the complementary sensitivity function $T(s)$ (with $s = jw$), is defined from resonant peak M_r is revealing of the overshoot first percent for the step response in tracing or in regulation. Indeed, this peak and overshoot are strongly correlated (Lanusse *et al.*, 2011). The damping proportion is characterized as the damping of the oscillatory method of these reactions.
CRONE control method has continued to evolve since the 1990's and now has three different generation design methods that are successively extending the application field as follows (Lanusse *et al.*, 2013):

A. First generation CRONE control strategy

This method is an open-circle gain crossover frequency dependent on a steady stage controller got from band-restricted genuine fractional differentiation. In this, the plant uncertainty is increase like and ostensible irritated plant stages are steady (CRONE Research Group, 2010). It is characterized inside a frequency scope of about the crossover frequency from the fractional exchange capacity of a request indispensable differentiation.

$$C_F(s) = C_0 s^n \quad (2.13)$$

This technique is an open-circle gain crossover frequency $\frac{n\pi}{2}$ subject to a consistent stage controller got from band-confined authentic fractional differentiation. In this, the plant uncertainty is increment like and apparent aggravated plant stages are consistent (CRONE Research Group, 2010). It is described inside a frequency extent of about the crossover frequency from the fractional trade limit of a demand vital differentiation.

$$C_F(s) = C_0\left(\frac{1+s/w_1}{1+s/w_h}\right)^n \quad (2.14)$$

with $\omega_l < \omega_A$ *and* $\omega_h < \omega_B$

For the CRONE controller need order-n_1 band-limited to handle the steady state error, and provide a control effort to level band-limited integrator with a order - n_F low-pass

filter. The first generation CRONE controller transfer function is defined as (CRONE Research Group, 2010):

$$C_F(s) - C_v\left(\frac{W_1}{s}+1\right)^{n_1}\left(\frac{1+s/W_1}{1+s/W_h}\right)\frac{1}{(1+s/W_h)^{n_h}} \quad (2.15)$$

B. Second generation CRONE control strategy

This is connected to a framework, when varieties of the plant are increase like around crossover frequency ω_{cg}, the control dropped the plant phase variety, as for frequency. The controllers give a consistent open-loop phase (genuine fractional-arrange combination) whose Nichols locus is an opposite straight line called a frequency layout as appeared in Figure 2.5. This additionally speak to what Bode, in 1945, accepted could be conceivable (Bode, 1945). This model ensures the quality of phase and modulus edges and of booming peaks of corresponding sensitivity and sensitivity capacities.

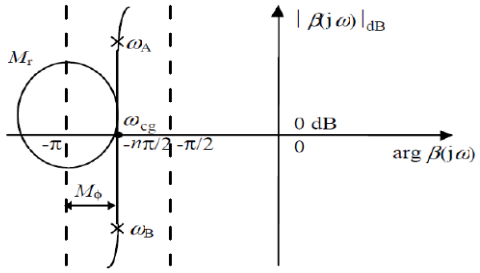


Figure 2.6: Frequency Template whose Nichols Locus is a Vertical Straight Line (Sabatier *et al.*, 2002)

The second generation CRONE approach describes the open-loop transfer function about the crossover frequency (W_{cg}) as the transfer function of a fractional-order integrator, as follows (Lanusse *et al.*, 2013)

$$\beta_F(s) = \left(\frac{W_{cg}}{s}\right)^{n_1} with\ n \in \mathbb{R} \quad (2.16)$$

For the nominal open-loop transfer function for the CRONE second generation controller to handled accuracy, robustness and control effort problems, using integer order n_l and n_h, is defined by (Lanusse *et al.*, 2013):

$$\beta(s) = k_0\left(\frac{1+\frac{s}{W_1}}{\frac{s}{W_1}}\right)^{n_1}\left(\frac{1+\frac{s}{W_h}}{1+\frac{s}{W_1}}\right)^{n}\frac{1}{\left(1+\frac{s}{W_h}\right)^{n_h}} \quad (2.17)$$

Where k_0 is a gain; w_l and w_h are the transitional low and high frequencies, respectively; n_l is the order at low frequencies; n is the order around crossover frequency w_u; n_h is the order at high frequencies;

When the nominal open-loop transfer is determined, the fractional controller $C_F(s)$ is defined by its frequency response (Lanusse *et al.*, 2013):

$$C_F(jw) = \frac{\beta(jw)}{G_0(jw)} \quad (2.18)$$

whose phase is variable and where $G_0(jw)$ designates the nominal frequency response of the plant.

The iterative conveyance of poles and zeros of the entire open circle fractional integrator exchange function β(s) results in sound request open circle fractional integrator exchange function $\beta_R(jw)$. Beginning with this attainable normal order Second generation CRONE controller $C_R(jw)$ is attained as (Lanusse *et al.*, 2013):

$$C_R(jw) = \left(\frac{\beta_0}{G(jw)}\right) \quad (2.19)$$

The aim of both the first and second generation CRONE controllers is to be robust to plant gain variations. Nevertheless, other types of model uncertainty, such as pole and zero misplacement are not considered. This inspired growth of third generation CRONE controller (Lanusse *et al.*, 2013).

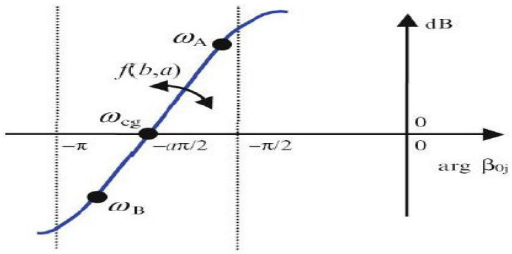
C. Third generation CRONE control strategy

The third generation CRONE controller is applied when the system frequency uncertainty domains are not gain-like. Besides they usually use complex non integer order integration over a selected frequency range $[\omega_A, \omega_B]$. This technique was employed to design controllers for plants by means of lightly damped modes (Oustaloup *et al.*, 1995) digital controllers and controllers for nonlinear systems whose nonlinear behaviors are taken into account by sets of equivalent linear behaviors (Pommier *et al.*, 2002).

The principle of the third generation CRONE control strategy consists of different stages namely optimization of open loop behavior, generalized template and optimal template.

i. Generalized template

In second order CRONE generation control strategy, the vertical straight line Nichols locus is replaced by any straight line portion called the summed up format appeared in Figure 2.6. The summed up layout is normally characterized by complex fractional order integration of order n, whose genuine part chooses its stage area at



recurrence, that is and whose fanciful part at that point chooses its situation to the vertical (Lanusse et al., 2013).

Figure 2.7: The Generalized Template (Lanusse *et al.*, 2013)

The complex fractional order integral transfer function is given by Equation 2.20 (Lanusse *et al.*, 2013):

$$\beta(s) = \left(cosh\left(b\frac{\pi}{2}\right)\right)^{sign(b)_a}\left(\frac{\omega_{cg}}{s}\right)^{a}\left((\omega_{cg})^{ib}\right)^{sign(b)} \quad (2.20)$$

$$with\ a + ib \in Y\ and\ s \in Y\ j.$$

But the complex fractional orders are respectively time domain and frequency domain complex plane. Mounting the equation 2.20 offers equation 2.21 (Lanusse *et al.*, 2013):

$$\beta(s) = \left(\left(\cosh\left(b\frac{\pi}{2}\right)\right)^{\sin h(b)_2}\left(\frac{\omega_{cg}}{s}\right)^{a}\left(\cos\left(b\ln\left(\frac{s}{\omega_{cg}}\right)\right)\right)^{\sin h}\right)^{\sin h} \quad (2.2)$$

by differentiating the $\beta(j\omega)$ magnitude and phase at frequency ω_{cg}, the angle of the generalized template to the vertical can be conveyed as a function of a and b (Lanusse *et al.*, 2013):

$$\frac{d(|\beta(jw)|_{dB})}{d(phase\beta(jw))} = \frac{-20a\sin h(b)}{ln(10)b\cos h\left(b\frac{\pi}{2}\right)} \quad (2.22)$$

The generalized band-limited template transfer function equation 2.22, is changed with a new general expression (Lanusse *et al.*, 2013):.

$$\beta(s) = e^{a\sin h(b)}\left(\frac{\omega_h}{s}+1\right)^{a}\left(\alpha_0\frac{1+\frac{s}{\omega_h}}{1+\frac{s}{\omega_1}}\right)^{a}\times\left(\left(R_{\omega_r}\frac{1+\frac{s}{\omega_h}}{1+\frac{s}{\omega_1}}\right)^{jb}\right)\left(\frac{s}{(1+\frac{s}{\omega_1})}\right)^{a\sin h(b)} \quad (2.23)$$

Where,

$$C = \cosh\left(b\left(\tan^{-1}\left(\frac{W_{cg}}{W_h}\right)\right)\right) \quad (2.24)$$

$$\alpha_0 = \left(\left(1+\left(\frac{\omega_r}{\omega}\right)^2\right)^{1/2}\right) \quad (2.25)$$

$$|b| < min\left(\frac{\pi}{2ln(\alpha_0)}, \frac{\pi}{2|ln(\alpha_0\frac{\omega_h}{\omega_1}|}\right) \quad (2.26)$$

Hence the third generation CRONE controller general form of is given by equation (2.27) (CRONE Research Group, 2010):

$$C(s) = \frac{\beta(s)}{G(s)} \quad (2.27)$$

ii. Optimal template

In CRONE control approach, it deals stability margin and performance, particularly the closed loop resonant peak M_r in the Nichols magnitude contour. If M_{rd} is the resonant peak for the nominal plant $G_0(jw)$, undefined quantity of open loop Nichols loci is tangent as shown in Figure 2.7, optimal template is selected from the M-contour for the nominal parametric state of the plant tangents the M_{rd} M-contour around echoing frequency $(\omega_r)$for the nominal plant state and it minimizes the deviations of the phase margin for the perturbed plant states (CRONE Research Group, 2010).

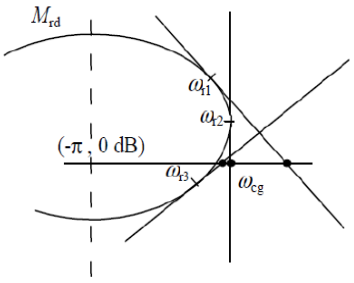


Figure 2.8: Infinite Number of Generalized Template Tangents (CRONE Research Group, 2010)

Figure 2.8 shows the optimal template positions of the open-loop uncertainty domains properly, so that they overlay the low stability margin areas as little as possible.

iii. Optimization of open loop behavior

The open-loop transfer function $\beta(s)$ has eight high level parameters as follows; $n_1, n_h, a, b, \omega_1, \omega_h, \omega_r$ and C. Where the control system designer fixed n_1 and n_h. So as to achieve the tangency condition, ω_r and C are stated. The four independent parameters of the cost function J are minimized based on resonant peak variation using nonlinear optimization algorithm and the four usual sensitivity functions $T(s), S(s), CS(s)$ and $GS(s)$ shaping constraints are fulfills using a set of Synthesizing a template through the optimization of three independent parameters (a tangency is imposed) from the four high-level parameters a, b, ω_1 and ω_h is the initial aim of the third generation CRONE control (CRONE Research Group, 2010). In order to determine the three optimal values from the following four parameters we optimized the generalized template (Figure 2.9) CRONE Research Group (2010):

i. The position of the template along the 0 Db axis is determined by optimal real integration order a.
ii. The angle to the vertical is determined using optimal imaginary integration order b which then determines its angle to the vertical;
iii. The length is determined using optimal corner frequencies ω_1 and ω_h

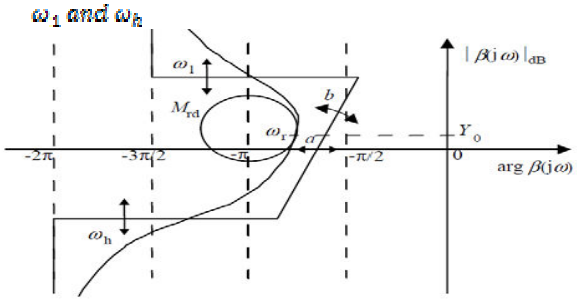


Figure 2.9: *Effects of Parametrs a, b, ω_1 and ω_h on the Asymptotic Nichols Locus* (CRONE Research Group, 2010).

The CRONE controller features are highlighted as follows:

i. The CRONE controller is based on frequency domain technique
ii. It has common unity feedback configuration.
iii. It can handle continuous time or discrete-time problem
iv. It can be applied to linear or non-linear problem
v. It can also be applied to perturbed linear time invariant system either Single Input Single Output (SISO) or Multi Input Multi Output (MIMO)

III. ROBUST CONTROL AND PERFORMANCE METRICS

Design of a controller which guaranteed some level of performance of the controlled system irrespective of the changes in the plant or process dynamics surrounded by a predefined class and the stability is guaranteed (Williams, 2007). To determine the response behaviour of the two controllers (PID and CRONE) and their robustness to unexpected shock inputs, time-domain analysis is carried out.

The dynamic system response to an input usually stated as a function of time. In time-domain analysis of such systems, the time response can be computed if the nature of the input is known and the mathematical model of the system can be

determined. The time response basically has two components: transient response and the steady-state response. Transient response is defined as that part of the system response that goes to zero as time goes to infinity and is dependent upon the system poles only and not on the type of input. It is therefore sufficient to analyze the transient response using a step input. The steady-state response indicates where the system output ends up as time approaches infinity and depends on system dynamics and the input quantity (Burns, 2001).

Performance metrics are metrics used to determine the performance improvement, effectiveness, efficiency, and appropriate levels of internal controls (Alagoz *et al.,* 2015) of a system. Therefore, the performance metrics to be considered in this work are:

A. Transient response: The transient response of a control system is one of the performance metrics for determining the settling time, rise time, overshoot and the steady state error of a given system in order to compare its performance with the other control technique. An optimal system is the system that has the smallest transient response and a better steady state performance (Chen, 1995).

The transient response of a control system consists of the settling time, rise time, overshoot and the steady state error. An optimal system is the system that has the smallest transient response and a better steady state performance (Chen, 1995). Integral of time multiplied by absolute error (ITAE) has the largest changes as ζ varies, therefore, has the best selectivity compared to integral of absolute error (IAE) and integral of square error (ISE) performance indices, in order for the system to track the reference input with a better transient performance. ITAE also yields a system with a faster response than the other criteria (Chen, 1995).

A limitation on the impelling sign or on the transfer speed of the subsequent frameworks must be forced; else, it is conceivable to structure a general framework to have an exhibition record as little as attractive, and the subsequent activating sign will approach unendingness (Andinet, 2011). The issue is to locate an ideal implementable with the ideal following execution as in (Chen, 1995):

$$\lim_{t\to\infty}|y(t)-r(t)| = 0 \qquad (2.28)$$

where $y(t)$ is the plant output, and $r(t)$ is the reference input. ITAE is defined by (Chen, 1995):

$$J = \int_0^{\infty} t|e(t)|dt \qquad (2.29)$$

where $e(t)$ is the error between the reference input and the plant output at time t.

The transient-response characteristics of a control system to a unit step input are usually specified by the following: Peak overshoot, M_p, peak time, t_p, settling time, t_s, steady-state error, e_{ss}, rise time, t_r and delay time, t_d as in Figure 2.10 (Rahim, 2009).

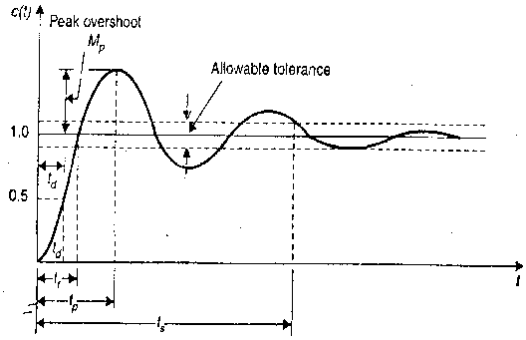


Figure 2.10: Transient Response Characteristics of a System (Rahim, 2009).

a. **Delay time (t_d):** Time required for the response to reach within 50% of the final steady-state value
b. **Rise time (t_r):** Time required for the response to rise from 0 to 90% of the final steady-state value
c. **Peak time (t_p):** Time required for the peak overshoot of the response to occur
d. **Settling time (t_s):** Time required for the response to reach and stay within 2% or 5% of its final steady-state value.
e. **Peak overshoot (M_p):** Normalized difference between the time response peak and the steady output and is defined as:

$$\%M_p = \frac{c(t_p)-c(\infty)}{c(\infty)} \times 100\% \qquad (2.30)$$

Where $c(t_p)$ is the time response peak and c(∞) is the steady output

f. **Steady-state error (e_{ss}):** Error between the actual output and desired output as 't' tends to infinity and is defined as:

$$e_{ss} = \lim_{t\to\infty}[r(t) - c(t)] \qquad (2.31)$$

A. The measures of interest in this research are peak overshoot, settling time and steady-state error

B. Robustness: System tolerates perturbations in dynamics. It is the low sensitivity to disturbances or perturbations not necessarily considered in the design and analysis of systems. If the process/plant dynamics of a robust system change, its performance is not expected to deteriorate to an unacceptable level, i.e. the system should be able to withstand those effects when executing the tasks for which it was designed. The disturbances and consequently the dynamic behaviour of a system are simulated and analysed using the standard test signals: impulse (sudden shock), step (sudden change), ramp (constant velocity) and parabolic (constant acceleration) as shown in Figure 2.11 (Rahim, 2009).

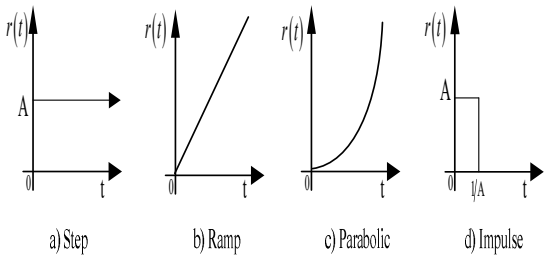


Figure 2.11: Standard Test Signals (Rahim, 2009)

The impulse signal is the choice for testing the robustness of the controllers (CRONE and PID) while the step signal is for the time response analysis in this work

C. Non-integer toolbox: This a toolbox for MATLAB got ready for creating Fractional order controllers for single-input single-yield (SISO) plant and assess their execution. Its code can be openly downloaded and adjusted. It was created because of the way that planning controllers that make utilized of fractional math request running a few calculations which are not found in the toolbox spread with MATLAB or some other normal toolbox. (Cois et al., 2002; Oustaloup et al., 2002). A great deal of these calculations considers building whole number order plant that rough somewhat the conduct of the fractional order plants. Since there are (really) numerous choices, in both the frequency and discrete time area, realizing them all effectively is an unavoidable takeoff point for checking, in significant, the relative advantages and terrible characteristics of each. Distinctive calculations concern strategies for fitting parameters: those unequivocal to fractional math ought to in like manner be realized deliberately (there is clearly no convincing motivation to execute systems like least squares improvement that may in like manner be used hence: MATLAB starting at now has respectable capacities with respect to that).

The toolbox is allowed to support scholarly utilization of fractional control, and in light of the fact that so it is simpler to benefit from recommendations and upgrades from different clients. Its code was kept as basic as conceivable to energize code reuse by different clients. It contains most understood commitments to this region from different scientists just as some from the creators' papers (Valério et al., 2002, 2003a, 2003b, 2004); fractional PIDs are, as far as we could possibly know, being methodically given in a toolbox to the first run through.

IV. SPEED CONTROL OF PERMANENT MAGNET DC MOTOR

A Speed control method

Speed control of PMDC drive has attracted considerable research owing to its outstanding speed control features. For the operation of the PMDC motor on a wide range of speed, it requires a controller for speed control of the drive to execute desired task and to eliminate the impact of disturbance torque that changes the speed. Hence, in designing speed controller for PMDC motor, the following control methods have to be considered.

i. Flux control method

Conventionally, in a DC motor Speed is conversely relative to the flux per shaft; thus, diminishing the flux, speed increments and the other way around. A rheostat is included arrangement with the field winding to direct the flux. Amassing extra resistance in series arrangement with the field winding builds the speed as it diminishes the flux. The current in the field is relatively very minor in shunt motors, cupper loss is small and, therefore, this technique is relatively effective. However, speed might be enhanced over the appraised an incentive by dropping flux with this method; it bound the most extreme speed as blurring of flux outside the breaking point will hurtfully irritate the compensation.

ii. Armature control method

The relationship between DC drive speeds with back e.m.f E_b is proportional one, difference between the applied voltage and the drop across the armature coil i.e

$$E_b = V - I_a R_a \qquad (2.8)$$

When the applied voltage V and the armature resistance R_a are retained unchanged, the speed and armature current I_a will have proportional relationship. Accordingly, including a resistance in series with the armature Ia diminishes making speed diminish too. The higher extent of the resistance in series with armature, the more it causes diminish in speed.

iii. Series- parallel control technique

B. The series-parallel control technique is widely used in electric traction, where two or more motors are mechanically coupled in series. The series connection of motor is employed when motors are operated with low speed and a parallel connection for high speed. When the motor is in series, it has the same current passing through them, while voltage across each motor is divided. For parallel connection, the voltage all through the motors is the same while the current is divided.

iv. Variable resistance in series with armature

By methods for including resistance in series with the armature, the rise in resistance will reduce voltage across the armature. And, hereafter, proportionally reduce the speed.

However, different control algorithms exist for speed control of PMDC motor. Proportional Integral Derivative (PID) controllers have been commonly used for speed control of PMDC motor (Aravind *et al*, 2017).

B. Permanent magnet DC motor Model

Permanent Magnet Direct Current Motor has a simple construction ever since the field winding of a DC motor is substituted by permanent magnet (PM). As a result, they

have turn out to be one of the most frequently used machines in many applications such as electric vehicles, robotic manipulators, trams or motion controller devices e.t.c (Cathen, 2001). Some of these applications require robustness, reliability and precision so that the process of tuning the controller parameters is of big importance because it helps in disturbance attenuation.

The design of disturbance torque controller, a suitable mathematical model needs to be first developed. A PMDC motor is considered in this research work. Circuit diagram representing the dynamics of PMDC motor is as shown in Figure 2.2.

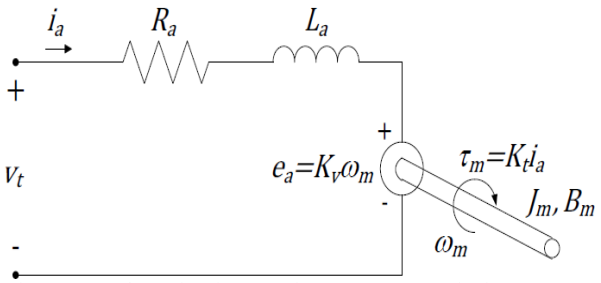


Figure 2.2: Schematic Diagram of PMDC Motor (Grignion, 2012)

From Figure 2.2 above, V_t is the input voltage, i_a is the armature current, R_a is the armature resistance, L_a is the inductance resistance, e_a is the back e.m.f, T_m is the torque, J_m is the inertia, K_b is the back e.m.f. constant, B_m is the damping ratio, K_t is the torque constant and ω_m is the angular speed. The following represent the nominal no load dynamics of PMDC motor (Grignion, 2012):

$$v_t = R_a i_a + L_a \frac{di_a}{dt} + K_v \omega_m \quad (2.1)$$

$$K_t i_a = J_m \frac{d\omega_m}{dt} + B_m \omega_m \quad (2.2)$$

Where equation 2.1 represents the electrical characteristics of the PMDC motor obtained using Kirchhoff's voltage law and the mechanical characteristic of PMDC motor is obtained using Newton's second law of motion as shown in equation 2.2.

Generally, considering i_a and ω_m to be the state variables and the system output measurement to be armature current. The following represent the state space model of the PMDC motor. Grignion (2012):

$$\dot{x} = Ax + Bu \quad (2.3)$$

$$y = Cx \quad (2.4)$$

Where

$$A = \begin{bmatrix} -\frac{R_a}{L_a} & -\frac{K_v}{L_a} \\ \frac{K_t}{J_m} & -\frac{B_m}{J_m} \end{bmatrix}, B = \begin{bmatrix} \frac{1}{L_a} \\ 0 \end{bmatrix}, \quad C = [1 \quad 0],$$

$$x = \begin{bmatrix} i_a \\ \omega_m \end{bmatrix}, \qquad u = v_t \quad y = i_a \quad (2.5)$$

If we consider the present of disturbances and uncertainties in the system in the form of a disturbance torque and a disturbance voltage, Equations 2.1 and 2.2are written as (Grignion, 2012):

$$v_d = \Delta R_a i_a + \Delta L_a \frac{di_a}{dt} + \Delta K_v \omega_m \quad (2.6)$$

$$T_d = -\Delta K_t i_a + \Delta J_m \frac{d\omega_m}{dt} + \Delta B_m \omega_m + T_L \quad (2.7)$$

Where v_d is disturbance voltage and T_d is disturbance torque Figure 2.3 shows the block diagram of PMDC Motor with disturbance torque due to electrical characteristics (v_d) and that of mechanical characteristic (T_d) of the PMDC Motor.

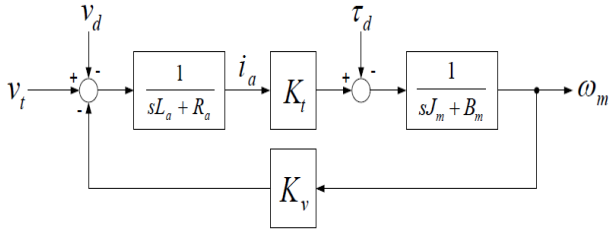


Figure 2.3: Block Diagram of PMDC Motor (Grignion, 2012)

i. Permanent Magnet DC Motor Transfer function

The transfer function of PMDC motor is obtained from the state space model equation (2.5). It is obtained by taking the Laplace transform with zero initial conditions. From equations (2.3) and (2.4) we have;

$$y(s) = [C(sI - A)^{-1} B]u(s) \quad (3.1)$$

The transfer function $\frac{Y(S)}{U(S)} = G(s)$ (3.2)

$$G(s) = \frac{\omega(S)}{V(S)} = \frac{K_t}{(sJ+B)(R+sL)+K_t K_b} \quad (3.3)$$

Where G(s) is the plant transfer function, $\omega(s)$ **is** speed of PMDC motor as output, V(s) is the input voltage, K_t is the torque constant, K_b is the back electromotive force (e.m.f.) constant, R is the armature resistance, L is the Armature Inductance, J is the Inertia and B is the damping ration. Table 3.1 shows the values of PMDC motor parameters used in this work.

Table 3.1: Permanent Magnet DC Motor Parameters

Parameter	**Value**	**Unit**
K_t	0.072	Nm/A
J	0.0005	Kgm^2/rad
B	0.00021	Nms/rad
R	0.5	ohms (Ω)
L	0.002	H
K_b	0.072	Vs/rad

Substituting the parameters of the PMDC motor we have;

$$G(s) = \frac{\omega(s)}{V(s)} = \frac{0.072}{0.002s^2 + 0.002504s + 0.06234}$$

V. CRONE CONTROLLER DESIGN

The crone controller was developed using Simulink design and crone control tool box respectively.

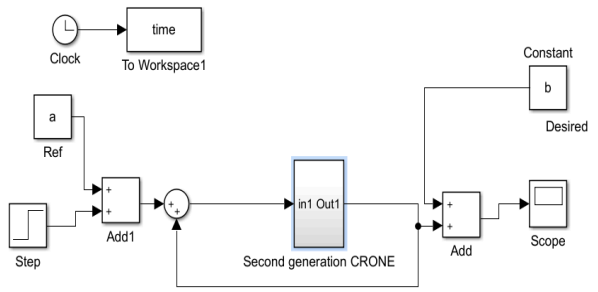


Figure 6: Simulink Diagram for Speed Control using Second Generation CRONE Controller

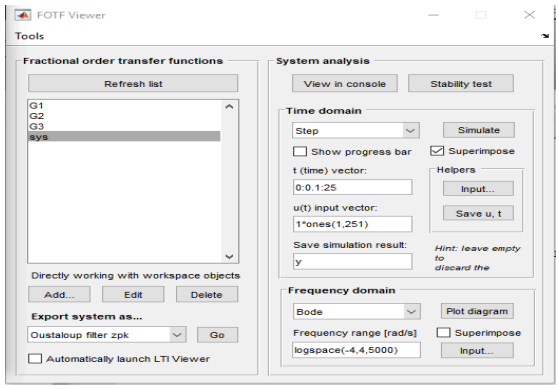


Figure 7: Implementation of Speed Control using Second Generation CRONE Controller

VI. RESULT AND DISCUSSION

C. Time Response for PMDC motor

This section shows the step response of PMDC motor obtained from the transfer function in equation (3.7).

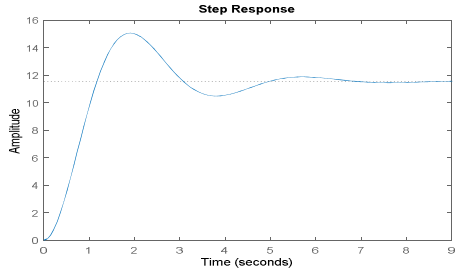


Figure 4.1: Step Response of PMDC Motor

Figure 4.1 shows the step response of the PMDC motor without controller. From the step response, the following characteristics were obtained: settling time of 6.2s, rise time of 0.793 s, steady state error of 0.0001m and overshoot of 30.4% were obtained. This clearly shows the need for a controller so that the system can meet up with the design specification (settling time of less than 1second and overshoot of less than 2%) irrespective of the external disturbance. From Figure 4.1, it is observed that the system model exceeded the desired speed of 1rad/s and which take more than 2s to settle down. It is clearly observed that the open loop test will not satisfy the desired design requirements. So, the system needs closed loop robust controller in order to achieved the design requirements.

B. Time Response for Second Generation CRONE Based Speed Controller

The response of second-generation CRONE based speed controller for PMDC motor system obtained from Simulink model of Figure 3.3. A nominal unit step signal was applied and the time response is as shown in Figure 4.3

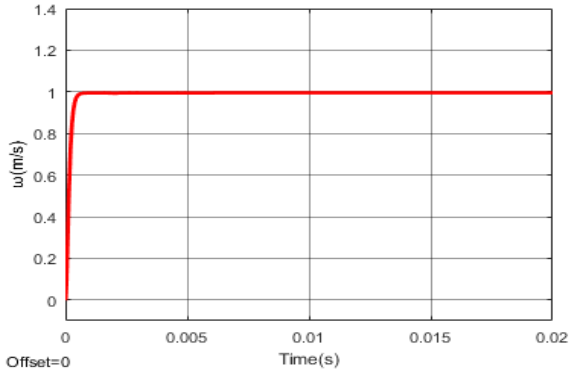


Figure 4.3: Response for CRONE controller base speed control of PMDC

Figure 4.3 shows the time response of the CRONE based speed controller, the model response and the controlled variable. The controlled variable was able to settle at 0.0018s with overshoot of 0.001%. This implies that the CRONE controller is adopted in speed control of PMDC motor to stabilize speed of the PMDC motor as well as to exhibit better dynamic behavior such as nullification of peak overshoot and get less settling time. Hence, it is suitable for control of system where robustness is desired.

C. Sensitivity of the controller

Since most disturbances and noise occur respectively at high and low frequencies, there is need to visualize the effectiveness of the CRONE controller in a frequency domain. Thus, bode plot is one of the frequency domain specification tools that determine the margin of stability with respect to both gain and phase margin. Figure 5 presents the bode plot of the sensitivity function of the PMDC system under the action of the CRONE controller.

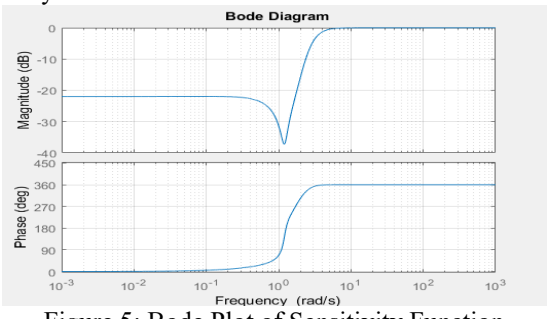


Figure 5: Bode Plot of Sensitivity Function

From Figure 5, the characteristics of the Bode plot of the sensitivity function are: Gain Margin of 34.6dB, Phase Margin of 90^0 and peak response of 0dB. Thus, since both values of the stability margins are positive and the magnitude of the sensitivity function has no peaks, it signified how second-generation CRONE controller is able to stabilize the system in the present of external disturbance. This justify that CRONE controller's robustness to disturbance rejection, which means that the effects of disturbance and the noise signals on the system are minimized.

i. Robustness of the Controller

The second CRONE controller was subjected to robustness test by building uncertain state space model and analyzing the feedback control system with uncertain element based on parameter variation. By specifying uncertain physical parameters in order to see the effect of random and worst-case parameter variation using functions (usample and robstab). The uncer J, l, R, kt and kb) were specified a 25% and 30% respectively to assess the robustness of the second generation CRONE controller.

(b) Bode plot

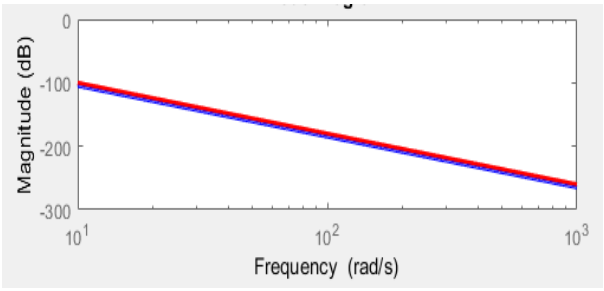
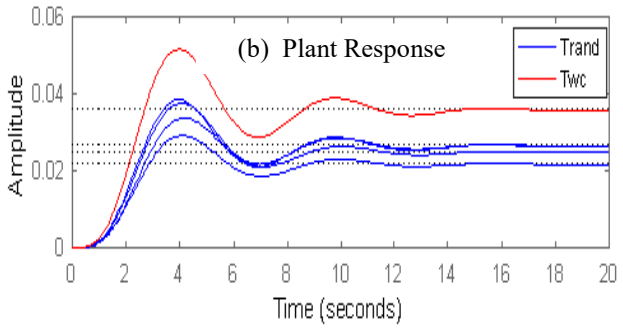


Figure 6: Bode and Plant Response

From the robustness analysis, the result shows that the system is robustly stable for the modeled uncertainty by tolerating up to 303% of the modeled uncertainty, with destabilizing perturbation amounting to 234% of the modeled uncertainty. This perturbation causes instability at the frequency 1.16rad/seconds. The sensitivity with respect each uncertain element is; 8% for B. Increasing B by 25% decreases the margin by 2%. 3% for J. Increasing J by 25% decreases the

(a) Bode Plot

reasing R by



25% decreases the margin by 13%. 16% for kb. Increasing kb by 25% decreases the margin by 4%. 12% for kt. Increasing kt by 25% decreases the margin by 3%. 1% for l. Increasing l by 25% decreases the margin by 0%. Therefore, from the analysis, the second generation CRONE controller perform robustly for the specified uncertainty on B, J, l, R, kt and kb.

VII. CONCLUSION

The development of a CRONE controller based speed control of PMDC Motor has been presented. The second generation CRONE based-controller was used for the speed control of Permanent Magnet DC Motor model with a better performance.

REFERENCES

[1] Alagoz, B.B., Deniz, F.N., Keles, C., & Tan, N. (2015). Disturbance Rejection Performance Analyses of Close Loop Control Systems by Reference to Disturbance ratio. ISA transactions, 55,63-71.

[2] Andinet, N.H. (2011). Design of Fuzzy Sliding Mode Controller for the Ball and Plate Syste. (Masters of Science Thesis), Addis Ababa Institute of Technology, Addis Ababa., Published.

[3] Aravind M. A., Niranjan Saikumar, Dinesh N. S. (2017). Optimal Position Control of a DC Motor Using LQG with EKF. International Conference on Mechanical System and Control Engineering.IEEE Trans. 978-1-5090-6530

[4] Banga, V., Kaur, J., Kumar, R.,&Singh, Y. (2011). Modeling and Simulation of Robotic Arm Movement Using Soft Computing. World Academy of Science, Engineering and Technology, 51, 616-619.

[5] Bode, H.W. (1945). Network Analysis and Feedback Amplifier Design. International Conference on Computer Engineering and Control Engineering. IEEE Trans. 978-1-5090-6530

[6] Bon-Ho B., Seun-ki, S.A. (2016). Compensation Method for time Delay of full-digital Synchronous Reluctance Machine. IEEE Trans. Ind Appl, 52, 3990-4000.

[7] Bhattacharya*; Apellat; Keel (2000). Robust Control-The Parametric Approach (PDF). Prentice Hall PTR. ISBN 0-13-781576-X.*

[8] Cathey, C.C. (2001). Electric Machines Analysis and Design Applying MATLAB

[9] Charles, M., Okoro, R., Ikposhi, I., & Oku, D. (2015). Reliable Control of PMDC Motor Speed Usind MATLAB. International Journal of Scientific & Engineering Research, 5(12), 208-216.

[10] Chen, C.T. (1995). Analog and Digital Control System Design: Transfer-Function, State-space, and Algebraic Methods. New York Orlando Austin San Antoro Toronto Landon Sydney Tokyo.

[11] Cois, O., Lanusse, P., Melchior, P., Dancla, F. and Oustaloup, A. (2002). Fractional systems toolbox for Matlab: applications in system identification and Crone CSD. In: 41st IEEE conference on Decision and Control, Las Vegas.

[12] CRONE Research Group, (2010). CRONE Control design Module User's Guide: Version.

[13] David, P.J. (2009). Sensitivity Analysis: Strategies, Method Concepts, Examples. University of Western Australia, Crawley. Engineering Technical Conferences and Computers and Information in Engineering Conference.

[14] Gamazo J.C., Vazque Sanchez E.J., Gomez G.L. (2010). Position and Speed Synchronous Reluctance Machine. IEEE Tan. Ind. Appl. 52, 3990-40000.

[15] Geest V., Geest, M.J., Polinder, H., Ferrera, J.A., Christman, M. (2015). Power Density Limits and Design Trends of High-speed Permanent Magnet Synchronous Machines. IEEE Trans. Electrification. 1,266-276.

[16] Zwyssig, C., Kolar, J.W., Round, S. (2009). Mega-speed Drive Systems Pushing beyond Million.IEEE Trans, Mechatronics. 564-574.

Validation of Aerodynamic Coefficients for Flight Control System of a Medium Altitude Long Endurance Unmanned Aerial Vehicle

Mudashir Olalekan Kadiri
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
kadirimudashir@gmail.com

Ameer Mohammed
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
https://orcid.org/0000-00022699-3608

Sunday Sanusi
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
toslez001@yahoo.com

***Abstract*—Stability and maneuverability have been the major trade-off in the design of aircraft both for civil or military usage. Hence, the need to design a suitable system for the control of the flight control surfaces (Elevator, Rudder and Aileron) on the aircraft. The flight control system (FCS) design helps resolve both the flying and handling qualities of the Unmanned Aerial Vehicles (UAV); it makes use of Stability Augmentation System (SAS) as a feedback mechanism tool for the FCS. Therefore, this paper discusses the validation of FCS for a Medium Altitude Long Endurance (MALE) UAV, whose service ceiling is 15,000 ft., having an endurance of up to 20 hours. A six degrees of freedom (DOF) aircraft model was used to test both the longitudinal and lateral/directional stabilities of the aircraft. Conceptual design parameters were used to generate both the dimensionless and derivatives of the MALE UAV using the AVL software. The results generated from AVL were validated using the digital DATCOM applications and both applications gave satisfactory results though with slight variations in the values of the coefficients obtained from both applications.**

Keywords— Flight Control System, Longitudinal Stability, Lateral/Directional Stability, Manoeuvrability, Stability Augmentation System, Unmanned Aerial Vehicles.

I. INTRODUCTION

Unmanned Aerial Vehicles (UAV) have been around for decades and are deemed to be the next major discovery in aviation industry. As the name suggests, these variants of aircraft do not have an on-board crew to actually fly them. They are remotely piloted from the ground control station. The first experiment involving flying without on-board pilot(s) can be traced back to early 20th century [1]. The Flight Control System (FCS) is essential for UAV to carry out flight missions with minimal or even without interference from human pilots [2]. The classical single-input/single-output (SISO) feedback control method, that is, Proportional Derivatives (PD) or Proportional Integral Derivatives (PID) control are one of the most common choices because of their simplicity in structure with fewer requirements on the accuracy of the dynamical model of the UAV, compared to the analytical method or the experimental method [3]. Hence, the MALE UAV flight control system shall incorporate a PID to improve the gain of the control system which shall be used for the future design of the SAS and shall be used to determine the stability of the UAV.

Flight control systems (FCS) have experienced a lot of changes over the past few years ranging from the use of warping wings and control surfaces by means of wires attached to the flying control stick in the cockpit, to the use of computer-based flight control system which provide improvements in stability and safety [4]. FCS are the devices that govern the attitude of an aircraft, it also affects the flight plan followed by the aircraft. The primary flight controls utilize elevator for pitching, aileron for rolling and the rudder for yawing. In manned aircraft, these surfaces are operated by the pilots in the flight deck or remotely piloted system in UAV. FCS consists of the flight control surfaces, connecting links and other necessary operating mechanisms to control the direction of an aircraft in flight [5]. The elevator acts in the lateral axis and causes the aircraft to **pitch,** while the aileron acts in the longitudinal axis and makes the aircraft **roll** and the rudder act in the vertical direction and makes the aircraft to **yaw** [6]. Hence, there is a direct relationship between the control system and control surfaces, which is described in Fig. 1. The reference trajectory is the input to the aircraft aerodynamics, this inputs are the pitch, roll, yaw and the throttle. These inputs cause the motion of the aircraft along any of the control axes (lateral, longitudinal and directional) through the movement of the control surfaces (Aileron, Elevator and Rudder). The output of the control surfaces are fed into the aircraft dynamics which is used to influence aircraft stability. The UAV produces three linear motions (heave, sway and surge) as well as three rotational motions (roll, yaw and pitch). The SAS is used to ensure effective feedback for the system.

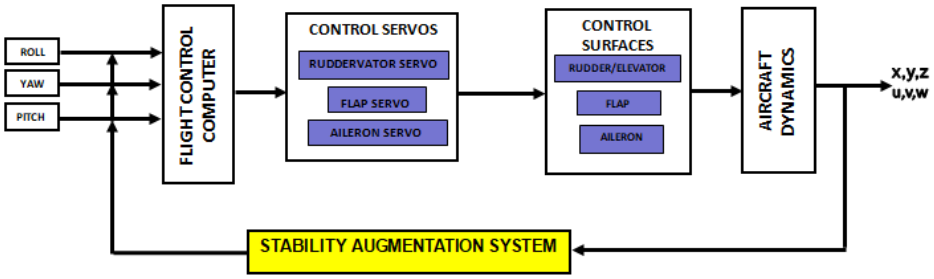


Fig.1: Block Diagram of FCS

The major motivating factor in the choice of the use of simulating software for the parameters of this UAV is to reduce the costs, time, and restrictions involved in the prediction of static and dynamic stability derivatives of the aircraft for its flight control system [7]. Computational efficiency as well as safety were the reasons why SIMULINK was adopted for modelling and simulating the behavior of the FCS. To ensure that the results are more robust, in addition to SIMULINK, the AVL results were validated using Digital Datcom. This helps with improving the reliability of the design.

The rest of the paper is organized as follows. Section II gives more detail on the selected validation platforms, namely, AVL and Digital Datcom. Section III presents the procedure on how to generate the coefficients of the aircraft. Section III compares the coefficients obtained using AVL and Digital Datcom Software. Concluding remarks are provided in Section IV.

II. Validation Platforms: AVL vs Digital Datcom

The Athena Vortex Lattice (AVL) is an open source code developed at Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. It is a program for the aerodynamic and flight dynamic analysis of rigid aircraft of arbitrary configuration, the flight dynamics analysis combines a full linearization of the aerodynamic model about any phase of flight [8]. The configuration definition is geometry-driven and the camberline is NACA dependent, while it produces its aerodynamics coefficients output as a function of forces and moments in a body or stability axes.

Digital DATCOM is a computer program that implements the methods specified in the United States Air Force stability and control manual to calculate the static stability, control and dynamic derivatives characteristics of a fixed-wing aircraft in accordance with specified flight condition [9]. The geometry of the aircraft is used as the input file while it produces corresponding dimensionless stability derivatives. The major limitations of the use of Digital DATCOM is that the program cannot analyse three lifting surfaces at once, however, this problem can be address by super-positioning of the lifting surfaces. Also, the dynamic derivatives of the aircraft are not suitable for not straight-tapered wings aircraft [10]. However, since the MALE UAV has a straight and tapered wing, the program can be easily used to obtain its aerodynamic coefficient.

III. Static and Dynamic Coefficients

The first step in the design of an FCS is to input the specifications of the UAV into a software which is used to generate the coefficients of the aircraft. The conceptual specifications from a light surveillance UAV has an empty weight of 450 kg, maximum take-off speed of 48 m/s, maximum operating weight of 600 kg, cruise speed of 44 m/s with a range of 5500 km, endurance of 20 hours and maximum ceiling of 15,000 ft. Based on user specifications, the airfoil section of the UAV has a root of NACA 2415 while the tip is made of SD 7032 with a mean aerodynamic chord of 0.90 m.

TABLE I. Longitudinal Coefficients using AVL

LIFT COEFFICIENT		PITCH MOMEMNT COEFFICIENT		DRAG COEFFICIENT	
CL	0.76530	Cm	-0.00730	Cd	0.05056
$Cl_{\bar{a}}$	5.47020	Cm_a	-0.32801	Cd_{da}	0.00033
Cl_q	7.55010	Cm_q	-4.45001	Cd_{de}	0.00048
$Cl_{\bar{d}e}$	0.00363	Cm_{de}	0.00622		
Cl_M	0.00011				

TABLE III. Longitudinal Coefficients using Digital Datcom

LIFT COEFFICIENT		PITCH MOMEMNT COEFFICIENT		DRAG COEFFICIENT	
CL	0.76390	Cm	-0.00742	Cd	0.04730
$Cl_{\bar{a}}$	5.44822	Cm_a	-0.32771	Cd_{da}	0.00033
Cl_q	7.57320	Cm_q	-4.43993	Cd_{de}	0.00050
$Cl_{\bar{d}e}$	0.00352	Cm_{de}	0.00022		
Cl_M	0.00012				

TABLE IV. LATERAL/DIRECTIONAL COEFFICIENTS USING DIGITAL DATCOM

YAW MOMENT COEFFICIENT		ROLL MOMEMNT COEFFICIENT		SIDEFORCE COEFFICIENT	
Cnb	0.02823	Clb	0.00862	Cyb	-0.16352
Cnp	0	Clp	-0.55771	Cyp	0
Cnr	0	Clr	0.17985	Cyr	0.00782
$Cn_{đa}$	0.00005	$Cl_{đa}$	0.00227	$Cy_{đa}$	0.00012
Cn_{dr}	0.00024	Cl_{dr}	0.00014		

The longitudinal stability coefficient which affect the UAV are pitch coefficient, lift coefficient and drag coefficient while the yaw moment coefficient, roll moment coefficient and the side-force affect the UAV in the lateral/directional stability axis [11]. Table I and Table II summarize the longitudinal coefficients and lateral/directional coefficients respectively of the MALE UAV using the AVL. While Table III and Table IV give the coefficients of both longitudinal and lateral/directional stability using the Digital DATCOM software.

IV. COMPARISON OF MAJOR COEFFICIENTS OBTAINED FROM AVL AND DIGITAL DATCOM

The six aerodynamic coefficients for the design of the FCS for a MALE UAV can best be summarized using Table V. These coefficients are the lift, drag, pitching moment, yawing moment, rolling moment and the side force of the aircraft. Lift is a force that opposes the weight of an aircraft and helps keeps the aircraft in the air [12]. It is mostly generated by the wing of the aircraft and acts through the centre of pressure of the aircraft. Also, the drag can be seen as the force that act opposite to direction of motion, it is caused by both friction and the atmospheric air pressure [13]. From Table V shown

TABLE II. LATERAL/DIRECTIONAL COEFFICIENTS USING AVL

YAW MOMENT COEFFICIENT		ROLL MOMEMNT COEFFICIENT		SIDEFORCE COEFFICIENT	
Cnb	0.02800	Clb	0.00831	Cyb	-0.16300
Cnp	-0.06402	Clp	-0.55852	Cyp	0.17401
Cnr	-0.01271	Clr	0.18341	Cyr	0.00802
$Cn_{đa}$	0.00005	$Cl_{đa}$	0.00250	$Cy_{đa}$	0.00038
Cn_{dr}	0.00023	Cl_{dr}	0.00013		

below, it can be observing that there is a slight variation in both the AVL and digital DATCOM coefficient, the percentage variation in the Lift-Drag ratio is 3.22. The lift-drag ratio of the coefficient using the AVL gives **15.14** while the lift-drag ratio of the coefficient using the Digital DATCOM gives **16.15** with a percentage error of 6.67 which is within the acceptable percentage error of 10 specified in [14]. According to [15], an aircraft with a higher Lift to Drag ratio are more efficient than those with lower lift to Drag ratio. Also, L/D also determines the gliding ratio ratio of the aircraft as well as the gliding range which affect the aircraft performance at the landing phase of flight [6].

V. CONCLUSION

This paper presents the validation of the aerodynamics coefficients of the flight control system for a MALE UAV based on the specification from a light surveillance UAV [10]. The data assumed the aircraft to have six degrees of freedom. This data was used on the AVL and Digital DATCOM models to obtain the aerodynamics coefficient of the UAV, these coefficients were used to determine the efficiency of the aircraft. Furthermore, a comparison of both AVL and Digital DATCOM coefficients was done and the results show some

TABLE V. AERODYNAMIC COEFFICIENTS USING BOTH AVL AND DIGITAL DATCOM

Symbol	Parameter	AVL Values	Digital DATCOM Values	Percentage Error
Cl	Lift	0.76530	0.76390	0.18
Cd	Drag	0.05056	0.04730	0.33
Cm	Pitch Moment	-0.0073	-0.00742	1.64
Cnƀ	Yaw Moment	0.02800	0.02823	-0.82
Cl_{b}	Roll Moment	0.00831	0.00862	-3.73
Cy_{b}	Side Force	-0.16300	-0.16352	-0.32

consistency between the two models for generating aerodynamic coefficients of the UAV. However, the Digital DATCOM values were found to be more reliable due to its higher values for lift- to- drag ratio compared to the AVL. Which shows that the UAV will be more aerodynamically efficient if the values obtained using the Digital DATCOM were used for the UAV design.

APPENDIX

The symbols and notations used in the paper are defined as follows. The symbols and notations are presented in the order they appear in the paper.

Symbol	Parameter
Cl	Coefficient of Lift
Cd	Coefficient of Drag
Cm	Pitch Moment Coefficient
Cnƀ	Yaw Moment Coefficient
Cl_{b}	Roll Moment Coefficient
Cy_{b}	Side Force
$Cl_{đ}$	Lift Curve slope with respect to Aileron
Cl_{q}	Lift Curve slope with respect to Angle of Attack
$Cl_{đe}$	Lift Curve slope with respect to elevator deflection
Cl_{M}	Lift Curve slope with respect to Pitching Moment
Cm_{a}	Pitch Moment Coefficient due to Aileron

$\mathbf{Cm_q}$	Pitch Moment Coefficient due to Angle of attack
$\mathbf{Cm_{de}}$	Pitch Moment Coefficient due to elevator deflection
$\mathbf{Cd_{da}}$	Coefficient of Drag due to Aileron deflection
$\mathbf{Cd_{de}}$	Coefficient of Drag due to elevator deflection
Cnr	Yaw Moment Coefficient due to rudder
$\mathbf{Cn_{da}}$	Yaw Moment Coefficient due to Aileron deflection
$\mathbf{Cn_{dr}}$	Yaw Moment Coefficient due to Rudder deflection
Clp	Roll Moment Coefficient due to Angle of Attack
Clr	Roll Moment Coefficient due to rudder
$\mathbf{Cl_{da}}$	Roll Moment Coefficient due to Aileron deflection
$\mathbf{Cl_{dr}}$	Roll Moment Coefficient due to Rudder deflection
Cyp	Side Force effect due to Angle of attack
Cyr	Side Force due to Rudder movement
$\mathbf{Cy_{da}}$	Side Force due to Aileron deflection

ACKNOWLEDGMENT

The authors thank Dr Nkem Ofodile of the Air Force Research and Development Centre, Nigerian Air Force Base Kaduna for the valuable discussions they had with her while developing the manuscript.

REFERENCES

D. P. Raymer, *Aircraft Design: A Conceptual Approach*. AIAA Education series, 2004.

Roger W. Pratt, *Flight control systems: practical issues in design and implementation edited*. IEE Control Engineering Series, 2013.

Atul Garg, Rezawana, L and Tonoy Chowdhury, "Evolution of Aircraft Flight Control System and Fly-By –Light," *IJETAE*, vol. 3, no. 12, 2013.

M.V. Cook, *Aircraft Modeling; practical issues in design and implementation*. Hertfordshire: IEE-IAAA, 2000.

R. C. Nelson, *Flight stability and automatic control*, 2nd ed. Boston: WCB/McGraw-Hill, 1998.

M. V. Cook, *Flight Dynamics Principles*, 2nd ed. Oxford: Elsevier, 2007.

F. L. Stevens, Eric Johnson and Lewis, *Aircraft control and simulation:Dynamics Control Design, and Autonomous Systems*, 2nd ed. New Jersey: Wiley, 2003.

W. H. Mason, "Software for Areodynamics and Aircraft Design." [Online]. Available: http://www.dept.aoe.vt.edu/~mason/Mason_f/MRsoft.html. [Accessed: 12-May-2019].

R. Fink, *USAF Stability and Control DATCOM*. Ohio: Wright Patterson, 1978.

USAF, "Stability and Control for Digital DATCOM," 2017.

K. Athulathmudali, "Flight Control Software Design for Small High Endurance Science Flight Vehicle," Cranfield, 2008.

T, Sablan, "Theory of flight," Massachussetts Institute of Technology, 2007.

A. S. Sheriff, "Problems in Flight Dynamics and its solution," *World J. Eng. Res. Technol.*, 2017.

J. Shen, Y. Su, Q. Liang, and X. Zhu, "Calculation and Identification of the Aerodynamic Parameters for Small-Scaled Fixed-Wing UAVs.," *Sensors (Basel)*., vol. 18, no. 1, Jan. 2018.

S. Salenko and AD, Obuhosky *Flight Dynamics*. NSTU Press, 2014.

Flight Performance of Solar-Powered Unmanned Aerial Vehicles: A Review

Aminu Hamza
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
aminuhmz@gmail.com

Ameer Mohammed
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
https://orcid.org/0000-00022699-3608

Abubakar Isah
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
dankade01@gmail.com

***Abstract*—Unmanned aerial vehicles (UAV) have applications in military, reconnaissance, disaster management, communication support, geographical survey and photography among others. Their use in everyday life is increasing. A common shortcoming of UAV is limited endurance. Most UAVs make use of conventional fuel for their propulsion as well as for powering on-board equipment. Relative to the size of these UAVs, the amount of fuel they can carry limits their endurance. Hence, there is a need for alternative ways to enhance their endurance as well to free the environment from pollutants. Sub-Saharan Africa is endowed with abundant solar radiation in which solar-powered UAVs have the potential to achieve optimum flight performance. Therefore, this paper reviews the flight performance of a solar powered UAV and also gives insights into the prospects of using a solar-powered UAV in achieving optimum flight performance**

Keywords—Flight Performance, Solar-Powered, Unmanned Aerial Vehicles

I. INTRODUCTION

Unmanned aerial vehicles (UAV) are aircraft that are operated autonomously or remotely piloted [1] [2]. Thus, human presence is completely removed in the aircraft itself. Interest in their use today is increasing as their application cuts across various human endeavors such as military [3] [4], humanitarian operations, disaster management [5] [6], communication support [4] [7] [8] [9], geographical survey and photography [4] [10], agriculture [4], remote sensing[9] [11] and payload delivery [4] [12] and so on. However, their major constraint is endurance. The endurance of an UAV is how long it can stay airborne. Relative to their size, the amount of fuel they carried limits their endurance. Most UAV make use of conventional fuel for their propulsion as well as for powering onboard equipment [13]. Hence, there is a need for alternative ways to enhance their endurance as well to free the environment from pollutants. The available option is to turn to alternative sources of energy. Renewable energy sources are a smart alternative to conventional fuel for UAV propulsion. These sources include solar, hydrogen fuel cells and high capacity storage systems [10]. Combination of conventional fuel and one of these alternative sources improves the endurance of a larger UAV while on the other hand an alternative source may be used for smaller UAV. At present, the ability to fly without conventional fuel is the main emphasis [1].

Insecurity, illegal cross border migration, terrorism as well as human trafficking have caused serious instability and economic hardship is Sub-Saharan Africa [14] . In order to curb this menace, there is a need for constant surveillance of migration routes, borders so as to survey movement of people across these countries. Aerial surveillance can be achieved effectively using UAV. Operating long endurance UAV on conventional fuel can be costly; it can also have adverse

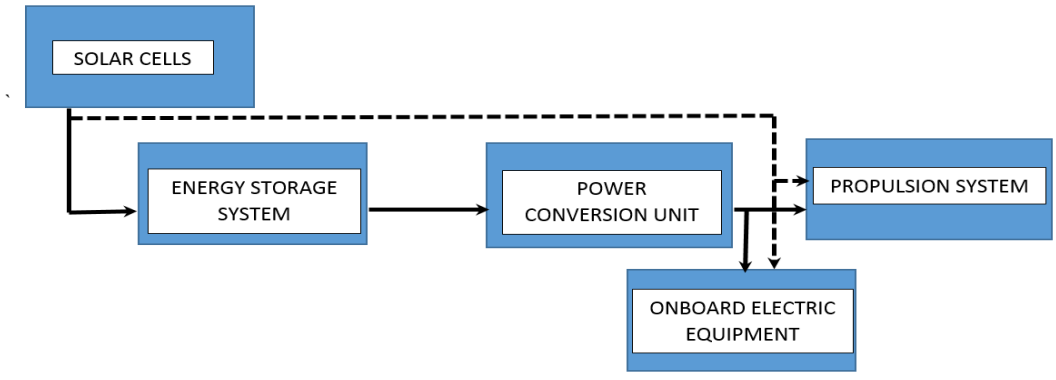


Fig.1: Block Diagram of a Solar-Powered UAV

TABLE I. LIST OF THE MOST PROMINENT SOLAR-POWERED UAV

UAV	Number of PV Cells	Cell efficiency (%)	Energy conversion efficiency (%)	Power (W)	Altitude (ft)	Endurance (hr)	Weight (N)	Year
SoLong [17]	76	-	96	225	-	48	12.8	2005
Zephyr 7[21]	96	19.6	90	320	70740	336	28.0	2010
Helios [23]	62120	-	-	40000	96863	30	720	2001
Zephyr 8[33]	-	28	-	-	70000	630	62	2018
Atlantik Solar 2 [28]	88	23.7	-	450	-	28	6.93	2015
Aquila [34]	-	-	-	5000	60000	2760	400	2016

environmental effects. To overcome some of these challenges, alternative energy sources can be adopted. Africa is blessed with an abundance of renewable energy resources such as solar energy, which can be harnessed even during flight. It is estimated that Africa receives in excess of 2000 kWh of global solar radiation annually which is the highest levels of annual global solar radiation [15]. This makes solar energy the most viable option for powering UAV in Sub-Saharan Africa. The UAV can be used for various activities such as continuous intelligence gathering, surveillance, reconnaissance as well as interdiction. Hence, this paper is focused on studying the flight performance of solar-powered UAVs with the aim of utilizing solar radiation to power UAVs in Sub-Saharan Africa. A typical block diagram of solar-powered UAV is shown in Fig 1. Section II gives more background information on solar-powered UAV. Section III discuss the performance and parameters of interest in solar-powered UAV. Section IV highlights their challenges and prospects in Sub-Saharan Africa, as well as any supporting technology. Concluding remarks are provided in Section V.

II. SOLAR-POWERED UAV

Solar energy has become one of the unlimited forms of renewable energy sources available that can be utilized to increase the endurance of UAV without adding substantial mass or requiring a substantial increase in size of the fuel system [1]. Sunrise I was the first solar-powered UAV to have flown. It flew for about 20 minutes at an altitude of around 100 m during its inaugural flight. It had a power output of 450 W using 4096 solar cells. This version was further improved to Sunrise II and tested with 4480 solar cells delivering about 600 W which is 14% efficiency of the solar cells [16][17].

Since then, many model airplane builders tried to fly with solar energy. Solar Challenger aircraft was built with 16,128 solar cells produced 2.5 KW achieving an endurance of 5 hours 23 minutes covering a distance of 262.3 km and reaching a service ceiling of 9 km, with solar energy as its only power source and no onboard energy storage system [16]. In 1993, the US government initiated AeroVironment Inc. project to study the feasibility of long duration solar electric that will fly above 19.812 km after the success of Solar Challenger. The project took the direction of solar propulsion with the Pathfinder that achieved its first flight in same year. The pathfinder achieved the record for a solar-powered aircraft, exceeding the altitude of the Solar Challenger. It reached an altitude of 15 km and two years later, it set another of record to 21 km [17][18].

Various works of research have resulted in the development of solar-powered UAV. In 2001 [19], Helios set the record for steady state sustained flight at an altitude of 29.5 km. In terms of day and night flight, SoLong was launched in 2005 it was airborne for 48 hours at a stretch. Another solar-powered UAV called Zephyr also set the record in 2007 for a steady-state flight lasting for over two weeks. With this, of course, there is quite notable progress in terms of solar-powered UAV [20]. Table I summarises some of the performance characteristics of the prominent solar-powered UAVs of the last two decades [19][21] [22][23].

III. FLIGHT PERFORMANCE OF SOLAR-POWERED UAV

Performance is a term used to describe the capability of an aircraft to accomplish a definite task that makes it useful for certain purposes [24]. Flight performance essentially deals with the behavior of an aircraft at cruise condition. This mainly involves optimizing endurance and range. Endurance is defined as how long an aircraft can stay in-flight on one full fuel tank. The range, on the other hand, is a measure of how far an aircraft can go on one full fuel tank [25]. The endurance and range of a solar powered UAV can calculated using equation I and II respectively [26]. A UAV has some limit to the amount of fuel tit can carry during a mission. This puts some restriction on their performance. To achieve an optimum flight performance, renewable energy sources have to be utilized. Solar energy is free, clean and practically inexhaustible, if optimally harnessed it can maximize the endurance and range of solar-powered UAV. In (1) and (2) $\boldsymbol{E}$ is the total energy in the battery, □ is propeller efficiency, $\boldsymbol{P_{Rmin}}$ Power required minimum, $\boldsymbol{T_{Rmin}}$ minimum thrust required minimum.

$$T_E = \frac{E}{\eta P_{Rmin}} \quad (\square\square\square)$$

$$R = \frac{\eta E}{T_{Rmin}} \quad (\square\square\square)$$

Karthik et al. investigated the performance of a solar-powered UAV [27]. The required power for the cruise was estimated and based on the estimated power, a conceptual design of the UAV was developed. In the study, the average energy efficiency was found to be 11%, while the exergy efficiency, as well as the conversion efficiency of the solar, were found to be 10% respectively. Details of the findings are shown in Table II. Oettershagen and colleagues [28], built a hand launch, small scale solar powered UAV with the aim of achieving perpetual flight endurance. The performance analysis solely focuses on the state of charge of the batteries at the start of the flight and after the flight. The energetic margin achieved after completing 28hrs flight shows significant improvement as compared to other UAV of the same category. However, the UAV considered in this paper is developed precisely to allow multi-day continuous flight in a 4-month window centered around summer solstice (June 21st) at a geographical latitude of 45^{o}N. Despite the fact that the time taken to reach operational altitude is greater because of low clouds and other weather phenomenon affecting the available solar energy; nevertheless, it recorded an endurance of 28 hrs as well as improved energetic margin closer to perpetual flight. The low clouds decrease solar irradiation. Its flight characteristics are summarized in Table III.

Park et al. [29] used a virtual system to evaluate the performance of main parts of the solar-powered UAV which are a solar panel and battery pack. The essence was to ascertain the feasibility of long term endurance. The system monitors energy production and consumption as the flight continues. The results of this simulation were compared with experimental results using MATLAB for validation. The two results show a similar trend for produced and consumed power. However, there is inconsistency in the amount of battery power left after the 13 hrs flight. This is due to the difference in actual and nominal efficiencies used as optical losses were neglected. Summary of findings are shown in Table IV.

In [30], Ozcan et al. focused on estimating the required solar irradiation to maintain the UAV in flight for 3.5 hrs. The flight path of the UAV is also taken into consideration so as to accumulate maximum solar energy available. A total of 218.7 MJ of solar was incident on the solar panels but only 18% of it was converted to electrical energy. This power is enough to power the UAV for 4 hrs based on the design, since no means of storage is provided in the design phase. The energy produced is being consumed instantaneously. The obtained performance characteristics are summarized in Table V.

The work by Mattar and colleagues investigates how the geometry of an airfoil used in making the wing of a solar-powered UAV can enhance its performance [31]. The geometry of airfoil was optimized using the Pattern Search (PS) algorithm. The solar panels were placed/arranged on the optimized airfoil of the wing. This gives the solar panels better exposure to solar irradiation which increases solar energy production. Similarly, the optimized airfoil improves the aerodynamic characteristic of the UAV. This also enhances the glide performance which minimizes energy consumption and thus increasing the range of the UAV

TABLE II. SUMMARY OF KEY PERFORMANCE PARAMETERS FOR [28]

S/N	Parameter	Ratings
1.	Energy efficiency	11 (%)
2.	Exergy efficiency	10 (%)
3.	Conversion efficiency	10 (%)

TABLE IV. SUMMARY OF KEY PERFORMANCE PARAMETERS FOR [29]

S/N	Parameter	Ratings
1.	Max power produced	30W
2.	Max power consumption	22W
3.	Battery power remained	220Ah
4	Endurance	13hrs

TABLE V. SUMMARY OF KEY PERFORMANCE PARAMETERS FOR [30]

S/N	Parameter	Ratings
1.	Power produced	2733 W
2.	Power required	750W
3.	Solar efficiency	<18%
4	Endurance	4hrs

TABLE VI. COMPARING KEY PERFORMANCE INDICATORS OF SOLAR-POWERED UAV

Ref.	Efficiency of solar panel	Energy storage system behavior	Optimization of wing airfoil (Aerodynamic efficiency)	Energy conversion efficiency	Exergy Solar irradiation in geographical location
[28]	-	✓	-	✓	✓
[27]	✓	✓	-	✓	✓
[29]	✓	✓	-	-	✓
[30]	✓	-	✓	-	✓
[31]	-	-	✓	-	-
Future	✓	✓	✓	✓	✓

IV. Discussion

From the foregoing, it can be understood that several important factors/parameters are being considered in describing the performance of solar-powered UAV [1]. These are;

- Efficiency of the solar panel.
- Behavior of the energy storage systems.
- Energy conversion efficiency.
- Type of geometry of the wing airfoil (Aerodynamic efficiency).
- Exergy of solar irradiation in the geographical location of interest.

The following literature [27] [28] and [30] placed more emphasis on the characteristic of the energy storage system and conversion efficiency while [27] and [31] focused on optimization of wing geometry to maximum solar irradiation. In addition, [28] and [29] concentrated on improving energy storage and power conversion efficiency. Table VI gives a summary of the various performance indicators various implementations have focused on when evaluating the performance of solar-powered UAV. From the foregoing discussion, it is evident that there is need for improvement of the power conversion efficiency of solar and storing capacity of batteries used for Solar-Powered UAV. Also, a suitable optimization method/tool that will give higher aerodynamic efficiency of wing airfoil is good topic for researchers

a. Prospects and Challenges in Sub-Saharan Africa

Harnessing solar energy in Sub-Saharan Africa is vital not just to power UAV but also economic development and poverty alleviation. With the highest level of solar irradiation globally, and policy initiatives that target the reduction in cost of photovoltaic module (solar cell) by 50% [15] [32] , tapping into solar resources can put an end to challenges that border on military reconnaissance and humanitarian operations, especially in Sub-Saharan Africa. This will whittle down military budgets and improve the productivity of forces in tackling some of this challenges. Other socioeconomic challenges are global warming, its effect can be mitigated with the adaptation of solar-powered UAV. Reduction in CO_2 emission levels may lead to cleaner environments. For UAVs to be fully deployed for military and humanitarian operations in Sub-Saharan Africa, the main challenge seems to be funding for research and development, putting in place operational guidelines, as well as ensuring political will by governments.

b. Supporting Technology

Recent breakthrough of solar cell technology across Sub-Saharan Africa which involves embedding solar cells using resilient and composite materials that blend into the aircraft aesthetic can be adopted. In addition, foldable PV cells can be incorporated so as to be compatible with adaptive or morphing wings that adjust flight control surfaces in response to changing flight conditions. The prospects of supporting technology for UAV in Nigeria are improving; the development of PV cell manufacturing facility in Maiduguri Nigeria, as well as more focus on UAV design and optimization in institutes like Air force Institute of Technology. As such, supporting technology for UAV will not be an impediment in realizing solar powered UAV technology.

V. Conclusion

Various approaches have been adopted by different works in describing the optimum performance of solar-powered UAV. It can be inferred that solar power can serve as an alternative power source to conventional fuel for propulsion of UAV to improve its endurance. This is ever more relevant in Sub-Saharan Africa, considering that it has the highest solar radiation in the world. This makes them practicable in number of applications including surveillance and reconnaissance, humanitarian operations; in addition, they can be useful in curbing the menace of insurgency and other forms of instability in the region. Furthermore, it will also reduce the effect of global warming and harmful gases emitted to the environment as by-products of conventional fuel combustion. Future work should focus on improving the efficiency of solar cells, storage capacity of batteries as well as improved optimization method for optimizing geometry of wing airfoil for higher aerodynamic efficiency.

Acknowledgment

The authors thank Engr. Suleiman Babani of Electrical Engineering Department, Bayero University Kano for the valuable discussions they had with him while developing the manuscript.

Reference

K. R. B. Sri, P. Aneesh, K. Bhanu, and M. Natarajan, "Design analysis of solar-powered unmanned aerial vehicle," *J. Aerosp. Technol. Manag.*, vol. 8, no. 4, pp. 397–407, 2016.

J. Gundlach, *Designing Unmanned Aircraft Systems: A Comprehensive Approach*. Virginia: American Institute of Aeronautics and Astronautics, Inc, 2012.

J. D. Barton, "Fundamentals of small unmanned aircraft flight," *Johns Hopkins APL Tech. Dig. (Applied Phys. Lab.*, vol. 31, no. 2, pp. 132–149, 2012.

A. Hobbs, "Unmanned Aircraft Systems," in *Human Factors in Aviation*, 2010, pp. 505–531.

Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, 2016.

Y.-C. Wang, D. Sheu, C. Tsai, H.-J. Huang, and Y.-H. Lai, "Performance Analysis and Flight Testing for Low Altitude Long Endurance Unmanned Aerial Vehicles *," *J. Aeronaut. Astronaut. Aviat. Ser. A*, vol. 45, no. 1, pp. 37–044, 2013.

B. Ozturk, M. J. Burton, and W. W. Hoburg, "Design of an Unmanned Aerial Vehicle for Long-Endurance Communication Support," in *American Institute of Aeronautics and Astronautics Inc*, 2017, no. June, pp. 1–18.

C. Liu, M. Ding, C. Ma, Q. Li, Z. Lin, and Y. C. Liang, "Performance analysis for practical unmanned aerial vehicle networks with LoS/NLoS Transmissions," in *2018 IEEE International Conference on Communications Workshops, ICC Workshops 2018 - Proceedings*, 2018, pp. 1–6.

E. Yanmaz, S. Yahyanejad, B. Rinner, H. Hellwagner, and C. Bettstetter, “Drone networks: Communications, coordination, and sensing,” *Ad Hoc Networks*, vol. 68, no. September, pp. 1–15, 2018.

J. Meyer, F. Du, and W. Clarke, “Design Considerations for Long Endurance Unmanned Aerial Vehicles,” in *Aerial Vehicles*, 2012.

H. Shakhatreh *et al.*, “Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges,” *IEEE Access*, vol. 7. pp. 48572–48634, 2019.

E. Cano, R. Horton, C. Liljegren, and D. Bulanon, “Comparison of Small Unmanned Aerial Vehicles Performance Using Image Processing,” *J. Imaging*, vol. 3, no. 1, p. 4, 2017.

J. Renau, F. Sánchez, A. Lozano, J. Barroso, and F. Barreras, “Analysis of the performance of a passive hybrid powerplant to power a lightweight unmanned aerial vehicle for a high altitude mission,” *J. Power Sources*, vol. 356, pp. 124–132, 2017.

Linda Etim, “Terrorism and Instability in Sub Saharan Africa,” 2016, pp. 16–22.

D. A. Quansah, M. S. Adaramola, and L. D. Mensah, “Solar Photovoltaics in Sub-Saharan Africa - Addressing Barriers, Unlocking Potential,” *Energy Procedia*, vol. 106, pp. 97–110, 2016.

B. J. Robert Boucher and I. Cincinnati, “History of Solar Flight AIAA/SAE/ASME,” in *20th Joint Propulsion Conference*, 1984, pp. 1–22.

A. Noth, “Design of Solar Powered Airplanes for Continuous Flight Sky-Sailor Design Sky-Sailor Prototype,” 2008.

A. Noth, “History of solar flight,” *Autonomous System lab*, no. July, pp. 1–7, 2013.

S. N. Kane, A. Mishra, and A. K. Dutta, “Preface: International Conference on Recent Trends in Physics (ICRTP 2016),” *Journal of Physics: Conference Series*, vol. 755, no. 1, 2016.

W. Harasani *et al.*, “Conceptual Design Features of the Sun Falcon 2 a Long Duration UAV,” *MAYFEB J. Mech. Eng.*, vol. 1, pp. 1–21, 2016.

IHS, “High-flying bird : Zephyr remains in the vanguard of solar-powered flight,” *Jane’s International Defence Review*, pp. 1–7, 2017.

G. Kumar, S. Sepat, and S. Bansal, “Review paper of Solar Powered UAV,” *Int. J. Sci. Eng. Researc*, vol. 6, no. 2, pp. 41–44, 2015.

L. J. Ehernberger, C. Donohue, and E. H. Teets Jr., “A review of solar-powered aircraft flight activity at the pacific missile range test facility, Kauai, Hawaii,” *Conf. Aviat. Range, Aerosp. Meteorol.*, p. p 1063-1069, 2004.

FAA, “Pilot’s Handbook of Aeronautical Knowledge - Chapter 11 - Aircraft Performance,” 2016, p. 28.

J. D. Anderson, *Aircraft perfmance and design*, Tata-McGra. McGraw Hill, 1999.

[26] P. R. Grant, “Fundsmentals of UAV Performamnce.” pp. 1–51, 2016.

B. S. Karthik Reddy and A. Poondla, “Performance analysis of solar powered Unmanned Aerial Vehicle,” *Renew. Energy*, vol. 104, pp. 20–29, 2017.

K. R. B. Sri, “results , performance analysis and model validation,” *J. Aerosp. Technol. Manag.*, vol. 4, pp. 397–407, 2016.

H. B. Park, J. S. Lee, and K. H. Yu, “Experiment and evaluation of solar powered UAV by virtual flight system,” in *2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan, SICE 2015*, 2015, pp. 1052–1057.

G. Ozcan *et al.*, “Solar irradiation estimation on a solar powered UAV over a mission course,” *2015 Int. Conf. Unmanned Aircr. Syst. ICUAS 2015*, vol. 1, pp. 102–108, 2015.

A. K. Mattar, C. Vaidya, and A. M. Saraf, “Airfoil selection, optimization and analysis for a solar-powered unmanned aerial vehicles,” in *IEEE Aerospace Conference Proceedings*, 2016, vol. 2016-June, pp. 1–9.

J. Amankwah-Amoah, “Solar Energy in Sub-Saharan Africa: The Challenges and Opportunities of Technological Leapfrogging,” *Thunderbird Int. Bus. Rev.*, vol. 57, no. 1, pp. 15–31, 2015.

T. Westbrook and H. Rostvik, “Solar Energy and Battery Development for Unmanned Aerial Vehicles: Leading to Increased Proliferation?,” 2018.

A. Kumar, “AQUILA (THE SOLAR POWERED DRONE),” *Int. J. Ind. Electron. Electr. Eng.*, vol. 4, no. 10, pp. 36–43, 2016.

VI.

Electric Power-Assisted Steering: A Review

Abubakar Isah
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
dankade01@gmail.com

Ameer Mohammed
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
https://orcid.org/0000-00022699-3608

Aminu Hamza
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
aminuhmz@gmail.com

***Abstract*—for safe operation of a vehicle, the driver must be able to maintain absolute control of the vehicle's serious operating dynamics. To achieve this, extensively power steering systems have been employed today. This has been as a result of the increase in the front axle loads of vehicles, making it require more power to turn its wheels. In addition, there is a need for vehicles with more responsive steering properties. As a result, the quest for higher efficiency of the system necessitates the need for a more flexible and efficacious system. Electric Power-Assisted Steering (EPAS) systems can serve as alternatives to hydraulic power-assisted steering systems. The EPAS system offered the desired flexibility, engine and space efficiency. In addition, no pollution is emitted to the environment because of the omission of all hydraulic components. However, it has issues with system inertia and vehicle stability. Therefore, the requirement for vehicle steering stability becomes strict. As such, the focus of this paper is to give an overview of the EPAS system, its different configurations and subsystems, in addition to prospects and challenges of adopting the technology in vehicles.**

Keywords—Actuators, Controllers, Electric Power Assisted Steering, Sensors.

I. INTRODUCTION

One of the most critical safety factors for vehicle control is its steering system. Over the past few years, vehicle steering systems have been periodically updated due to the increasing front axle loads. This makes them require more power to turn their wheels [1][2]. There is also a need for vehicles with more responsive steering characteristics [3]. For this reason, there has been increasing incorporation of electronic control technology in power steering systems [2]. The power steering system on a passenger vehicle is mostly Hydraulic Power Assisted Steering (HPAS) or Electric Power Assisted Steering (EPAS) system. In comparison to HPAS, EPAS is safer, harmless to the environment and saves more energy [4][5]. The system does not have any belt-driven steering pump that is continuously running with the engine, so it has less weight and the motor consumes energy only when the steering wheel is turned by the driver; this leads to improvement in fuel efficiency [6][7][8]. The unique characteristic of the EPAS system is that it offers a way in appreciating minimum energy consumption in electric vehicle (EV) subsystem [9]. The increase in fuel economy of vehicles equipped with electric power steering varies by vehicle and engine size but normally differ from 1 to 3 miles per gallon [10]. An electric motor is used in the system to assist the driver when negotiating a turn. This is achieved via a sensor that detects the location and torque applied to the column of the steering. An assistive torque is applied by a computer module through the assist motor connected to either the steering gear or column of the steering [11][12]. Depending on the conditions of deriving, this offers different amounts of assistance to be obtained [2]. In spite of the fact that EPAS systems provide the desired flexibility, free the surrounding from toxic, provides engine and space efficiency. It comes with certain issues which include system inertia and vehicle stability [13]. Inertia is an issue that heavily influences the vehicle stability [14]. As a result, the requirement for vehicle steering stability becomes stringent and in addition to inertia, the vehicle steering instability may be attributed to many external disturbances and system uncertainties. More so, increasing the responsiveness of the steering system decreases its stability [14]. A test study carried out in [14] by a simple combination of variables were each variable; level of the main damping, the servo motor gear ratio, as well as the virtual model reference rack mass, was varied periodically to identify their effect on critical speed. It shows that for an increase in desired rack mass and damping, the stability of the vehicle increased. In [5], it was noted that damping improves the performance of vehicles running in a straight line. However, varying the servo motor gear ratio in [14] did not produce any clear trends.

Therefore, the focus of this paper is to give an overview of the EPAS system, its different configurations, controllers, sensors and actuators. The rest of the paper is organized as follows. Section II details the design requirement of an EPAS system. Section III gives background information on the various EPAS configurations. Section IV presents and discusses the various sensors, actuators and controllers used in an EPAS system. Section V highlights the challenges and prospects of EPAS technology. Concluding remarks are provided in Section VI.

II. DESIGN REQUIREMENT OF EPAS SYSTEM

With the idea that the designs of the EPAS system differ amongst manufacturers; nevertheless, the essential components for designing an EPAS system must include:

i. Torque sensor;
ii. Electric motor;
iii. Rotational angle sensor;
iv. Controller;
v. Vehicle speed sensor and;

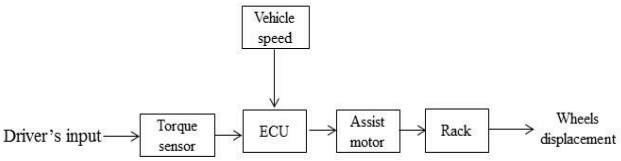


Fig. 1. Torque flow of EPAS system

vi. Coupling between motor and steering mechanism [10][13].

Undoubtedly, the most important element of EPAS is the torque sensor because it detects and process the driver's effort applied to steer the wheels of the vehicle and directed it to the vehicle electric control unit (ECU) [6]. Based on the pre-established characteristic curves, steering the direction, and vehicle speed the ECU defines the necessary motor current. The characteristic curves are discussed in section IV. The ECU then regulates the electric motor voltage to produce the desired current while succeeding the desired steering. To accomplish this the assist motor may be positioned at a different number of positions [10][13][15] and [16]. The motor controller is fundamentally employed to control the torque supplied to the steering mechanism. The vehicle speed is required to fine-tune the sensitivity of the torque controller and as soon as the torque of the steering wheel increases, the torque of the motor increases to ensure the usability of the steering. However, at higher speeds, the torque is reduced in order to obtain a better road feel [16]. The angle of rotation of the steering wheel must also be used to adjust the sensitivity and the performance around the null position of the steering wheel [2][6][10] and [13]. Fig. 1. shows how torque flows in the EPAS system. Table I show the specifications of the electromechanical components of a typical pinion-type (P-EPAS) system

III. EPAS System Configurations

Principally, the EPAS system has three major configurations that have been extensively used: pinion-type (P-EPAS), rack-type (R-EPAS) and column-type (C-EPAS) configurations [11]. The type represents the location where the assisting motor is positioned. In [15] the effect of motor position of different configurations of EPAS systems on the performance of the system were studied. The study revealed

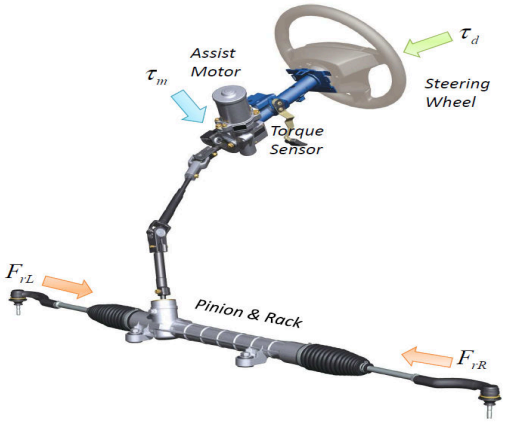


Fig. 2. Column assist configuration [7].

TABLE I. Electromechanical Specifications of a Typical Pinion-Type (P-EPAS) System [2][10].

ITEM	SPECIFICATIONS
Rack Force	7747 N
Rack stroke	144 mm
Stroke ratio	45.335 mm/rev
Rack and Pinion	
Type	2.3
Number of teeth	6
Reducer	
Type	Worm and resin wheel
Reduction gear ratio	15.1
Motor	
Type	Brushed DC motor
Rated voltage	12 V
Rated current	65 A
Rated torque	3.4 Nm
Rated rotational speed	1,180 rev/min

that changing the assist unit position from the steering wheel (C-Type) to the wheels increases the steering sensitivity and efficiency. The various configurations are deliberated in the next sections.

a. Colum-type assist configuration (C-EPAS)

In this configuration, the unit responsible for the assist is positioned at the steering column inside the cabin. This configuration is the world first commercialized EPS designed in the year 1988 for larger vehicles [7][16][17]. For compact vehicles, this configuration is perfect with small engine section: power-assist unit is integrated into the steering column and thus, the assist unit is waterproof. The configuration has superior safety, comfort and environmental performance. Currently, the C-type configuration is normally equipped in a compact car which has a load of about 6 kN [11][17]. Fig. 2 shows the column-type assist configuration.

b. Pinion-type assist configuration (P-EPAS)

In this configuration, the assist unit is integrated to the pinion shaft. The configuration design is light, compact adapted to the engine compartment and the system motor output is higher [17][10]. Fig. 3 shows the pinion-type assist configuration.

c. Rack assist configuration (R-EPAS)

Rack configuration provides excellent steering performance as a result of minimal friction between the time of turning the steering wheel and start of rack movement [16]. In this configuration, the assisting unit is integrated into the rack shaft and located in the engine compartment. This configuration is used for medium to large-size vehicles with a high force acting on the rack of about 8 kN [17]. Fig. 4 shows the rack-type assist configuration. Table II summarizes the load requirements for each configuration of the EPAS system.

TABLE II. LOAD REQUIREMENT OF EPAS [1][11].

Configuration	Vehicle type	Load (kN)
C-EPAS	Small	6
P-EPAS	Medium	8
R-EPAS	Large	12

IV. SENSORS, ACTUATORS AND CONTROLLERS

a. Sensors

Torque sensors form a significant part of EPAS system. They detect the applied torque quantity from the steering wheel by the driver and its direction [10]. It works by measuring the torque on which the current EPAS control strategies are based on [17]. The choice of a torque sensor influences the performance of the EPAS system. The common types of torque sensors employed in EPAS systems are torsion bar, optical torque sensor, surface acoustic torque sensor, strain gauge, piezoelectric torque sensor, magnetostrictive and magnetic torque sensor [18]. Table III shows the characteristic of these sensors. From the table, it can be seen that optical sensors have the best performance indicators.

b. Controllers

The EPAS system is inherently unstable and as such a controller is usually designed and integrated into the system so as to achieve stability. However, designing an EPAS system entails solving a problematic tracking control under the existence of disturbance and uncertainty [15]. Therefore, the performance of the EPAS requires that the control must deal with the transmissibility observed from the external sources of instabilities to the driver to achieve a good steering feel [8]. The common classical type of controller used is the Proportional Integral and Derivative (PID) controller which have been broadly and effectively employed in the EPAS systems. The advantages of PID control algorithms over other control methods is its simple structure

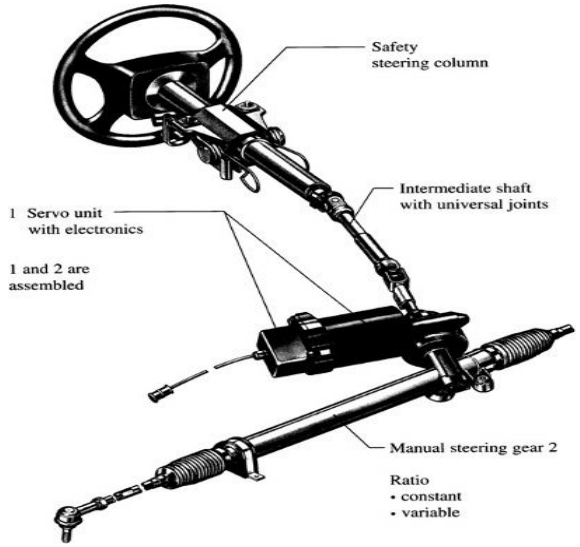


Fig. 3. Pinion assist configuration [1].

TABLE III. CHARACTERISTICS OF TORQUE SENSORS

Sensor	Sensitivity	Response	Precision	Cost	Weight
Strain gauge	-	Slow	Low	-	Very heavy
Torsion bar	Good	-	High	-	-
Optical sensor	Good	Fast	-	Low	Very light
Piezoelectric	-	-	-	-	-
Magnetic sensor	Good	Fast	-	-	-

and low implementation cost. Nevertheless, it has poor transient performance when external disturbances or uncertainties are in existence [15]. The work in [19] demonstrated that under double lane changes and returning condition, the designed curve under PID control can provide higher performance in control stability. The work in [4] investigated the robustness and stability of EPAS controlled by Active Disturbance Rejection Controller (ADRC) through frequency-domain analyses on a C-type EPAS system. The result shows that the steering wheel can be turn by the driver with the desired steering torque, which is free from load torques that tend to vary with the change of driving conditions.

On the other hand, model-based controllers such as Linear Quadratic Gaussian (LQG), fuzzy control, the robust controller ($\mathbf{H}_{\infty}$) etc., are also used as EPAS controllers. The former control method has been used in designing controllers of EPAS system and can be used to reduce the number of required sensors and offer to assist in the existence of external disturbance and measurement noise [15]. To increase robustness and performance of the EPAS system, the robust control method, $\mathbf{H}_{\infty}$ has been frequently used in controller design [15], [20]. The synthesis was used in [20]; it was found to lessen the effect of disturbances on the outputs. In addition, Dannohl et al. established an improved robust controller, $\mathbf{H}_{\infty}$ for R-type EPAS system and found the

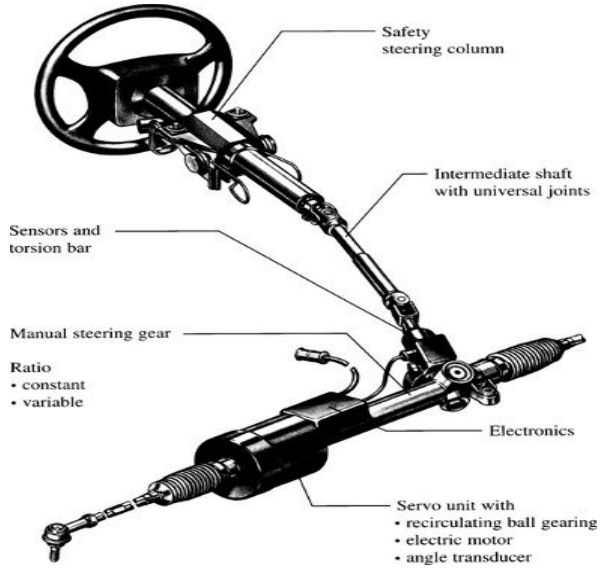


Fig. 4. Rack assist configuration [1].

improved controller boosted the performance of the system in contrast with the original $\mathbf{H}_{\infty}$ controller [15]. A two-stage

TABLE IV. EFFECT OF PID CONTROLLER GAINS [22].

Controller Response	Rise Time	Overshoot	Settling Times	Steady State Error
K_p,	Decrease	Increase	Small Change	Decrease
K_i,	Decrease	Increase	Increase	Eliminate
K_d	Small Change	Decrease	Decrease	Small Change

controller for a C-type EPAS system was developed according to [7], [15] and [21]. The first stage is a PID controller to increase motor response, while the second stage is a $\mathbf{H_{ss}}$ controller which describes the preferred assist torque based on the road feels performance requirements. Another aspect of EPAS controller that is getting much attention recently is the tuning of the controller [9][22]. The procedure of choosing the controller factors to meet an optimal performance requirement so that less power will be consumed during its operation is known as controller tuning [22]. Today, many studies have been carried out to optimize the energy consumption of the PID controller used in the EPAS system. The work in [9] suggested that the best gain parameters for the controller should be chosen using optimization techniques to save time and cost as against the traditional methods. The work in [22] recommended guidelines for tuning PID controllers (i.e. setting the values K_p, K_i, K_d) established either on investigational step responses or on the K_p values that result in minimal stability when proportional control action is used alone. The guidelines, which are briefly presented in Table IV suggest a set of values of K_p, K_i and K_d that will aid the stable operation of the system. However, the resulting system may exhibit a large maximum overshoot in the step response, which is generally undesirable. In such a case an iterative fine tuning is required until tolerable results are gotten [22]. Table IV summarizes the effect of various PID controller gains on the system.

Some of the techniques used in tuning the controller are Genetic algorithm (GA), fuzzy logic, particle swamp optimization (PSO) among others. Via simulation of a simplified social system, PSO was advanced and has been established to be strong in solving problems of continuous nonlinear optimization. It has features of easily implemented technique, stable characteristic convergence and good computational efficiency [23]. In [9], PSO was used and the outcomes of the analysis established on the hybrid PSO-PID controller demonstrated better performance and effectiveness contrasting the conventional PID controller. In addition, the work in [24] proposed PSO optimization technique to improve motor current tracking performance in EPAS system.Based on the result obtained, power consumption was decreased by 0.023% compared to Ant Colony Optimization (ACO). The results in [23] done for PID and PSO-PID based show that PSO-PID controller was able to control the current almost accurately, produce a smooth motor power and hence it shows potential in optimizing energy efficiency of EPAS system as against the PID method. On the other hand, in the case of dynamic vehicle velocity, there is need for it to be further refined. In [22], a fuzzy-PID controller was designed for an EPAS system and it significantly improves the system manoeuvrability, steering wheel returnability, system performance and robustness. The control strategy proved to be more effective than the traditional PID controllers. In

TABLE V. CHARACTERISTICS OF SOME EPAS MOTORS

Motors	Power (W)	Rotational speed	Noise	Inertia	Rotor Position Sensor
BDC	500	Low	High	High	None
BLDC	>500	High	Low	Low	Mandatory
BLAC	>500	High	High	Low	Mandatory

[25], an EPAS system based on brushless DC (BLDC) motor was implemented. It combines the traditional PID control with a fuzzy controller to make the system performance robust. The motor torque ripple was insignificant, the dynamic response of the system was fast, and a good assist characteristic was attained. In [11], genetic algorithm (GA) was used with a C-type EPAS system which employed a BLDC. The result proved that the tuned controller minimizes energy consumption as compared to PID controller.

c. Actuators

Generally, an electric motor is used for actuation of the EPAS system for the provision of assist torque. It receives signals from the controller to execute commands [7]. To determine target motor current levels from assist characteristic or boost curve, usually the EPAS system depends on vehicle velocity and driver torque measurements [26][25]. The characteristics curve of the assist is an established graph of motor current against torque of the steering wheel and tentative generated velocity of the vehicle to produce the assistance torque desired [26][25]. The EPAS assist characteristic commonly used are straight, polygonal/broken or curve line type [25][27]. Since the changing between motor current and steering wheel torque is linear, therefore the linear type of assist characteristics is simple and easy to control, which is used normally [25]. The graphs are shown in Fig. 5.

Different types of electric motors have been employed for the EPAS system from DC motor to brushless motors; brush DC (BDC) motor, BLDC and AC (BLAC) motor [11]. Either way, the selected motor for the system should produce a smooth torque with minimum ripple, high efficiency, low inertia, fault-tolerance capability and the minimum package size and weight [11]. The major criteria for selecting the motor of an EPAS system is power while other requirements are secondary [23]. The use of BDC on an EPAS needs a low cost and a simple control [26]. However, the brushes require high maintenance, low overall power density and Electromagnetic Interference (EMI)

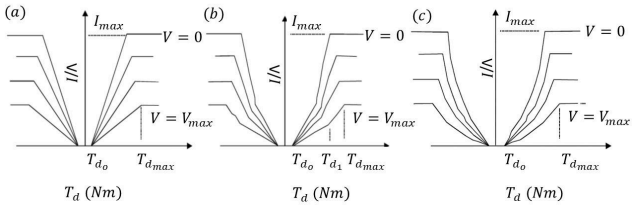


Fig. 5. Assist characteristic or boost curve [19].

problems related to commutator arcing are some of the disadvantages associated with it. These shortcomings

encourage some manufacturers to apply a BLDC on EPAS system [26]. The motor controller employed presently for the BLDC is more complex than the traditional as it necessitates a rotor position sensor (resolver). Therefore, BLDC has the disadvantage to be more expansive but undeniably it is a high-_ powered EPAS motor, realizing the low moment of inertia, low torque ripple, low cogging and frictional loss torque while reducing noise characteristics [28]. Table V shows the types and their characteristics. It is clear from Table V that BLDC outperforms other motors as EPAS actuator and thus recommended.

V. Prospects and Challenges of EPAS

a. Challenges

EPAS system has evolved since when it was first successfully implemented. Its demonstrated significant advantages over HPAS however, it still faces numerous challenges needed to be mastered such as system inertia, stability and energy consumption to obtain an accepted replacement of the HPAS systems [14][29]. Vehicles energy demand is greatly influenced by the drive system and the number of add-on systems which must be satisfied completely by the battery and the charging system [2][12]. Being an energy-consuming system, studies carried out on EPAS in [9][11][12][24] and [26] provides different energy optimization methods. In addition, irrespective of how it is configured the upcoming new technology of 42-volt will offer high power to vehicle electrical systems including the EPAS system [2][13][29]. With the fact that noteworthy progress has been achieved and realized on stability however, much is required to be done on actuation, steering feel, functional safety and sensing sub-system. According to [29], EPAS development effort seems to be too high, its steering feel may not be adequate and functional safety might be difficult to realize. Moreover, the existing generation of EPAS boost cannot provide the feel and handling that drivers anticipate in larger and heavier vehicles [2][13]. The least expensive in all of the EPAS configuration is the C-type configuration however, is not suited for use on large and heavy vehicles which is a greater challenge. The R-type configuration is more expensive which is more suitable for smaller vehicles application [2][13]. Motor noise and power-to-weight ratio of EPAS actuation system still require more research. The EPAS system requires an acceptable design motor with greater efficiency and forbearance of temperature. Since the derivers inputs are usually specific course adjustments followed by periods of idleness. Additionally, in an under-hood environment, the motors must as well work for prolonged periods that can withstand excessive high temperature of about 150 degrees Celsius with little or maintenance-free [30].

b. Prospects and Future Potentials

Even with the highlighted above, the EPAS system has the potential to give drivers a wonderful driving experience and feeling. It will also save a lot of cost for car manufacturers as the material required to implement the EPAS system is less compared to the HPAS system resulting in production cost. Furthermore, fuel consumed by cars with EPAS is also less as compared to HPAS. This translates to savings in fuel as well as a reduction in carbon dioxide emission to the environment.

VI. Conclusion

This paper reviewed the EPAS system. It looks into its various configurations, torque sensors, controllers and actuators. The types torque sensor used in EPAS provides numerous research opportunities; torsion bar and optical sensors are the most used sensors. Performance of other types of torque sensors can be investigated so as to come up with the most optimum in terms of complexity and speed. Finally, suggested development in EPAS should entail optimization of controller and actuation systems.

References

P. D.-I. J. Reimpell, D.-I. H. Stoll, and P. D.-I. J. W. Betzler, *The Automotive Chassis: Engineering Principles*, Second. Oxford: Butterworth-Heinemann, 2001.

V. N. Sulakhe, M. A. Ghodeswar, and M. D. G. Gite, "Electric Power Assisted Steering," *Int. J. Eng. Res. Appl.*, vol. 3, no. 6, pp. 661–666, 2013.

Z. Li, S. Shi, and C. Zhang, "Active return control strategy of electric power steering system based on disturbance observer," *J. Eng. Sci. Technol. Rev.*, vol. 11, no. 3, pp. 61–69, 2018.

P. B. Kandula, "Dynamics and control of an electric power assist steering system," Cleveland State University, 2010.

Z. Ma and C. Zhan, "System Stability and Control Strategy of Electric Power Steering," in *International Conference on Logistics Engineering, Management and Computer Science*, 2015, no. Lemcs, pp. 303–306.

V. Mandal, D. Chandra, S. Yadu, V. Roy, and J. Yadav, "ELECTRIC POWER ASSISTED STEERING," *IJARIIE*, vol. 2, no. 2, pp. 911–914, 2016.

K. Yamamoto *et al.*, "A New Control Design for an Optimized Electric Power Steering System Electric Power Steering System," in *20th IFAC World Congress*, 2017.

M. Marouf, Alaa Djemai and P. Sentouh, Chouki Pudlo, "A New Control Strategy of an Electric Power Assisted Steering System," IEEE Trans. Veh. Technol., vol. 8, no. 61, pp. 3574–3589, 2012.

R. A. Hanifah, S. F. Toha, and S. Ahmad, "Energy Consumption Optimization with PSO Scheme for Electric Power Steering System," *Int. J. Comput. Theory Eng.*, vol. 7, no. 4, pp. 297–301, 2014.

M. F. Rahman, "ELECTRIC POWER ASSISTED STEERING SYSTEM FOR AUTOMOBILE," Encycl. Life Support Syst., vol. III, 2013.

M. K. Hassan, N. A. M. Azubir, H. M. I. Nizam, S. F. Toha, and B. S. K. K. Ibrahim, "Optimal design of Electric Power Assisted Steering system (EPAS) using GA-PID method," *Procedia Eng.*, vol. 41, pp. 614–621, 2012.

J. Römer, P. Kautzmann, and G. Frey, Michael Frank, "Reducing Energy Demand Using Wheel-Individual Electric Drives to Substitute EPS-Systems," *Energies*, vol. 11, 2018.

V. Chahare, I. Mhaisne, and D. Dongre, "An Overview on Future Electric Steering System : A Project Approach," *Int. Res. J. Eng. Technol.*, vol. 4, no. 7, pp. 2152–2159, 2017.

A. Gröndahl, "Functional Modelling and Simulation of an Electric Power Assisted Steering Development of an advanced engineering tool," Chalmers University of Technology. Gothenburg, Sweden, 2018.

N. Mehrabi, "Dynamics and Model-Based Control of Electric Power Steering Systems," University of Waterloo, Ontario, Canada, 2014.

C. IJEKT, "Electric Power Steering Systems," *IJEKT Corporation*, p. 9, 2015.

K. Yamamoto, "Control of Electromechanical Systems, Application to Electric Power Steering Systems," Communaute Universite Grenoble Alpes, 2016.

L. J D Turner, J D Austin, "A review of current sensor technologies and applications within automotive and traffic control systems," in *Instn Mech Engrs*, 1999, vol. 214, pp. 589–614.

H. Chen, Y. Yang "Study on Electric Power Steering System Based on ADAMS and MATLAB," *J. Softw. Eng.*, vol. 9, no. 8, pp. 868–876, 2015.

L. Yi Chabaan, Rakan C. Wang, “Control of electrical power assist systems: H design, torque estimation and structural stability,” *J. Soc. Automot. Eng. Japan*, vol. 22, pp. 435–444, 2001.

A. Natarajan, “Design and Control of Electric Power Assisted Steering (EPAS) Electronic Control Unit (ECU),” in *Proceedings of Sixth IRAJ International Conference*, 2013, no. October, pp. 15–19.

A. Gupta, S. Kumar, and R. Purohit, “PID Control of Electric Power Steering System,” *Int. J. Res. Publ.*, vol. 4, no. 8, pp. 8–11, 2015.

A. Hanifah, A R Toha, S F Ahmad, “Hybrid PID and PSO-based control for electric power assist steering system for electric vehicle,” in *IOP Conference Series: Materials Science and Engineering OPEN*, 2013.

A. A. Bahmanshiri, “PARTICLE SWARM OPTIMIZATION TECHNIQUE OF CURRENT TRACKING CONTROLLER FOR ELECTRIC POWER-ASSISTED STEERING SYSTEM,” University Putra Malaysia, 2015.

L. Feng, R. Shen, and X. Wu, “Research on control strategy of the vehicle electric power steering system based on brushless DC motor,” in *International Conference on Advances in Energy and Environmental Science*, 2015, no. Icaees, pp. 503–506.

K. A. Danapalasingam, “Optimisation of energy in electric power-assisted steering systems,” *Int. J. Electr. Hybrid Veh.*, vol. 5, no. May 2013.

G. Shi, S. Zhao, and J. Min, “Simulation Analysis for Electric Power Steering Control System Based On Permanent Magnetism Synchronization Motor,” in *2nd International Conference on Electronic & Mechanical Engineering and Information Technology*, 2012, pp. 1778–1783.

K. A. Danapalasingam, “Energy Optimization of Brushless DC Motor in Electric Power-Assisted,” *J. Teknol.*, vol. 1, pp. 1–5, 2014.

E. Lutz, “Future Trends for Automotive Steering Systems,” JTEKT Eng. J. English Ed., no. 1013, pp. 2–7, 2016.

Freescale Semiconductor, “Steering — Electronic Power Assisted.” Freescale Semiconductor, Inc., p. 4, 200

Towards Morphing Wing Technology in Aircraft for Improved Flight Performance

Emmanuel Attah Zebedee
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
emmanuelattah97@gmail.com

Ameer Mohammed
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
https://orcid.org/0000-00022699-3608

Muhammad Sharrif Lawal
Department of Aerospace Engineering,
Air Force Institute of Technology
Kaduna, Nigeria
mslawalg@gmail.com

***Abstract*—Conventional wing structures have no adaptation to changing flight conditions – they are optimized for specific flight conditions. An aerodynamically efficient configuration in one flight condition may be inefficient in another. As such, this work explores the architecture of a morphing wing and assesses the possibility of achieving a functional deformable wing by determining suitable sensors, actuators and controllers for implementation. The morphing wing technology adopts a closed-loop control whereby real-time sensing and monitoring of the wing shape is used to determine if the deformed shape is optimum under the flight condition. For this to be achievable, different sensors and their potential for use in the determination of pressure distribution are explored. Optical sensors were found to be the most reliable. To enable control of the morphing wing during flight, actuators were investigated. Piezoelectric actuators were found to be the most suitable for an adaptive and flexible wing structure due to their weight, precision and speed. In terms of controllers, fuzzy controllers were found to be the most computationally efficient. A combination consisting of optical sensors, piezo-electrical actuators and fuzzy controllers was found to be the most efficient for implementing the adaptive wing system. The benefits of the morphing wing technology depend on the development of robust and reliable sensors, actuators and controllers that facilitate the operation of the flexible wing structure.**

Keywords*—*Actuators, Controllers, Flight Performance, Morphing Wings, Sensors.

I. Introduction

Modern aircraft designs are optimized for a specific flight condition, their performance however over the entire flight mission is a compromise – a given wing configuration which is aerodynamically efficient in one flight condition may perform poorly in other flight phase[1]. The conventional wing box structures are rigid and have no adaptation for changing flight conditions. The movable surfaces on the wings which are largely the control surfaces, flaps, slats, spoilers, and some trailing edge devices such aileron and elevator, have limited degrees to which they can change the overall shape and consequently provide limited benefits compared to those obtainable from deformable or adaptable wing structures [1], [2]. Hitherto, several innovative concept of varying complexity for adapting the wing structure have been developed, one of which is the variable camber flaps developed with the view to optimize take-off and landing performances [3], [4]. Other concepts include wings capable of varying the angle of twist [5] while some have overall camber-morphing capabilities [6]-[10]. The germane thrust of these concepts is to achieve efficient aerodynamic flow control and substantial reduction in drag, from variable chord wing segments [11], to wings with variable aspect ratio [12] for planform control during aircraft mission. To enlist the associated benefits such as expanding aircraft flight envelope, reduction in profile drag to improve aircraft overall efficiency, endurance and range, reduction in vibrations and many more, it will be necessary to design structures capable of changing their shape continuously, while withstanding operative structural and aerodynamic loads.

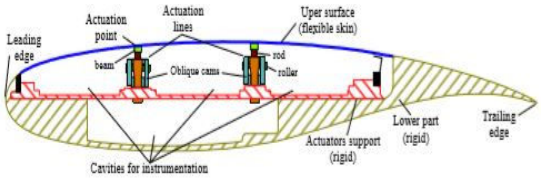


Fig.1: Cross Section of a Morphing Wing Model [12].

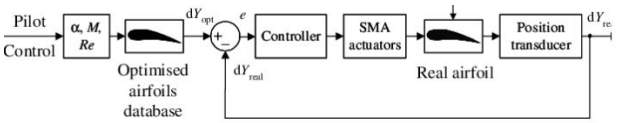


Fig. 2: Block Diagram illustrating the flow process and components of a morphing wing.

Morphing wings are design to have flexible and adaptive architectures, to ensure aircraft continuously achieve optimum flight performance [13]. "Metamorphic" solutions can be applied to various structural components with reference to several geometrical parameters in order to optimize specific performances: plan form alteration (span, sweep, and chord) [14-16], out-of-plane transformation (twist, dihedral and span-wise bending) [12], and airfoil adjustment (camber and thickness) [14]. The overwhelming arguments, that have led the research towards the adaptable wing structures, are essentially two: To improve aircraft overall efficiency (in terms of lift to drag ratio, pitching moment, gliding factor, power factor) and secondly to reduce fuel consumption [14]. To develop a morphing wing, we need a creative structural design that allows for extensive deformation as well as a high-performance sensor, actuator and controller combination [15]. Fig. 1 shows a cross-section of a morphing wing model. The lower part of the airfoil is essentially rigid, while the upper surface is deformable along the leading edge. Deformation of the upper surface is achieved by means of beams pressed against the upper surface at contact (actuation) point.

The objective of this paper is to analyse the architecture of a morphing wing, develop mechanisms to assess the possibility of achieving a functional deformable wing and the determination of suitable sensors, actuator and controllers for the actualization of the same. Fig. 2, shows a block diagram illustrative of the entire process and the requisite components. The process essentially starts when the pilot induces certain command in respond to a flight condition or maneuvers, this action correspondingly causes a change in either the angle of attack, Mach number or the Reynold number of the flow. Based on the pressure distribution over the airfoil, the system searches through the database of the various airfoil configuration to determine the most suitable airfoil shape for such flight condition. The controller is then prompted to send signals to the actuator for appropriate modification of the wing structure to conform to the preselected airfoil configuration that guarantees maximum performance for such flight condition.

The rest of the paper is organized as follows. Section II presents the sensing and measurement techniques that can be adopted. Section III gives more detail on actuators for morphing wing. Section IV provides the potential controllers that can be used. Future perspectives are provided in Section V with concluding remarks presented in Section VI.

II. Sensing and Measurement

A. Sensing Techniques

One of the greatest advantages of a morphing wing is drag reduction, consequently lesser fuel consumption. Drag reduction is achieved by changing the shape of the flexible skin, so that the laminar-to-turbulent transition point moves closer to the wing trailing edge based on the pressure distribution along the wing chord, as is shown in Fig. 1. This action, delays stall at the leading edge of the wing to a higher angle of attack (AOA), thus enabling the pilot to fly at a higher AOA with increased lift generation and little drag as possible. From the forgoing, a critical component of a morphing wing is the pressure sensors. A comparative analysis of different sensors and their potential for use in the determination of pressure distribution over a morphing wing surface is highlighted in Table I.

B. Sensing and Measurement Technology

Sensing systems are employed in morphing wing design with the sole objective of measuring aerodynamic loads via pressure distribution. From the measured sensory data applicable fluid mechanics and aerodynamic principle– principally Bernoulli's principle is used to calculate an optimum wing shape that guarantees best performance for such flight conditions. This initiates a chain of events in which the flexible skin and deformable structure of the wing are driven by means of actuators and control system to produce the desired deformation with a resultant wing shape that is aerodynamically optimize, efficient and ideal for the particular flight condition. Sensing system measures in real time, the wing shape and flight loads synchronously to actualize a closed loop control. The crucial role sensing system play in the morphing process, cannot be overemphasized, this is largely due to the fact that a real time sensing and monitoring of the wing shape is vital towards making an informed judgement as to whether the deformed shape is optimum for a specific flight condition. Thus, it is imperative to state that in order to guarantee the flight performance of a flexible morphing wing, the aerodynamic shape and pressure distribution over the surface of the wing must be monitored in real time using efficient and effective pressure and shape sensing technologies.

In the past decade, a number of sensing and measurement technologies have been developed and implemented in monitoring the geometrical shape of the wing. These technologies employ peculiar measuring techniques such as photo elastic method, vision measurement technique, the strain sheet measuring method, laser scan measuring system and the three-coordinate measuring technique, etcetera [5]–[9]. These techniques are rather conventional and while they are quite suitable for testing and measurement of an aircraft wing shape under static condition or during taxi, they are however incapable of real time monitoring of aerodynamic shape of a morphing wing during flight. suitable for the ground testing of the wing shape, but they are inapplicable for the real-time monitoring of the aerodynamic shape of the morphing wing during flight. This presents a unique challenge that necessitates the developing effective methods for an active, real time

monitoring and measurement of the wing shape in flight, to obtain optimum performance of the morphing wing aircraft in order to avail its benefits. A number of research work have elucidated a preference for optical fiber sensing technique as a promising candidate for the realization of a real time monitoring of an aircraft wing shape in flight [10]. The preference for optical fibre shape sensing technique over other conventional methods are succinctly listed below:

1. The optical fiber sensors, have some inherent incredible characteristics such as high sensitivity, small volume, short response time and lightweight. An interesting feature is the fact that they can be embedded into the flexible skin of the morphing wing, and able sensitively and rapidly measure the deformation of the wing using the strain parameter.to.
2. The distributed optical fiber sensing network is unique feature of these the type of sensor can be gainfully employed in the real-time measurement of the flexible deformation of the morphing wing, and the high rate and large capacity transmission of the measured data. This makes a case of the possibility a real time monitoring of the morphing wing shape in flight is realizeable.

TABLE III. COMPARATIVE ANALYSIS OF DIFFERENT SENSORS

Sensors	Limit Switch Sensor	Photoelectric Sensor	Inductive Sensor	Capacitive Sensor	Ultrasonic Sensor	Optical fiber sensor
Advantages	• High current capability • Low cost • Low-tech sensing	• Compatible with most materials • Durable • Longest sensing range • Very fast response time	• Resistant to harsh environments • Durable • Easy to install and use	• Can detect through barriers • Sensitive to non-metallic targets	• CoSmpatible with most materials	• High sensitivity, • Small in size • Vert fast response time • Lightweight
Disadvantages	• Requires physical contact with target • Very slow response • Contact bounce	• Lens subject to contamination • Sensing range affected by colour and reflectivity of target	• Distance Limitations	• Very sensitive to extreme environmental changes	• Resolution • Repeatability • Sensitive to temperature changes	• More fragile than other sensors • They are expensive • Requires precise installation
Applications	• Interlocking • Basic end-of travel sensing	• Packaging • Materials handling and parts detection	• Industrial • Machine tools • Senses metal-only targets	• Level sensing	• Anti-collision • Doors • Web brake level control	• Real time sensing and monitoring
Practicability for use	• Bulky	• Portable	• Bulky (but cheap)	• Bulky (with sensitivity and specificity)	• Portable	• Portable

3. Using Optical fiber sensor, the 3-D shape of the morphing wing can be visualized in real time during the deformation process by ccombing the optical fiber shape sensing technique with the three-dimensional (3-D) reconstruction and graph visualization system, This enables visualization and monitoring of the wing shape possible.

III. TRENDS ON ACTUATORS

To enable control of the morphing wing during flight, actuators are employed. The most common is the Smart Shape Memory Alloy (SMA) actuators which objectively deform wing camber to move the transition point from laminar to turbulent flow closer to the trailing edge [16]. This promotes a large laminar region on the wing surface, and thus reduces drag over an operating range of flow conditions characterized by Mach numbers, airspeeds, and AOA. The SMA actuator wires are made of nickel-titanium, they contract like muscles when driven electrically. SMA actuator have the ability to associate deflections with applied force, and they are capable of deforming into a variety of shapes and sizes, extremely useful

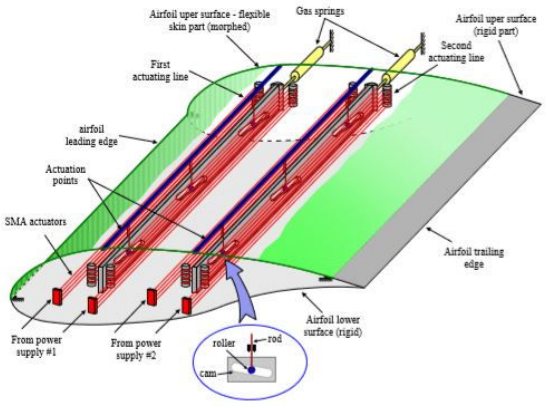


Fig. 3: Morphing Wing structure showing SMA Actuators and Recall (Gas springs) [11].

to achieve actuation goals of a morphing wing system. A morphing wing structure is usually made of composite consisting of two actuation lines as shown in Fig. 3. In Fig. 3, each of these actuation lines uses three shape memory alloys wires as actuators, which are connected to a current controllable power supply. Also, each line contains a cam, which moves in translation relative to the structure. The cam is responsible for the movement of a rod related on the roller and on the skin. Recall of the morphing wing is achieved by means of a gas spring. When the SMA heats up the actuator contracts and the cam is displaced to the right, this results in the rise of the roller and the displacement of the flexible skin upwards. In contrast, cooling the SMA causes a movement of the cam to the left, and a consequent movement of the skin down. This horizontal displacement of the actuator is converted into a vertical displacement at a rate 3:1 or 4:1 depending on the type of material used for the flexible outer skin. The following sections present the most common actuators that can be used with the morphing wing.

A. Pneumatic, Hydraulic and Electrical Actuators

Pneumatic linear actuators convert energy in the form of compressed air into mechanical motion. The motion can be rotary or linear depending on application requirement. Electric linear actuators take the rotational force of a motor (electrical energy) and convert it into linear movement (torque). A comparison of different actuators that could be potentially be used for morphing wing is in Table II. From Table II, hydraulic and pneumatic actuators cannot be used for a flexible wing design due to their actuation mechanism. Most of these actuators requires extra devices like compressors, power, pipes, and/or pumps which will consequently result in increased weight and reduced reliability. Furthermore, incorporating such actuators to a deformable wing could be cumbersome. Other more promising techniques such as piezo and bimorph actuation are currently being adopted in the development of adaptive and deformable wing structures. These are discussed in the proceeding section.

B. Other Actuators

TABLE IV. COMPARISON OF DIFFERENT ACTUATORS

Characteristics	Pneumatic Actuator	Hydraulic Actuator	Electrical Actuator
Performance	• Simple system composition • Moderate lifetime guarantees easy to replace if need be • Able to handle shock loads • High noise levels • Low Efficiency • Excellent Reliability • High amount of maintenance	• Moderately complex system composition • With proper maintenance it can last for a long lifetime • Explosion-proof, Shock-proof, Spark-proof • Hydraulic fluid leaks and disposal • Low efficiency • Good reliability • High user maintenance throughout the life of the system	• Control system and Motion component can work together in multiple complex configuration • With proper maintenance it can last for a long lifetime • Able to handle shock loads • Minimal noise levels • High efficiency • Good reliability • Little to no maintenance except for when replacements are necessary
Power	• High power • High load ratings	• Very high power • Extremely high load ratings	• High power • Can be high depending on the speed and positioning accuracy
Control	• Simple valves • Very difficult to achieve position accuracy	• Mid stroke positioning requires additional components and user support	• Flexibility of motion control capabilities with electronic controller • Positioning capabilities and velocity control allows for synchronization
Response rate	• Very high speeds • Very high acceleration	• Moderate speed • Very high acceleration	• Moderate speed • Moderate acceleration
Utilities	• Compressor, Power, Pipes	• Pump, Power, Pipes	• Power only option
Cost	• Low purchasing cost • Moderate operating cost • Low maintenance cost	• High purchasing cost • High operating cost • High maintenance cost	• High purchasing cost • Low operating cost • Low maintenance cost

1) Piezo-Actuators

Piezo electric materials produce electricity when stress is applied and conversely produces deformation when voltage is supplied. The later effect is used for linear actuation which can directly convert the input electrical energy into output mechanical energy and linear movement. Piezo-actuators have excellent operating bandwidth and can generate large forces from a compact size. They also display extreme positional accuracy of several nanometers. The problem however with piezo-actuators is that their application is limited because they have very small displacements. Nevertheless, their high precision and accuracy makes them potential candidates for use as sensors in wing morphing.

2) Bimorph Actuators

A bimorph is a cantilever used for actuation or sensing which consist of two active layers. During operation, one of the layers extends while the other compresses. This causes the element to bend. Some configurations have a passive layer between the two active layers. Bimorphs can be thermomechanical or can be piezo-bimorphs. Bimorphs have higher stroke value; however, with very limited force and low frequency. This makes them potentially useful as sensors for wing morphing.

IV. CONTROL OF A MORPHING WING

Controllers are used to control the deformation of the flexible wing. In Fig. 4, for each of the two actuation lines in the morphed composite, the open loop control architecture uses a controller which takes as reference value the required displacement of the actuators from a database of optimized airfoil shape stored in the computer memory. The structure of the actuation lines is identical, thus both lines are connected to the same controller. For feedback the position signal from a linear variable differential transducer (LVDT) is connected to the oblique cam sliding rod. This method however, does not take direct information from the pressure sensors regarding the wind flow characteristics and hence classed as an "open-loop control".

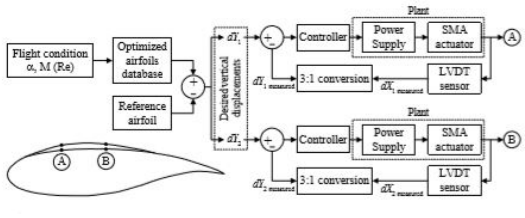


Fig. 4: Open loop Control Architecture [12].

In Fig. 5, for each flight condition, a pair of optimal vertical deflections (dY1opt, dY2opt) for the two actuation lines is quite ostensible. The SMA actuators morph the airfoil until the vertical deflections of the two actuation lines (dY1real, dY2real) became equal to the required optimal deflections (dY1opt, dY2opt). A two-position transducer is used to measure the vertical deflections of the real airfoil, while the controller's role is to initiate and send a command to supply an electrical current signal to the SMA actuators. The electrical signal generated is based on the error signals (e), between the required optimal vertical displacements and the obtained displacements. The designed controller as stated earlier is valid for both actuation lines, which are practically identical. The PD fuzzy logic controller holds a lot of promise for implementation in morphed composite structure. It has as output the voltage that controls the Power Supply output current, while having two inputs, first the error (difference between the desired and measured vertical displacement) and second, the change in error (the first derivative of the error) [17]. Another of such prodigious controller is the Mamdani type, widely accepted for capturing expert knowledge. It finds enormous application, due to its simple structure of "min-max" operations [18]. Generally speaking, fuzzy controllers (PD and Mamdani type controllers) have the potential to achieve a level of expertise close to (and possibly better than) human expertise in actuator control. However, the capabilities of fuzzy control are dependent on the level of sophistication of its rules and

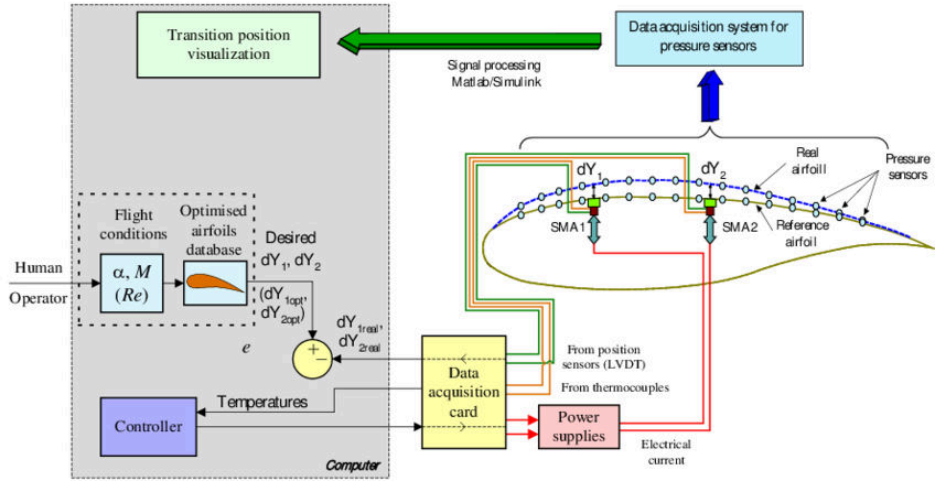


Fig. 5: Architecture of an Open Loop Morphing Wing model [11].

input signal. This makes the case of sensory data and its pre-processing critical to the application of these sorts of controllers.

V. FUTURE PERSPECTIVES

The enormous benefits of the morphing wing hinges on the development of robust and reliable sensors, actuators and controllers that facilitates the operation of the flexible wing structure. A classic and tragic example of the critical role of sensors is the recent crash of Boeing 737 MAX Ethiopian airline due to faulty sensor reading relied on by the Maneuverability Control Augmentation System (MCAS). MCAS is an automated safety feature designed to prevent the plane from entering into a stall, or losing lift by trimming the aircraft stabilizer nose down. It relies on Angle of attack sensors which tells it to automatically point the nose of the plane down if the airplane is in danger of going into a stall.

To overcome tragedies such as that of the Boeing 737 MAX, there needs to be more redundancy and robustness in the sensors. For this, multiple sensors can be adopted for use in the morphing wing. Where by one sensors measures short term (fast) changes and the other measures long term (slow) changes in flight performance. By blending the right sensing modalities, the efficiency of the adaptive wing may be improved. So far, optical sensors have proven to be the most reliable compared to other categories of sensors. Their greatest advantages are high sensitivity, electrical passiveness, freedom from electromagnetic interference, wide dynamic range and multiplexing capabilities. Thus, optical sensors hold a lot of promise and application in this area and are increasingly being adopted and thus suggest that a more thorough research work devoted to finding their application in sensing pressure in a deformable wing would prove beneficial. On actuators, piezoelectric actuator are best suited for an adaptive and flexible wing structure due to their light weight and their actuating precision and fast reaction. They have excellent responsiveness and conversion efficiency from electrical to mechanical energy. These properties not only allow them to be employed in morphed wing technology but also for ultra-small-scale precision motion devices, micro actuator medical tools and in dynamically driven, high temperature actuators. The cost and difficulty of manufacture of this sort of actuators is rather quite high. Due to their high cost, an alternative can be bimorphs. Bimorphs provide higher stroke values compared to piezo-actuators, even though they have limited force and frequency.

VIII. CONCLUSION

The novel and emerging technology of deformable, adaptive and flexible wing structure was presented as the solution for aerodynamically efficient flights. For this to be achieved, the right sensors, actuators and controllers need to be interfaced. During flight, pressure sensors are used to detect pressure distribution over the wing. These are relayed to the actuators through the controllers to enable the adjustment of wing camber. Based on the analysis in this work, a combination consisting of optical sensors, piezo-electrical actuators and fuzzy controllers was found to be the most efficient for implementing the adaptive wing system. The potentials of the adaptive wing technology are limitless; thus, future study is imperative to truly harness the benefits they may offer.

REFERENCES

[10] B. Jenett, S. Calisch, D. Cellucci, et al., “Digital morphing wing: active wing shaping concept using composite lattice-based cellular structures, ” Soft Robot. Vol 4(1) pp. 33–48, 2017.

[11] M.D. Luca, S. Mintchev, G. Heitz, et al., “Bioinspired morphing wings for extended flight envelope and roll control of small drones,” Interface Focus 7, 20160092, 2017.

[12] S.A. Meguid, Y. Su, Y. Wang, “Complete morphing wing design using flexible-rib system,” Int. J. Mech. Mater. pp. 159–171, 2017.

[13] S. Kota, R. Osborn, G. Ervin, D. Maric, “Mission adaptive compliant wing – design, fabrication and flight test,” RTO Applied Vehicle Technology Panel (AVT) Symposium 2009.

[14] Y. Li, Y. Wang, C. Wen, “Temperature and strain sensing properties of the zinc coated FBG,” Optik—Int. J. Light. Electron Opt. 127 (16) 6463–6469, 2016.

[15] Y. Chen, Y. Cheng, H. Liu, “Application of improved wavelet adaptive threshold de-noising algorithm in FBG demodulation,” Optik—Int. J. Light Electron. Opt. 132 (16) 243–248, 2017.

[16] V. Mishra, M. Lohar, A. Amphawan, “Improvement in temperature sensitivity of FBG by coating of different materials,” Optik—Int. J. Light Electron. Opt. 127 (2) 825–828, 2016.

[17] A. Gautam, A. Kumar, R.R. Singh, V. Priye, “Optical sensing and monitoring architecture for pipelines using optical heterodyning and FBG filter,” Optik—Int. J. Light Electron Opt. 127 (20) 9161–9166, 2016.

[18] T. Tao, Q. Chen, J. Da, S. Feng, Y. Hu, C. Zuo, “Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system,” Opt. Express 24 (18) 20253–20269, 2016.

[19] J. Ge, A.E. James, X. Li, Y. Chen, K.W. Kwok, M.P. Fok, Bidirectional soft silicone curvature sensor based on off-centered embedded fiber bragg grating, IEEE Photonics Technol. Lett. 28 (20) 2237–2240, 2016.

[20] Pecora, R., Magnifico, M., Amoroso, F., Lecce, L., Bellucci, M., Dimino, I., Concilio, A. and Ciminello, M., SARISTU - Smart Intelligent Aircraft Structures, chapter 2.2 “Structural Design of an Adaptive Wing Trailing Edge for Large Aeroplanes”, Springer International Publishing AG, Switzerland, 2015.

[21] Barbarino, S., Bilgen, O., Ajaj, R.M., Friswell, M.I., Inman D.J., “A Review of Morphing Aircraft”, Journal of Intelligent Material Systems and Structures, Vol. 22, pp. 823 -877, 2011.

[22] Monner, H. P., Bein, T., Hanselka, H., and Breitbach, E., “Design Aspects of the Adaptive Wing -The Elastic Trailing Edge and the Local Spoiler Bump”, Proceedings of Royal Aeronautical Society Symposium on Multidisciplinary Design and Optimization, Royal Society Publishing, London, pp. 15.1–15.9, 1998.e

[23] Pecora, R., Barbarino, S., Concilio, A., Lecce, L., and Russo, S., “Design and Functional Test of a Morphing High-Lift Device for a Regional Aircraft”, Journal of Intelligent Material Systems and Structures, Vol. 22, No. 10, pp. 1005–1023, 2011.

[24] Pecora, R., Amoroso, F., and Lecce, L., “Effectiveness of Wing Twist Morphing in Roll Control”, Journal of Aircraft, Vol. 49, issue 6, pp. 1666-1674, 2012.

[25] M. Di Luca, S. Mintchev, G. Heitz, F. Noca and D. Floreano, “Bioinspired morhing wings for extended flight enevelop and roll control of small drones” interface focus 7: 20160092, 2017.

[26] Ameduri, S., Concilio, A., Pecora, R., “A SMA-based morphing flap: conceptual and advanced design”, Smart Structures and Sytems,Vol. 16, No. 3, 2015.

[27] Amendola, G., Dimino, I, Concilio, A., Pecora, R., Amoroso, F., "Actuation System Design for a Morphing Aileron", Applied Mechanics and Materials, Vol. 798, pp. 582-588, 2015.

Perkins, D. A., Reed, J. L., and Havens, E., "Morphing Wing Structures for Loitering Air Vehicles", Proceedings of the Forty-Fifth AIAA Conference on Structures, Structural Dynamics and Materials, AIAA Paper 2004-1888, April 2004.

Implementing Flash Event Discrimination Policy in Internet Protocol Traceback using Shark Smell Optimization Algorithm

Omoniyi Wale Salami
Department of Computer Engineering
ABU, Zaria

Ime Jarlath
Department of Computer Engineering
ABU, Zaria

Emmanuel Adewale Adedokun
Department of Computer Engineering

Muhammed Bashir Mu'Azu
Department of Computer Engineering
ABU, Zaria

Lukman Adewale Ajao
Department of Computer Engineering
ABU, Zaria

Abstract*—An Internet Protocol (IP) traceback tool can wrongly identify a Flash event (FE)flow path as a Denial of service (DoS) attack path when tracing a DoS attack source because of the symptomatic similarities between them. IP traceback scheme should be able to differentiate FE flow from DoS attack flow to avoid mistaking FE for DoS attack during a DoS attack traceback process. Discrimination policy was introduced into IP traceback scheme to address this challenge, but the discrimination policy can generate conflicting results. This work proposed improvement to the discrimination policy implementation by incorporating statistical analysis of attack packets distribution on each edge along the path to select the edges with the highest attack packets. The results of different tests carried out show that the proposed solution reduced attack path detection error by 32.35%, path detection accuracy by 22.19%, but increased convergence time by 0.65% than the benchmark solutions.*

Keywords*—Network forensic, Shark Smell Optimization, Nature Inspired Algorithm, IP Traceback, Flash Event*

IX. *INTRODUCTION*

Network attacks such as unauthorized use of restricted online assets without permission, stealing or gaining unauthorized access into a system, exposing private resources, or malicious disabling or altering or destroying services of a system on the network, are cybercrimes [1]. Denial of service (DoS) attack and its variants are the most powerful damaging network attack used to harm a business or organization [2]. Different countries of the world, including Nigeria, have enacted laws and policies to fight the scourge of cyberattacks. Examples are the United States Stop Online Piracy Act and Protect IP Act (SOPA/PIPA) [3]. The UK Data Protection Act [4], and the Nigerian cybercrime act 2015 [5]. To be able to institute legal proceedings against the perpetrator of DoS attacks in a court of law the perpetrator must be identified. Successful legal engagement against the perpetrator can deter others from the act. Also, it can make it possible to be able to compensate the victim. Successful legal action can only be achieved based on proven infallible facts that establish criminal offence against the accused. Network forensic professionals use Internet Protocol (IP) traceback tools to acquire network data that can be used as facts about an attack and also used to detect the source of the attack.

The process by which data packets are traced back to their source using available internet protocol information on the packets, e.g. source IP address, is called IP traceback. IP traceback scheme should be specifically designed to take into consideration the intricate attributes of DoS. It should be able to discriminate DoS from normal transactions on the network that may cause traffic surge on the network, such as flash event (FE) [6]. Flash event is symptomatically similar to Distributed DoS (DDoS) attack, which is a variant of DoS. Flash event (or flash crowd [6]), refers to a situation whereby majority of network users are accessing a particular network resource on a server [7]. A good example of flash event when legitimate traffic overwhelmed the network was the news of Michael Jackson's death on June 25, 2009. CNN reported that "Michael Jackson's death sees Twitter, TMZ, news sites struggle to cope with traffic" [8]. Both flash event and DDoS attacks generate heavy traffic from different sources to a particular server. Flash event can be mistaken for DoS attack because of their similarities [9]. Both of them cause network flooding and reduce throughput. Careful analysis of the FE traffic and DoS traffic can show some of their differences. Flash event characteristics like rate of request from the same source IP address, the timing between request packets arrivals, the sizes of request packets and their contents, and the relationship between packets should be different from those of DoS. Also, packets traffic features such as delays, throughput, packets sequences and entropy, and their randomness may show if they are from the same node or different nodes [10]. Although, their traffics patterns may look very similar. Usually, there are usually more packets per IP address in the DoS traffic than what may occur in the case of flash event traffic [9]. Thus, entropy will be wider for the flash event traffic than in DoS attack traffic. In the case of the DoS attack the attacker has good knowledge of the system it is attacking and deliberately overloads it by exploiting its known vulnerability. But flash

event is just an accidental phenomenon caused by users' attraction to some online items on a server.

Some of the major differences between flash event and DoS attack are;

1. In contrast to the purpose of DoS attack, flash event is not perpetrated to inflict bad consequences on the affected host but happens as a result of large visitors being attracted to the items on the host.

2. DoS attacker knows and targets the major metrics that can impact on the performance of the network it is attacking. It understands the impact of its action on the network. Users causing flash event are unaware of the effect of their activities on the network.

3. Correlation between DoS traffic packets is stronger than those of flash event traffic packets [9]. This is evident because flash event traffic packets generation is purely random and naturally unpredictable but DoS attack packets are generated automatically with or without deliberate randomization.

Although FE may cause significant latency on the network that may disrupt the network services [11]. Identifying the individuals at its sources is unwarranted. The solution to it is a better planning and upgrading of the capacity of the network to be able to cope with such situations. Thus, efficient IP traceback scheme should be able to differentiate flash event traffic from DoS attack traffic during the traceback process. The only solution that was found in literature that specifically address the problem of FE on IP traceback was SSOA-DoSTBK [12].

There are various methods for implementing IP Traceback scheme [13]. Some of the common methods available in the literature are; Packet marking that may be Deterministic Packet Marking (DPM) or Probabilistic Packet Marking (PPM). DPM has large convergence time and PPM requires large number of attack packets to reconstruct attack path. Packets marking method was used for Efficient Traceback Technique (ETT) for detecting DDoS attack in cloud-assisted healthcare environment [14]. The traceback solution that uses completion condition to determine the minimum required packets for IP traceback [3] was based on packets marking method as well.

In recent time some IP traceback schemes are developed based on nature inspired algorithms. Nature inspired algorithms are developed for complex computing where exact results may be hard to achieve or not determinable due to non-availability of enough input parameters for conventional computation [15]. Venkataramanan & Ravi [16] used particle swarm system for their IP traceback solution. Particle swarm system is an enhanced Particle Swarm Optimization (PSO) algorithm with local updating rule incorporated in addition to the global updating rule of the original PSO. Saini et al. [17] developed hybrid IP traceback mechanism using two nature inspired optimization algorithms, namely, Ant Colony Optimization and Particle Swarm Optimization. The two optimization algorithms are swarm algorithms and known to be effective in solving combinatorial optimization problems. The PSO was used to improve the convergence rate and reduce the computational complexity of ACO algorithm. The improved Ant Colony Optimization (ACO) algorithm solution was used for solving IP traceback by [18]. It used ant colony optimization algorithm for flow-based traceback. Most of the IP traceback solutions developed using nature inspired algorithms are flow-based solutions. They can easily mistake FE traffic surges found on the traced path for attack and return wrong attack paths. Salami et. al. [12] proposed SSOA-DoSTBK for mitigating the thechallenge of flash event causing error in IP traceback results. SSOA-DoSTBK used Shark Smell Optimization Algorithm to implement discrimination policy for discerning DoS attack flow from Flash Event flow during DoS attack source detection process. SSOA is a single agent nature inspired optimization algorithm that was developed to be used for obtaining optimal search results. The fundamental concepts of this algorithm were based on the superior ability of shark to find its prey in a large search space, the sea, within a limited period of time using its strong smell sense [19]. The algorithm comes in hand for a quick scrutiny of large data with close similarities to select the ones with the best fit within a limited time constrain. Like other nature inspired algorithms, the SSOA user-define parameters can be used to control its speed and results accuracy.

This research work developed an improved IP traceback scheme called iSSOA-DoSTBK. iSSOA-DoSTBK is an improvement to SSOA-DoSTBK [12]. The discernment policy used for distinguishing FE traffics from DoS traffics is found to be susceptible to generating conflicting results as explained in section II. The proposed solution mitigates the deficiency in the discrimination policy of SSOA-DoSTBK [12] using statistical analysis of the attack packets distribution along the flow path. This enhances detection of the genuine source of the attack by avoiding the situation of conflicting results that may mislead the IP traceback. This will ensure detecting the original source of a DoS attack in order to be able to take appropriate actions against the real perpetrator of the attack.

The contribution of this work is the incorporation of distribution analysis into the discernment policy implementation using variation of the number of attack packets on the nodes along the attack path to correctly distinguish attack traffics from other surges.

By mitigating the effects of discernment policy generating conflicting results the benefits are;

1. making IP traceback scheme to efficiently avoid flash event and other legitimate flows that may be symptomatically similar to an attack traffic.

2. enhances IP Traceback focus on attack data and obtain more accurate results.

3. effectively mitigating the effects of distinguishing the sudden traffic surge in a network from the DoS attack traffic during traceback process.

X. RESEARCH METHOD

The prototype of the iSSOA-DoSTBK proposed in this research was simulated on a laptop PC; x64-based processor,

Intel Core i3-3110M CPU 2.4 GHz, 8.0 GB RAM, running Ubuntu 18.04 LTS operating system. Network Simulator version 2 (NS2), ns-allinone-2.35, was used for the simulation.

TABLE VI. WEIGHT VALUES ASSIGNED TO DISCRIMINATION PARAMETERS

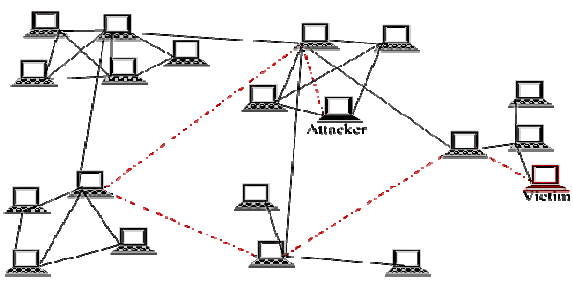
Parameter	NS2 Parameter	Value
ICMP	Pkt Type	4
IP-ID	Pkt Id	5
MAC-addresses	Not used	
Packet-Length	Pkt_sz	1
TTL	Not used	
VLAN-ID	Not used	
Dst IP address	Dst Node Id	4
Dst Port Msk AS	Not used	
Pkts	Packets Count	No Value (NV)
Active	Host log Time	1
NextHop	Next hop node	2

C++ and OTCL were used for the necessary programming. MS Excel 2016 was used to generate the charts.

A Waxman network topology model consisting of randomly placed 200 nodes at the coordinate points in the 20 × 20 meters network area was generated using random topology generator in NS2. The WaxMan's probability was used to establish connection links attributes, including transmission rates, bandwidth between nodes, and spatial distribution of nodes as;

$$p(i, j) = \eta . e^{\left(\frac{d(i,j)}{L\gamma}\right)} \quad (1)$$

The WaxMan's connectedness probability p(i, j) was calculated in (1) [18] to establish links between the nodes. The nodes distances apart, d(i,j) was set to 1 to connect each node to its one-hop distance neighbouring node. L is the longest possible distance between any two nodes. Different values of η were selected from [0.3, 0.6, 0.9, 1.2, 1.5] for different generations of 100 simulations each in that order. The value of γ = 0.1 was used for all values of η selected. Ping and messaging agents were set up using OTCL classes. Messaging was used for generating normal network transaction between other nodes. Ping agent was used to generate the ping of death attack used as the DoS attack by the randomly selected attackers on the randomly selected victims. Other communication agents including TCP, FTP, and CBR were established for different network transactions on the network. Fig. 1 shows a simple network with randomly connected nodes showing an attacker and a victim, and the attack path (red dotted line) through which the attack packets were transmitted. The attack paths and routing information were generated using Monte Carlo method [20].



Fig, 1. A randomly connected network

A. *The Discrimination policy*

The expression for determining the fitness of an edge been on attack path was developed based on set criteria as explained in the following.

A set of unique parameters of the attack packets flow shown in Table 1 [12] was used to set the discrimination policy. The column under "Parameter" shows the parameters of the original packets on the real network. The column under "NS2 Parameter" shows the equivalent of the real network packet parameters as it is in NS2. The "Value" column shows the weights assigned to the parameters in the discrimination policy.

The discrimination policy parameters were stored in an array, x_o

$$x_0 = [4,5,*,1,*,*,4,*,0,1,2] \quad (2)$$

The elements in x_o in (2) are arranged to corresponds to the position of the parameter assigned the weight in Table 1. The parameter with no weight assigned has value * in the array.

When reconstructing the attack path, each neighbouring node is examined with respect to the flow parameters stored in the array, x_o, and matching parameters were assigned values according to its corresponding weights. The value obtained for each parameter of packets flow records on each node for each edge *e* examined was stored in $x_{e,q}$.

$$\chi_e = \sum_{q=1}^{Q} (x_{e,q} \cap x_0) \quad (3)$$

Q is the total parameters in the discrimination policy

The sum of matching parameters of all the attack packets on the edge e is determined by (3) [12] as χ_e.

$$\chi_j = [\chi_e, \cdots, \chi_E] \quad (4)$$

$1 \le e \le E$, E is the total edges on the node

Each χ_e for each edge on node j on the attack path was stored as an element in the array χ_j in (4). It is considered that there may be two or more χ_e values in (4) that may be

accidentally the same. This is because different sets of numbers can sum up to the same value, e.g. the additions (5 + 5), (3 + 7), (6 + 4) all give 10. When there are multiple elements of χ_j that are the same, this will cause conflict of results if the values indicate multiple edges as equally likely to be attack path. There will be the need to further process the data to pick out the most likely solution among the equally likely solutions. Further steps were considered here to avoid this situation of results collision that may confuse the IP traceback in determining the correct attack path. The variations and distribution of the discriminating parameters on the nodes involved were incorporated as follows.

The average value, β_e, of all the χ_e of the edges connected to each node is calculated.

$$\beta_e = \frac{\chi_e}{E} \quad (5)$$

E is the total edges on the node

Equation (5) determines the most common set of values. It will be small if smaller values are more than bigger values, or big if otherwise. But it will still be the same for χ_e values that are equal if those χ_e values have the same number of terms E, i.e. nodes with same number of connected edges.

Since the occurrence of traffic flows in a packet switch network is random, it is assumed that the path that different attack packets will follow may not be the same. Thus, those χ_e with same values may not have exactly the same set of addends from the packets of the flows. Variance can show the differences in the terms in different χ_e. It will also magnify their differences that are very close.

$$\alpha_e = \sum_1^M \frac{(x_{e,m}-\mu)^2}{N} \quad (6)$$

M is the total number of matching parameters on edge e, *N* is the number of packets examined, μ is the mean of all matching parameters obtained for the edges on hop j

The variance of the distribution of the sums of χ_e on all edges was obtained with (6). The variance will be smaller for the edge that transmits a relatively constant number of attack packets in it traffics than the one that intermittently transmitted different numbers of the attack packets.

$$P_e = \frac{\chi_e}{\sum_{e=1}^{E}(\chi_e)} \quad (7)$$

The probability of edge e been on the attack path, P_e was calculated using (7).

$$f(\chi_j) = max\left[\left(\frac{\beta_e}{\alpha_e}\right).\left(\frac{\chi_{j,e}}{\alpha_e}\right).e^{\left(\left(\frac{\chi_{j,e}}{\alpha_e}\right)^{p_e}\right)}\right]_{e=1}^{E} \quad (8)$$

$\chi_{j,e} \geq 0$

Using (4) to (7) the fitness of the edge been on the attack path, $f(\chi_j)$, is determined from (8). In (8), $\left(\frac{\beta_e}{\alpha_e}\right)$ further differentiates edges with same values of χ_e based on the distributions of their matching parameters. It produces larger values for the edges that constantly carried a large number of attack packets than the one that intermittently transmitted it. As the shark advances from the victim towards the attacker, at every node (*j*) on the attack path, each edge e carrying packets to (*j*) was examined using (3) and (8). Equation (8) normalises the computation into a small range of values that can be easily and quickly handled.

B. The Shark Smell Optimization Algorithm

SSOA was proposed by Abedinia et al. [19]. The fundamental concepts of this algorithm were based on the superior ability of shark to find its prey in a large search space, the sea, within a short period of time using its strong smell sense. The assumptions considered for the modelling of the artificial shark by Abedinia et al. [19] are:

1. The injured fish (prey) ejecting the blood into the sea is relatively static compared to the shark's motion.
2. The blood is being continuously ejected and spreads with decreasing concentration outwardly. Thus, shark follows the increasing intensity of the blood smell from the point of the smell detection towards the injured fish.
3. Only one injured fish is ejecting blood in the search space at the time of the search.

A set of position vectors consisting of possible initial solutions is normally generated for the initial position of the shark. The set of the initial position vectors here was derived from (4). The shark's velocity to position *j* at stage k of iteration *i* is selected as the minimum between calculated value with the gradient function and the value weighted with the shark's natural velocity limiter at stage k, as;

$$|v_{i,j}^k| = min\left[\left|\eta_k . R1 . \frac{\partial(OF)}{\partial x_j}|_{x_{i,j}^k} + \alpha_k . R2 .\right|, |\beta_k . v_{i,j}^{k-1}|\right] \quad (9)$$

Equation (8) is the objective function (OF) in (9) [19]. The derivative, $\frac{\partial(OF)}{\partial x_j}$, was derived with respect to the elements of χ_j in (4). The parameters η_k, $R1$, α_k, $R2$ and β_k are optimization algorithm parameters that can be used to fine tune the performance of the algorithm. The default values of the parameters as used in [19] were adopted for this research. $\eta_k \in [0,1]$, $R1 \in [0,1]$, $\alpha_k = [0,1]$, $R2 \in [0,1]$ and $\beta_k = 4$

The new position of shark, Y_i^{k+1}, at every stage is dependent on its previous position, X_i^k, and the velocity (V^k) within the time interval (Δ_{tk}) moving from X_i^k to Y_i^{k+1}. Thus, the new position is given as

$$Y_i^{k+1} = X_i^k + V_i^k \cdot \Delta_{tk} \quad (10)$$

For convenience Δ_{tk} is set as $\Delta_{tk} = 1$ in equation (10) [19].

Shark execute a local search at each stage by moving round the position to exploit all the points. Shark uses the point with the highest smell to proceed to the next position. The shark exploitation is modelled as:

$$Z_i^{k+1,r} = Y_i^{k+1} + R3.Y_i^{k+1} \quad (11)$$

R3 in (11) is another random number with uniform distribution in the range [−1, +1], r is the local position counter in the local exploitation.

The point with the strongest smell among R positions in the local search, which is the optimal solution, is finally selected thus:

$$X_i^{k+1} = argmax\{OF(Y_i^{k+1}), OF(Z_i^{k+1,1}), \ldots, OF(Z_i^{k+1,R})\} \quad (12)$$

Equation (12) [19] returns the most probable edges that are segments of the attack path together with their end nodes.

C. iSSOA-DoSTBK Attack Source Traceback Process

i) Get attack packets attributes from an attack detector
ii) Set rules for the discrimination policy from the attack packets attributes
iii) Detect edges that carry ingress traffic to the present node
iv) Detect parameters of each packet on neighbouring node connected by each ingress edges that matched discrimination policy rules using (3)
v) Store matching parameters detected in step (iv) using (4)
vi) Calculate the average value of attack attributes found on each neighbouring node using (5)
vi) Calculate the variances of the attack packets attributes on each neighbouring node using (6)
vii) Calculate the probability, *Pe*, of edge e been on the attack path using (7)
viii) Estimate the attack packet distribution on each neighbouring node using (8)
ix) Select the ingress edge with highest attack packets using (12)
x) Add the edge selected in step (ix) and neighbouring node it connects to the attack path
xi) If this node is neither the attack source nor the closest router to the source of the attack packets move to the node on the other end of the edge selected in step (ix)
xii) Repeat steps (iii) to (xi)
xiii) Otherwise Stop
xiv) RETURN attack path and the nodes on the path

XI. Results and Discussion

The performance was measured in terms of correctness based on the number of attack packets available on the returned path as was used in [12], [18]. The results obtained for iSSOA-DoSTBK were compared with similar results obtained from ACS-IPTBK and SSOA-DoSTBK for the same tests under the same conditions. Performance in attack packets detection and convergence time were used as the metrics for comparing the iSSOA-DoSTBK with the benchmarks. These values form the data that were used for calculating *percentage improvement* and *correctness* used for plotting the results.

A. Performance Evaluation

The performance of the iSSOA-DoSTBK was calculated and compared to the ACS-IPTK and SSOA-DoSTBK that it was benchmarked with as follow;

To illustrate the computation of the comparison metrics, if the average of the results obtained from iSSOA-DoSTBK in a test is S_{AVE} and A_{AVE} was obtained from ACS-IPTBK in the same test, e.g. average of attack packets on paths returned by iSSOA-DoSTBK is S_{AVE} and for ACS-IPTBK is A_{AVE},

$$pecentage\ improvement = \frac{S_{AVE} - A_{AVE}}{A_{AVE}}\% \quad (13)$$

The comparison of the average performance of iSSOA-DoSTBK over ACS-IPTBK and SSOA-DoSTBK is calculated with (13) [12].

$$Correctness = \frac{\frac{Average\ attack\ packets\ on\ returned\ path}{Total\ packets\ routed\ on\ the\ path}}{} \times 100\% \quad (14)$$

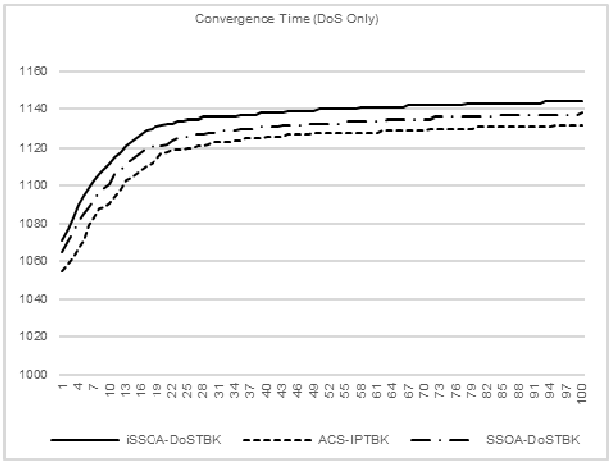
The performance of the schemes in a traceback of the simulated attacks under different conditions was measured in terms of correctness of the attack path returned by examining the number of attack packets on the returned path using (14) [12].

B. Analysis of the Results

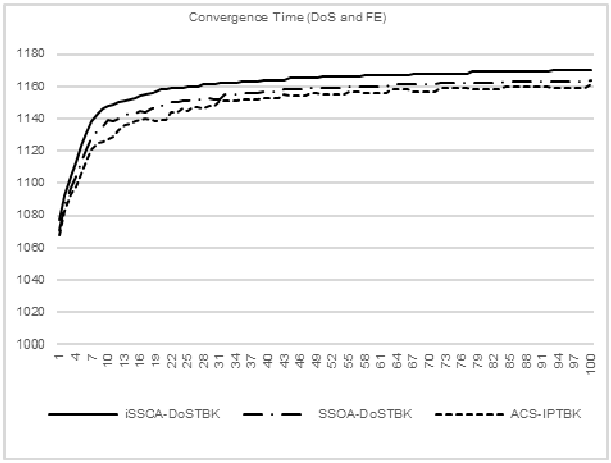
The tests were carried out under three conditions which are; (1) DoS source was traced when there is no flash event in the network (DoS only), (2) DoS source was traced when there were flash events in the network (DoS with FE), (3) source of DoS attack with spoofed packets was traced when there were flash events in the network (Spoofed DoS with FE). Two dimensional (2D) graphic plots and bar charts of the results are presented for a clearer view of the performance of the iSSOA-DoSTBK and benchmarks solutions.

1) Convergence Rate

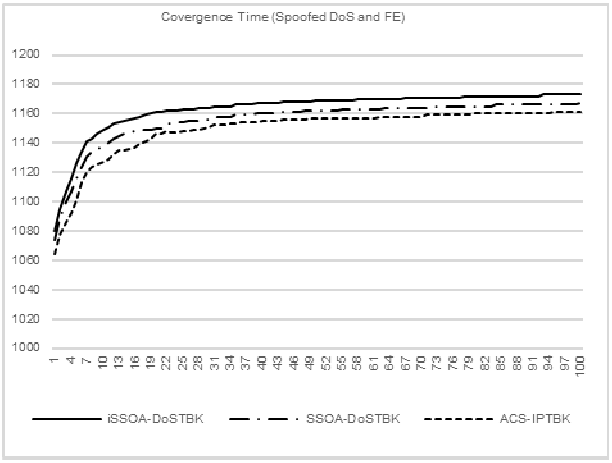
The time to converge by each of the solutions was compared in Fig. 2a(i) to 2a(iii). iSSOA-DoSTBK took a slightly longer time to converge as shown in the figures. The iSSOA-DoSTBK undertakes more computations to deeply scrutinize each node for the attack packets than the other two benchmark solutions. Similar to preying behaviour of shark whereby it traces a particular blood smell, iSSOA-DoSTBK used (3) to select the particular attack traffic. It used (5), (6) and (7) to check the distributions of the attack packets on each node it examined. It determined the node that transmitted most of the attack packets with (8). This process accounted for the longer time it took to return a solution.



(i). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK convergence time for tracing DoS in the network



(ii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK convergence time for tracing DoS with FE only in the network



(iii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK convergence time for tracing Spoofed DoS with FE in the network

Fig. 2a. Comparison of iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK Convergence Rate

As shown in Fig. 2b the average percentage increase in time taken by iSSOA-DoSTBK compared to the time taken by SSOA-DoSTBK were 0.65%, 0.64%, and 0.64% for DoS only, DoS and FE, and Spoofed DoS and FE tests respectively. While it was 1.24%, 1.03% and 1.22% respectively for the same tests compared to ACS-IPTBK Convergence time.

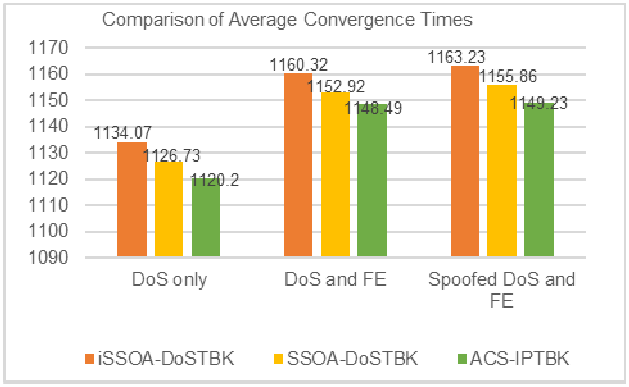
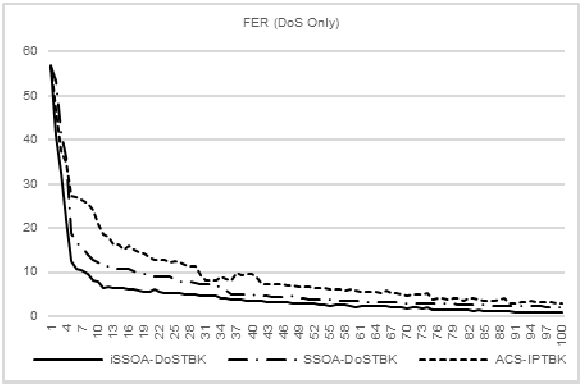


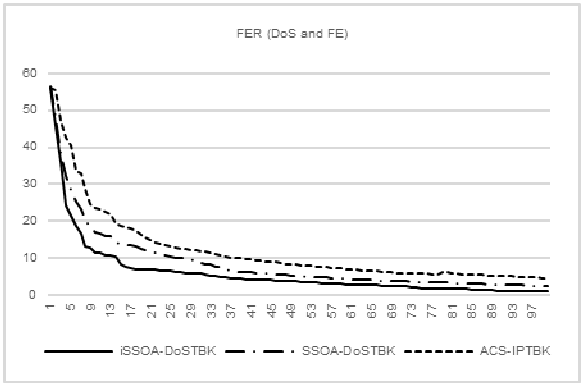
Fig. 2b. Comparison of average of times taken by iSSOA-DoSTBK, SSOA-DoSTBK, and ACS-IPTBK to return attack paths

2) False Error Rate

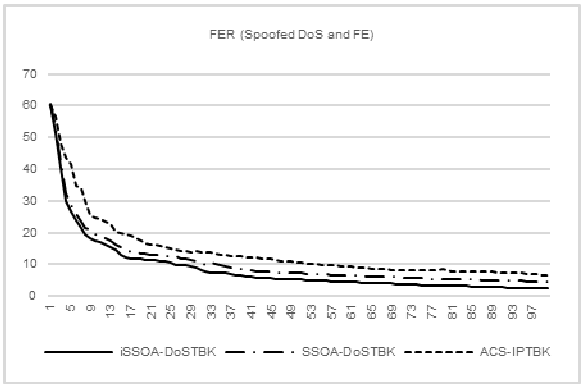
The false error rate indicates the number of attack path segments that are not detected and segments that are wrongly detected as part of the attack path by an IP traceback solution. This metric shows the ability of an IP traceback to detect true attack paths. The comparison of the results from the iSSOA-DoSTBK and the benchmark solutions are shown in Fig. 3a(i) to 3a(iii). The least errors were found in the paths returned by the iSSOA-DoSTBK because of the deeper scrutinization introduced into the solution.



(i). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK 2) False Error Rate for tracing DoS in the network



ii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK 2) False Error Rate for tracing DoS with FE in the network



iii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK 2) False Error Rate for tracing Spoofed DoS with FE in the network

Fig. 3a. Comparison of iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK False Error Rate Data.

The comparison of the average false error rates recorded for iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK are shown with bar chart in Fig. 3b. It can be seen that the false error rates for iSSOA-DoSTBK were less than those recorded from results of SSOA-DoSTBK by 32.35%, 29.80%, and 19.55% for DoS only, DoS and FE, and Spoofed DoS and FE tests respectively. The errors obtained from the results of ACS-IPTBK were higher than those obtained from iSSOA-DoSTBK by 51.51%, 49.75%, and 39.42% for the three tests respectively.

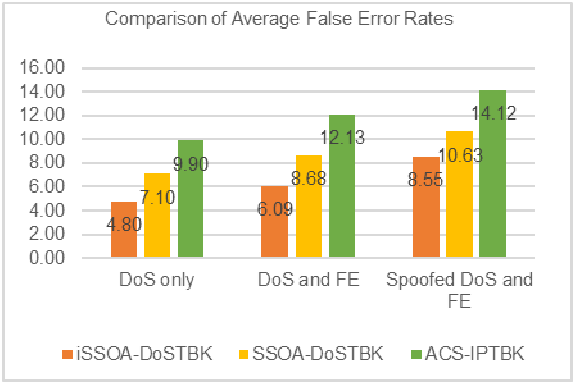
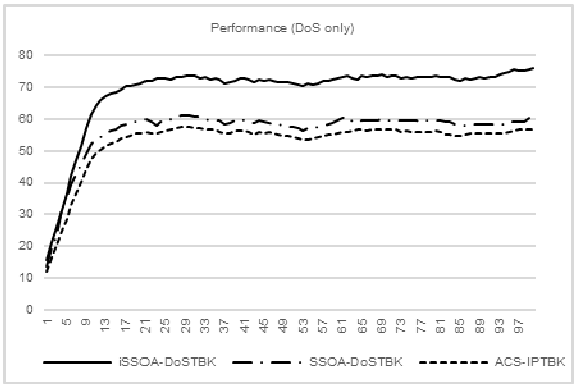


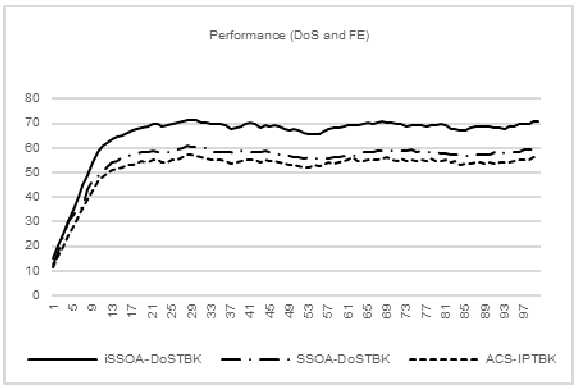
Fig. 3b. Comparison of Average False Error by iSSOA-DoSTBK, SSOA-DoSTBK, and ACS-IPTBK in returning attack paths

3) Performance

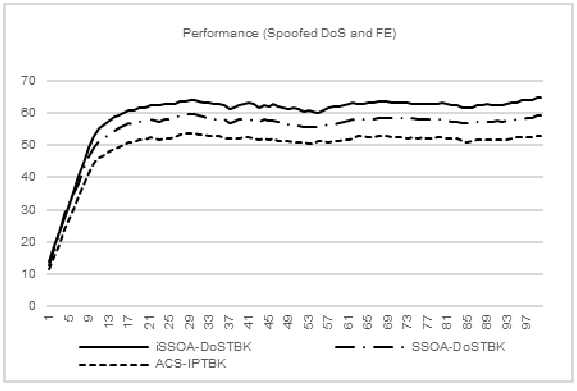
To further examine the accuracy of the attack paths returned by iSSOA-DoSTBK and the other benchmark solutions, the number of attack packets on the paths returned were compared. Fig. 4a(i) to 4a(iii) show the results of the comparison. The percentage of attack packets on the paths returned by iSSOA-DoSTBK were more than the attack packets found on the paths returned by the benchmark solutions as can be seen in the figures.



i). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK Performance for tracing DoS in the network



ii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK Performance for tracing DoS with FE in the network



iii). iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK Performance

for tracing Spoofed DoS with FE in the network

Fig. 4a. Comparison of iSSOA-DoSTBK, SSOA-DoSTBK and ACS-IPTBK Performance Data.

Fig. 4b shows the bar chart comparison of the correctness of the attack paths returned by iSSOA-DoSTBK and the other benchmark solutions in the tests conducted. The results of iSSOA-DoSTBK contained higher number of attack packets than those of SSOA-DoSTBK by 22.19%, 18.45%, and 8.14% for DoS only, DoS and FE, and Spoofed DoS and FE tests respectively. It is higher than the results of ACS-IPTBK by 30.21%, 26.09%, and 19.96% for the same tests respectively.

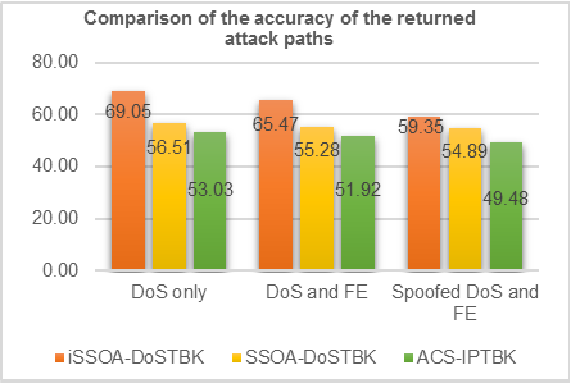


Fig. 4b. Comparison of the accuracy of the attack paths returned by iSSOA-DoSTBK, SSOA-DoSTBK, and ACS-IPTBK

XII. Conclusion

This work incorporates statistical analysis of attack packets distribution on attack path edges in iSSOA-DoSTBK to avoid the possibility of discrimination policy generating conflicting results. In SSOA-DoSTBK [12], discrimination policy was only used to compare packets on the nodes along the attack path with the known attack packets features. The edge on which the highest number of packets matching attack packets was found is then selected. This process may generate multiple edges on the same node that have equal number of attack packets, a situation that may make it difficult to select an edge that is a correct segment of the attack path. iSSOA-DoSTBK further examines the distribution of the attack packets on the nodes along the attack path to ascertain the edges that transmit attack packets more regularly during the flow time. This ensures the selection of the most probable segments of the attack path during the traceback. The results obtained from the tests conducted show improvements in terms of reduced error rates, and more accurate attack paths detection by iSSOA-DoSTBK over the SSOA-DoSTBK and ACS-IPTBK. The iSSOA-DoSTBK recorded errors lower by as much as 32.35% compared to errors returned by the SSOA-DoSTBK. The errors in its results were lower than the errors in the results of ACS-IPTBK by as much as 51.51%. But SSOA-DoSTBK returned attack paths a little faster than the iSSOA-DoSTBK by as much as 0.65%. ACS-IPTBK was faster to return attack paths than iSSOA-DoSTBK by as much 1.24%. It was noted that the time differences for returning attack paths between the iSSOA-DoSTBK and the benchmark solutions were minimum for the DoS and FE tests but highest for DoS only tests. This observation indicates that iSSOA-DoSTBK performs better than the other solutions it was compared with in terms of discerning network traffic surges caused by flash events from the DoS attacks flow. The benchmark solutions' method of attack paths detection are more of estimating most probable segments rather than a deep examination of the edges for the attack packets to get the correct segments. They were able to estimate the attack paths faster when DoS attacks were the only sources of heavy traffics. iSSOA-DoSTBK performs a more thorough examination of the nodes in all cases to ascertain their involvement in the attack packets rerouting. The thorough examination of the nodes made iSSOA-DoSTBK to be slightly slower. But it returned more accurate attack paths with up to 22.19% higher accuracy than SSOA-DoSTBK and 30.21% higher than ACS-IPTBK.

This work tests iSSOA-DoSTBK for detecting DoS attacks only. It has not been tested with distributed DoS. Also, adjusting the user defined parameters of SSOA to make the search process faster, which may reduce the accuracy of the results, has not been tested in this work. These options are still open areas for future research.

References

[1] ISO/IEC, "Information technology — Security techniques — Information security management systems — Overview and vocabulary," Geneva, Switzerland, 2009.

[2] D. S. N. Mary and A. T. Begum, "AN ALGORITHM FOR MODERATING DOS ATTACK IN WEB BASED APPLICATION," in *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*, 2017, pp. 26–31.

[3] S. Saurabh and A. S. Sairam, "Increasing Accuracy and Reliability of IP Traceback for DDoS Attack Using Completion Condition," *Int. J. Netw. Secur.*, vol. 18, no. 2, pp. 224–234, 2016.

[4] Data-Protection-Act-1998, "Data Protection Act 1998," *Data Protection Act 1998*, 1998. [Online]. Available: http://www.legislation.gov.uk/ukpga/1998/29/contents. [Accessed: 22-Sep-2018].

[5] Cybercrime Act 2015, "CYBERCRIMES (PROHIBITION, PREVENTION, ETC) ACT, 2015," *CENTRE FOR LAWS OF THE FEDERATION OF NIGERIA, 2015*, 2015. [Online]. Available: http://lawnigeria.com/LawsoftheFederation/Cyber-Crime-Act,-2015.html. [Accessed: 03-May-2018].

[6] A. Bhandari, A. L. Sangal, and K. Kumar, "Characterizing flash events and distributed denial-of-service attacks : an empirical investigation," vol. 9, no. March, pp. 2222–2239, 2016.

[7] A. Dhingra and M. Sachdeva, “Recent Flash Events : A Study,” in *International Conference on Communication, Computing & Systems (ICCCS-2014)*, 2014, pp. 94–99.

[8] R. Linnie and H. Nick, “Cable News Network,” *Cable Network News*, 26-Jun-2009.

[9] S. Chawla, M. Sachdeva, and S. Behal, “Discrimination of DDoS attacks and Flash Events using Pearson ’ s Product Moment Correlation Method,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 10, pp. 382–389, 2016.

[10] M. A. Mohamed, N. Jamil, A. F. Abidin, M. M. Din, W. W. N. S. Nik, and R. A. Mamat, “Entity-based Parameterization for Distinguishing Distributed Denial of Service from Flash Events,” *Int. J. Eng. Technol.*, vol. 7, no. 2.14, pp. 5–8, 2018.

[11] M. Sachdeva and K. Kumar, “A Traffic Cluster Entropy Based Approach to Distinguish DDoS Attacks from Flash Event Using DETER Testbed,” *ISRN Commun. Netw.*, vol. 2014, no. 259831, p. 15, 2014.

[12] O. W. Salami, I. J. Umoh, and E. A. Adedokun, “The Development of an Internet Protocol Traceback Scheme for the Detection of Denial of Service Attack,” *Niger. J. Technol. Dev.*, p. In Press.

[13] V. Murugesan, M. Shalinie, and N. Neethimani, “Brief Survey of IP Traceback Methodologies,” *Acta Polytech. Hungarica*, vol. 11, no. 9, pp. 197–216, 2014.

[14] R. Latif, H. Abbas, S. Latif, and A. Masood, “Distributed denial of service attack source detection using efficient traceback technique (ETT) in cloud-assisted healthcare environment,” *J. Med. Syst.*, vol. 40, no. 161, pp. 1–3, 2016.

[15] N. Siddique and H. Adeli, “Nature Inspired Computing: An Overview and Some Future Directions,” *Cognit. Comput.*, vol. 7, no. 6, pp. 706–714, 2015.

[16] N. Venkataramanan and T. Ravi, “An Enhanced IP Trace Back Mechanism by using Particle Swarm System,” *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 3, no. 9, pp. 48–51, 2017.

[17] A. Saini, C. Ramakrishna, and K. Sachin, “A Hybrid Approach to Address IP Traceback Problem using Nature Inspired Algorithm,” in *International Conference on Information Technology, Communications and Computing*, 2017.

[18] P. Wang, H.-T. Lin, and T.-S. Wang, “An improved ant colony system algorithm for solving the IP traceback problem,” *Inf. Sci. (Ny).*, vol. 326, pp. 172–187, 2016.

[19] O. Abedinia, N. Amjady, and A. Ghasemi, “A new metaheuristic algorithm based on shark smell optimization. Complexity,” *Complexity*, vol. 21, no. 5, pp. 97–116, 2016.

[20] M. J. Coates and R. D. Nowak, “NETWORK TOMOGRAPHY FOR INTERNAL DELAY ESTIMATION,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (Cat. No.01CH37221).*, 2001, pp. 3409–3412.

Modification of Bacterial Foraging Optimization Algorithm using Elite Opposition Strategy

Maliki Danlami
Department of Computer Engineering
Federal University of Technology
Minna, Nigeria
danlami.maliki@futminna.edu.ng

Muazu Mohammed Bashir
Department of Computer Engineering
Ahmadu Bello University Zaria
Kaduna, Nigeria
mbmuazu@abu.edu.ng

Olaniyi Olayemi Mikail
Department of Computer Engineering
Federal University of Technology
Minna, Nigeria
mikail.olaniyi@futminna.edu.ng

Jonathan Gana Kolo
Department of Electrical and Electronics Engineering
Federal University of Technology
Minna, Nigeria
jgkolo@futminna.edu.ng

Abstract— **This research work presents the modification of Bacterial Foraging Optimization Algorithm (BFOA) using the elite opposition strategy. The BFOA uses a random search strategy which affect it convergence performance due poor diversification in the search process and the possibility of Oscillatory behaviour towards the search process. The Elite Opposition BFOA is developed to provide more search space so as to enhance more exploitation. The Elite Opposition BFOA (EOBFOA) and the BFOA have been tested using twelve standard benchmark functions (Unimodal and Multimodal benchmark functions). From the simulation result obtained, the EOBFOA outperform BFOA by obtaining better global minimum solution**.

Keywords— **bacterial foraging optimization, elite opposition, benchmark test function, chemotaxis.**

I. INTRODUCTION

Optimization is the process of finding the best solution to certain problems base on either finding the maximum or minimum solution with in a certain boundary using a particular objective function. In the world of optimization, traditional optimization methods have been applied in finding the best solution around a specific domain, however the traditional methods (gradient base methods) experience difficulties in finding global optimum [11]. Technically, optimization algorithms can be classified into deterministic and stochastic optimization methods. The deterministic algorithms usually have better solution for a particular optimization problem when the same set of initial values are use at the initial stage of the algorithms. However, such method usually engaged in local search process and easily trapped in local optima. The stochastic optimization methods mostly use a random search process that can enable it escape from local optima and search for a good solution after certain number of iterations [6].

In 2002, Passion was inspired by the foraging behaviour of Escherichia Coli, and propose the Bacteria Foraging Optimization Algorithm (BFOA). The field of BFOA at present has attracted the attention of different researchers' in solving global optimization problem [4]. The BFOA based on social behaviour of the E.Coli bacterial has gain popularity and wider application in solving optimization problem ranging from robot coordination, distributed optimization and control [1]. One of the main challenges of BFOA is its poor convergence capability over multimodal and rough fitness application compared to other evolutionary algorithm such as Genetic Algorithm (GA) and Differential Evolution (DE) [12].

An Adaptive Bacterial Foraging Algorithm (ABFA) was applied in colour image enhancement using fuzzy entropy as an objective function. The ABFA technique optimized the objective function by varying the step size of the bacteria colony. The loss of unnecessary information from the image is reduce by placing constrain during the minimization of the entropy. The ABFA was also compared with the existing image enhancement technique (histogram equalization) and the ABFA outperformed the histogram equalization technique [10]. A multilevel Co-operative Bacterial Foraging Algorithm was applied in colour image segmentation that involved the combination of bacterial chemotaxis, cell-to-cell communication and adaptive scheme for the modification of the Bacterial Foraging Algorithm. A standard test image was used to evaluate the performance of the Co-operative Bacterial Foraging Algorithm with the traditional BFOA. The Co-operative Bacterial Foraging Algorithm outperformed the traditional BFOA in terms of finding a better threshold in less processing time [15].

Bacterial Foraging Optimization Algorithm was modified by varying the population of the bacteria for the purpose of image compression and applying it in fuzzy vector quantization to enable the reduction in average distortion estimation between reconstructed image and training image. The modified BFOA called the BFVPA ensure that the population size of the BFOA scale through variation in the stages of chemotaxis, swarming, elimination and communication sensing in the iteration process. BFVPA

perform better than BFOA when compare on PSNR for different images [9]. The same BFVPA was used in the optimization of palm print authentication. Both BFOA and BFVPA were used to select the combination of features that lead to the best performance in palm print base biometric identification system. The accuracy of the authentication system when using BFVPA technique is more than 97.65% when compare with BFOA with 96.65%. [8].

The performance of BFOA was improved using a social behaviour strategy by directing the movement of bacteria towards a lower value of cost function as against the usual random process of the algorithm. The new strategy enhances the lowering of number of iterations of the BFOA when applied to geometric transformation in image registration. The processing time of the algorithm with social behaviour strategy was 24% less when compared with BFOA [13]. Medical image alignment was performed between two images using BFOA. The optimization process was guided by similarity metrics as an objective function which measure the degree of resemblance between the source and target image. The author introduces h-BFOA as a modification to BFOA, due to some unnecessary chemotaxis steps that occurs as a result of oscillation in the position of the bacteria when they are close to the optimal value. The h-BFOA performs higher than the BFOA according to mean square error measure between the two algorithms [3].

The performance of bat algorithm was improved by modifying the algorithm with Elite Opposition Learning (OBL). The standard bat algorithm suffers from poor convergence and also been stuck in local minima. The modified bat base on OBL outperform the standard bat algorithm [17]. An improved BFOA algorithm based on Machine Learning frame work called the IBFO was used in the prediction of severity of somatization disorder. The BFOA was modified using the opposition base learning, the modified algorithm was better in terms of speed and accuracy of good solution [14]. Generally, the efficiency of any new or modified nature inspired optimization algorithm is usually tested using some set of standard benchmark test function which ranges from unimodal and multimodal benchmark test function [5]. The performance of one algorithm cannot be determined by the type of problem it solves most especially if the problems are specific with a different property. Therefore, to evaluate an optimization algorithm one most also identify the kind of problem it performs better [7].

II. Bacterial Foraging Optimization Algorithm (BFOA) Process

The Bacterial Foraging Optimization Algorithms is governed by four processes, which are chemotaxis, swarming, reproduction, elimination and dispersal [16].

A. Chemotaxis

Let j, k, and l denote the indices of chemotactic, reproduction and elimination dispersal events. The position of each bacterium in a group of S bacteria at jth chemotactic, kth reproduction steps and the lth elimination- dispersal event be represented in equation 1.

$$P(j,k,l) = \{\theta(j,k,l)\ i = 1,2,3 \ldots\ldots S \quad (1)$$

θ^i is the position of the ith bacterium, at jth chemotactic step, a kth reproduction step and lth elimination - dispersal event.

Let the cost of the ith bacterium at the location represented as $J(i,j,k,l)$.

The new position of the ith bacteria after a movement (tumbling) is defined as

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\Phi(j) \quad (2)$$

Where $C(i)$ indicates the number of steps taken in the random direction indicated by the tumble. If the cost $J(i,j+1,k,l)$ at location $\theta^i(j+1,k,l)$ is lower than the cost at $\theta^i(j,k,l)$, then the bacterium will move in the same direction with step size $C(i)$.

B. Swarming

In swarming the cell-to-cell communication is performed via information exchange through attractant to swarm to together or via repellent to isolate each other. The cell to cell communication signal by the ith bacterium is represented in equation (3).

$$\sum_{i=1}^{S} J_{cc}\left(\theta,\theta^i(j,k,l)\right), i = 1,2,3 \ldots\ldots\ldots S \quad (3)$$

Collectively such cell-to-cell attraction and repulsion is given by equation (4).

$$J_{cc}\left(\theta,\theta^i(j,k,l)\right) = \sum_{i=1}^{S} J_{cc}\left(\theta,\theta^i(j,k,l)\right) = a + b \quad (4)$$

$$a = \sum_{i=1}^{S}[-d_{attract}\exp(-w_{attract}\sum_{m=1}^{p}(\theta_m - \theta_m^i)^2)] \quad (5)$$

$$b = \sum_{i=1}^{S}[-h_{atrepellent}\exp(-w_{repell}\sum_{m=1}^{p}(\theta_m - \theta_m^i)^2)] \quad (6)$$

$d_{attract}$ =depth of attractant

$w_{attract}$ = width of attractant

$h_{repellent}$ =height of repellent

$w_{repellent}$ = weight of repellent

Where,

$J_{cc}(\theta,P(j,k,l))$ is the objective function value. in swarming, the individuals climb as in equation (7)

$$J(i,j,k,l) = J(i,j,k,l) + J_{cc}(\theta,P) \quad (7)$$

C. Reproduction

The reproduction of next generation bacteria cells occurs after certain number of chemotactic steps say Nc. Let the number of reproductive steps be Nre, the bacterium health is measured as the sum of its fitness value throughout its life

$$health = \sum_{j=1}^{Nc} f(i, j, k, l) \quad (8)$$

D. Elimination Dispersal

This event is also predefined as the number of (Ned) of elimination dispersal event of a bacterium. The (Ped) is the probability of elimination-dispersal event of a bacterium. The chemotaxis provides the basis for local search. In case of avoiding a halt in local optima, bacteria will tend to disperse to other location in the search space with respect to the probability (Ped).

III. MODIFICATION OF BFOA USING ELITE OPPOSITION STRATEGY

The opposition-based strategy is basically a computational intelligence strategy that considers the current individual and its opposite individual simultaneously in order to get a better approximation. The elite opposition base strategy, which is an extension of opposition base strategy is one of the effective approaches in enhancing the performance of Evolutionary Algorithms (EA). Let $P \epsilon [x, y]$ be a real number. The opposition number of $P(P^*)$ is defined in equation (9)

$$P^* = x + y - P \quad (9)$$

The main idea of elite based opposition is for any visible solution, calculate and evaluate the opposite solution at the same time, and chose the best individual. The equation for the elite base opposition is given in equation 10.

$$P1 = k(d_{ai} + d_{bi}) \quad (10)$$

$$x_{i,j}^* = P1 - x_{i,j} \quad (11)$$

Where i=1, 2, ……,N; , j= 1,2,……n and k is a random variable [2]. Where N is the population size, and $(d_{ai}\ d_{bi})$ is the dynamic bound of decision variable which can be obtained as follows:

$$d_{ai} = \min(x_{i,i}) \quad (12)$$

$$d_{bi} = \max(x_{i,i}) \quad (13)$$

The flowchart for the EOBFOA is shown in Fig. 1

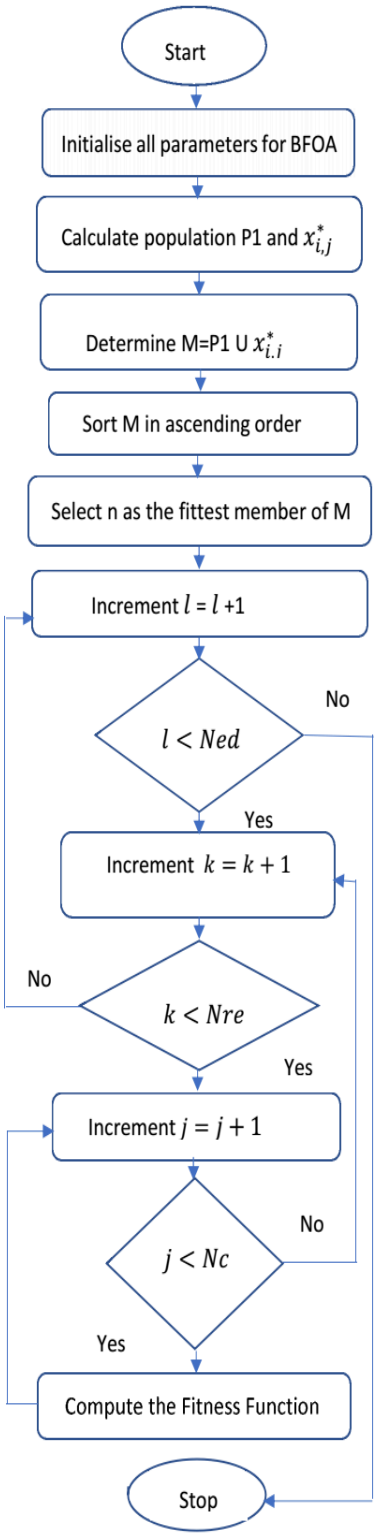


Fig. 1. Flowchart of Elite Opposition BFOA

III. PERFORMANCE EVALUATION OF BFOA AND EOBFOA

The performance of both algorithms was evaluated using twelve standard benchmark functions. The range of each test function and the Global Minimum (GM) are shown in table 1. The EOBFOA from analysis, performs better in all the test function compare to BFOA on both unimodal ($f1, f2, f3, f4, f5, f6$) and ($f7, f8, f9, f10, f11, f12$) benchmark test function.

TABLE 1: BENCHMARK FUNCTIONS FOR BFOA AND EOBFOA

Function	Equations	Range	GM
$f1$ (Griewank)	$f_1(x)=\frac{1}{4000}\sum_{i=1}^{n}(x_i^2)-\prod_{i=1}^{n}\cos(\frac{x_i}{\sqrt{i}})+1$	[-600 600]	$f(X^*)=0$
$f2$ (Sphere)	$f_2(x)=\sum_{i=1}^{n}x^2$	[-5.12 5.12]	$f(X^*)=0$
$f3$ (Zakrov)	$f_3(x)=\sum_{i=1}^{n}x_i^2+(\frac{1}{2}\sum_{i=1}^{n}ix_i)^2+(\frac{1}{2}\sum_{i=1}^{n}ix_i)^4$	[-5 10]	$f(X^*)=0$
$f4$ (Schwefel's 221)	$f_4(x)=f(x_1,\ldots,x_n)=max_{i=1,\ldots,n}\lvert x_i\rvert$	[-100 100]	$f(X^*)=0$
$f5$ (Schwefel's 222)	$f_5=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	[-100 100]	$f(X^*)=0$
$f6$ (Schwefel's 223)	$f_6(x)=f(x_1,\ldots,x_n)=\sum_{i=1}^{n}x_i^{10}$	[-10 10]	$f(X^*)=0$
$f7$ (Ackley)	$f_7(x)=-aexp\left(\sqrt{\frac{1}{d}\sum_{i=0}^{d}x_i^2}\right)-exp(\frac{1}{d}\sum_{i=1}^{d}cos(cx_i))+a+$	[-32.7 32.7]	$f(X^*)=0$
$f8$ (Drop wave)	$f_8(x)=-\frac{1+\cos(\sqrt{\sum_{i=1}^{n}x^2})}{(\frac{1}{2})(\sum_{i=1}^{n}x^2\ i)+2}$	[--5.2 5.2]	$f(X^*)=-1$
$f9$ (Goldstein and price)	$f_9=[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1$	[-2 2]	$f(X^*)=3$
$f10$ (Salomon)	$f_{10}=-\cos\left(2\pi\sqrt{\sum_{i=1}^{D}x_i^2}\right)+0.1\sqrt{\sum_{i=1}^{D}x_i^2}+1$	[-100 100]	$f(X^*)=0$
$f11$ (Rosenbrok)	$f_{11}(x)=\sum_{i=1}^{n-1}[(x_i-1)^2+100(x_i^2-x_{i+1})^2]$	[-5 10]	$f(X^*)=0$
$f12$ (Rastrigin)	$f_{11}=\sum_{i=1}^{n}[x_i^2-10\cos(2\pi x_i-x_{i+1})^2+10]$	[-5.12 5.12]	$f(X^*)=0$

A Unimodal Test Function

Most of the Unimodal Test function are continuous, convex and the function can be defined over n-dimensional space. Table 2 shows the results for unimodal test function base on statistical analysis. Fig. 2 to Fig. 6 shows the converging plot of BFOA and EOBFOA on different unimodal test function.

TABLE 2: UNIMODAL BENCHMARK FUNCTIONS

Algorithm	Metrics	$f1$	$f2$	$f3$	$f4$	$f5$	$f6$
BFOA	Best	5.7632E-05	8.8350E-05	3.0103E-04	0.0070	0.0101	2.5889E-19
	Worst	6.9968E-05	2.1180E-04	5.8613E-04	0.0298	0.0142	7.6261E-19
	Mean	6.3800E-05	1.5007E-04	4.4358E-04	0.0184	0.0122	3.8132E-19
	Std.	8.7227E-05	8.7292E-05	2.0159E-05	0.0161	0.0029	5.3923E-19
EOBFOA	Best	9.7350E-07	1.9324E-05	6.2877E-05	0.0081	0.0048	2.2573E-27
	Worst	5.2340E-05	2.2478E-04	1.7433E-04	0.0090	0.0049	1.0605E-23
	Mean	2.6657E-05	1.2205E-04	9.0307E-05	0.0086	0.0048	5.3036E-24
	Std.	3.6322E-05	1.4528E-04	1.1882E-04	6.1947E-04	7.4972E-04	7.4972E-24

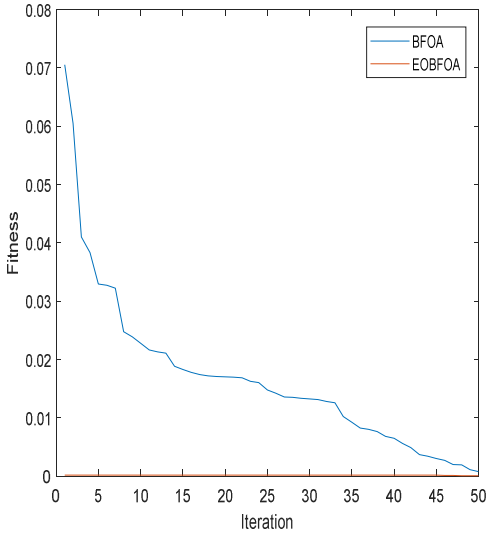


Fig. 2. Griewank BFOA and EOBFOA

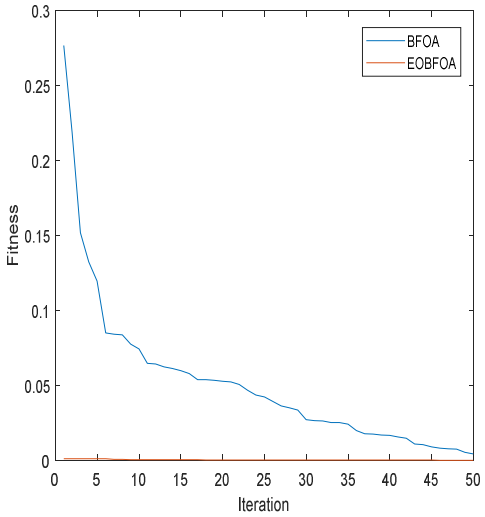


Fig. 2. Sphere BFOA and EOBFOA

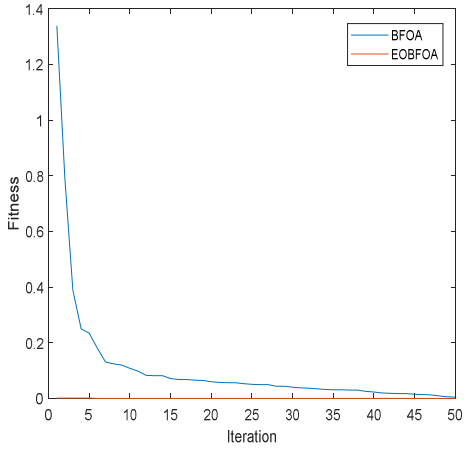


Fig. 3. Zakrov BFOA and EOBFOA

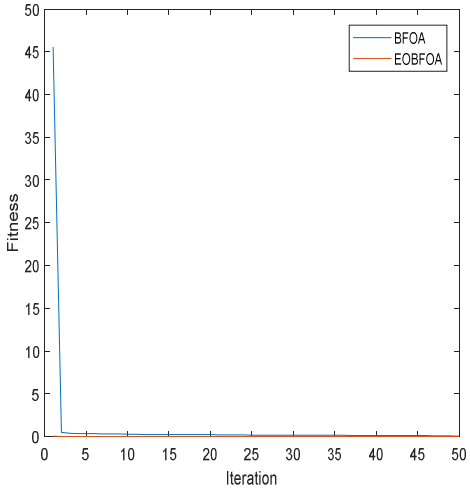


Fig. 4. Schwefel's 221 BFOA and EOBFOA

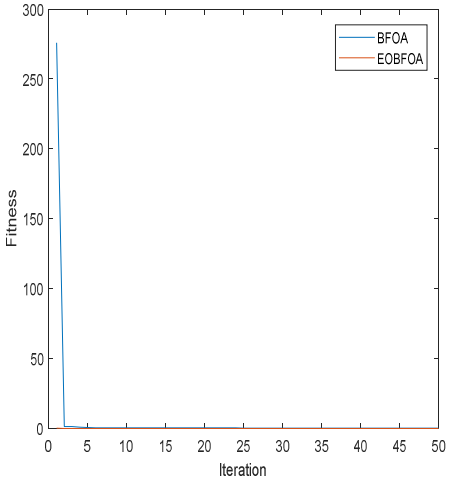


Fig. 5. Schwefel's 222 BFOA and EOBFOA

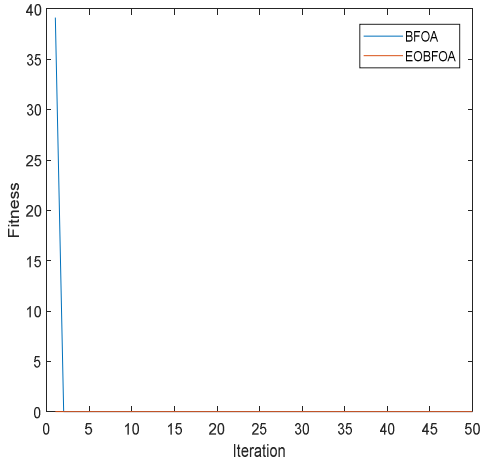


Fig. 6. Schwefel's 223 BFOA and EOBFOA

B. Multimodal Test Functions

The optimization test using multimodal benchmark functions was to check for the exploration performance of the BFOA and EOBFOA. Table 3, shows the simulation results for the using the multimodal test function and the

graphical representation of the convergence plot is also shown in Fig. 7 to Fig. 12.

TABLE 3: MULTIMODAL BENCHMARK FUNCTIONS

Algorithm	Metrics	f7	f8	f9	f10	f11	f12
BFOA	Best	0.0412	-0.9612	3.2677	8.6822E-04	0.1105	0.0872
	Worst	0.1968	-0.9638	3.2752	0.0053	0.1223	0.1093
	Mean	0.1190	-0.9747	3.2715	0.0032	0.1164	0.0983
	Std.	0.1100	0.0153	0.0053	0.0032	0.0084	0.01560
EOBFOA	Best	0.0146	-0.9923	3.0262	2.9487E-04	0.0013	0.0070
	Worst	0.0425	-0.9886	3.1375	0.0036	0.0124	0.0108
	Mean	0.0285	-0.9904	3.0819	0.0020	0.0068	0.0089
	Std.	0.0198	0.0026	0.0787	0.0024	0.0078	0.0027

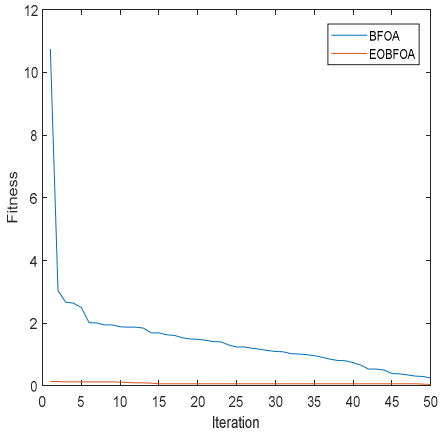


Fig. 7. Ackley BFOA and EOBFOA

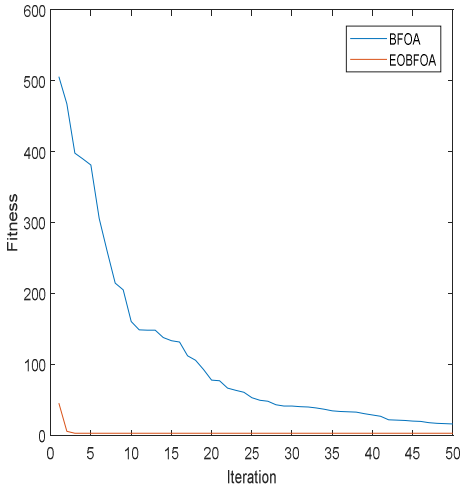


Fig. 9. Goldstein price BFOA and EOBFOA

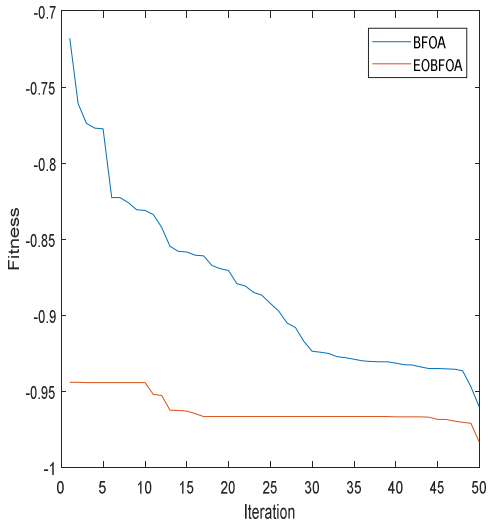


Fig. 8. Drop wave BFOA and EOBFOA

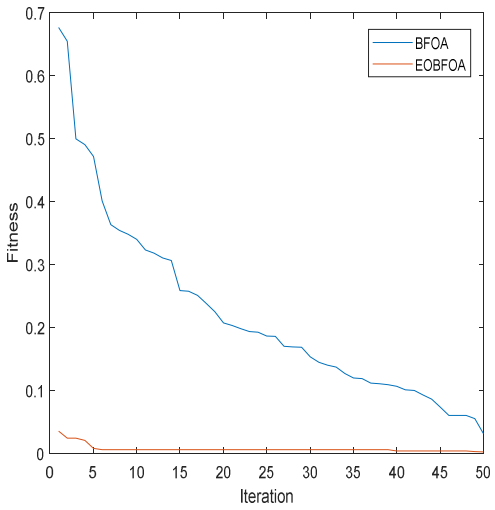


Fig. 10. Salomon BFOA and EOBFOA

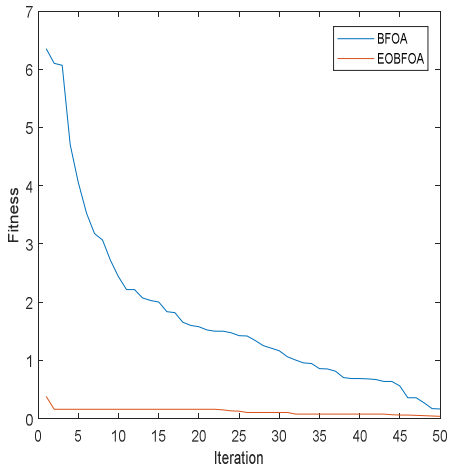


Fig. 11. Rosenbrok BFOA and EOBFOA

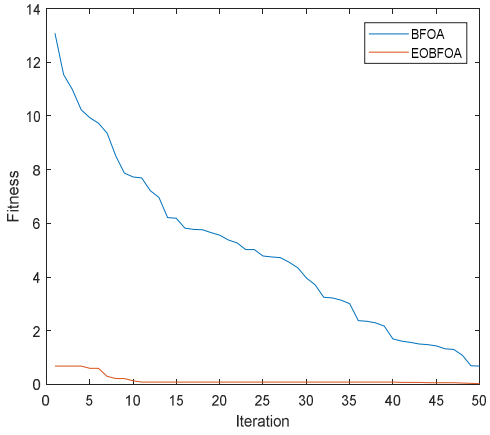


Fig. 12. Rastrigin BFOA and EOBFOA

IV. CONCLUSION

In this research, the EOBFOA is developed by modifying BFOA using the elite opposition strategy and tested on both unimodal and multimodal benchmark test function. The evaluation of both algorithms using best, worst, mean and standard deviation is performed, with the EOBFOA converging to a better solution than the BFOA. The convergence plot further shows that the elite opposition strategy provides more improvement to the standard BFOA by enhancing more diversification of the algorithm during search process. Further research will base on evaluating the performance of the EOBFOA at different parameter settings and also its application to medical images.

REFERENCES

[1] S. Amrinder, and Sonika, J. *A Systematic Way for Image Segmentation Based on Bacterial Foraging Optimization Techniques*. International Journal of Computer Science and Information Technologies. 2014.Vol. 5. Pp 130-133.

[2] Q. Chiwen, Z. Shi'an, F Yanming, and Wei, H. *Chicken Swarm Optimization Based on Elite Opposition-Based Learning*. Hindawi Journal of Mathematical Problems in Engineering. 2017. Pp 1-20.

[3] B. Enrique, V. Andrea, D. Sergio, and C. Oscar, *Bacterial Foraging Optimization for Intensity-Based Medical Image Registration*. 2015 IEEE Congress on Evolutionary Computation (CEC). 2015. Pp 2436-2443.

[4] L. Jun, D. Jianwu, B. Feng, and W. Jiansheng, *Analysis and Improvement of the Bacterial Foraging Optimization Algorithm*. Journal of Computing Science and Engineering. 2014. Vol. 8, No. 1. Pp 1-10.

[5] H. Kashif, N. Mohd., S. Mohd., C. Shi., and N. Rashid. *Common Bench Mark Functions for Metaheuristics Evaluation: A Review.* International Journal of Informatics Visualization. 2017. Pp 218-222.

[6] Z. Lina, L. Ligiang., Y. Xin-She, and D. Yuntao. *A Novel Hybrid Firefly Algorithm for Global Optimization*. https://doi.org/10.1371/journal.pone.0163230. 2016 Pp 1-17.

[7] J. Momin,. and Y. Xin-She, *A literature Survey of Bench Mark Functions for Global Optimization Problems. International.* Journal of mathematical Modelling and Numerical Optimization. 2013.Vol. 4. No. 2, Pp 150-194.

[8] S. Nandita, C. Amitava, and M. Sugata. *BFOA with Varying Population Based Feature Selection and Optimization Based Feature Selection and Optimization in Palm Print Authentication- A Comparative Study*. 2017 IEEE Calcutta Conference. Pp 236-240.

[9] S. Nandita, C. Amitava, and M. Sugata. *Fuzzy VQ Based Image Compression by Bacterial Foraging Optimization Algorithm with Varying Population*. 2014 International Conference on Control, instrumentation, energy & Communication (CIEC). Pp 641-645.

[10] P.V. Om, R. C. Rishi, and Abhinav, G. *An Adaptive Bacterial Foraging Algorithm for Color Image Enhancement*. Annual Conference on Information Science and System (CIES). 2016. Pp 1-6.

[11] H. A. Rahib, and T Mustafa.. *Experimental Study of Specific Bench Mark Function for Modified Monkey Algorithm*. 12th International Conference on Application of Fuzzy System and Computing, ICAFS 2016. Pp 595-602.

[12] A. Sasithradri,, and N.N. Singh. *Synergy of Adaptative Bacterial Foraging Algorithm and Particle Swarm Optimization Algorithm for Image Segmentation.* 2014. International Conference on Circuit, Power and Computing Technologies [ICCPCT]. Pp 1503-1506.

[13] B. Siviu-Ioan, R. Florin, L. Ramona, Cristina, D. N. and C Hariton,. *Social Behaviour in Bacterial Foraging Optimization Algorithm for Image Registration.* International Conference on System Theory Control and Computing. 2014. Pp 330-334.

[14] L.V. Xinen., C. Huiling,. Z. Qian, I. Xujie, H. Hui., and W. Gang. *An Improved Bacterial Foraging Optimization Algorithm Based Machine Learning Framework for Predicting the Severity of Somatization Disorder*. https://www.mdpi.com/1999-4893/11/2/17. 2018. Pp 1-18.

[15] L. Yang. H. Kunyuan,, Z. Yunlong, and Hanning, C. Color Image *Segmentation Using Multi Thresholding-Cooperative Bacterial Foraging Algorithm*. The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems. 2015. Pp 181-185.

[16] T.A. Yasmina,., and F. Hadria, *A Hybrid Bacterial Foraging Optimization Algorithm and a Radial Basic Function Network for Image Classification.* Journal of Information Processing Systems. 2017.Vol. 13, No. 2. Pp 215-235.

[17] H. Zaharudeen, M.B, Muazu, K.A Abibulah, and T.A Salawudeen. *Development of a Modified Bat Algorithm Using Elite Opposition-Based Learning*. 2017 IEEE 3rd International Conference on Electro-Tecnology for National Development (NIGERCON). Pp 144-151

Energy Efficient Learning Automata Based QL-RACH (EELA-RACH) Access Scheme for Cellular M2M Communications

Nasir A. Shinkafi
Communication Research Group, Department of Electrical Engineering Bayero University, Kano
P.M.B. 2011, Kano, Nigeria
nasir.shinkafi@umyu.edu.ng

Lawal Muhammad Bello
Communication Research Group, Department of Electrical Engineering Bayero University, Kano
P.M.B. 2011, Kano, Nigeria
lmbello.ele@buk.edu.ng

Dahiru Sani Shu'aibu
Communication Research Group, Department of Electrical Engineering Bayero University, Kano
P.M.B. 2011, Kano, Nigeria
dsshuaibu.ele@buk.edu.ng

Ibrahim Saidu
Department of Information and Communication Technology, Usmanu Danfodio University, Sokoto, P.M.B 2346, Sokoto, Nigeria
ibrahim.saidu@udusok.edu.ng

***Abstract*— this paper introduces an Energy Efficient Learning Automata Q-Learning Random Access Channel (EELA-RACH) Access Scheme to improve energy efficiency. The proposed EELA-RACH scheme employs a Distributed Learning Automata (DLA) technique based on Learning Automata (LA) feedback to minimize the energy consumed during updating Q-value and storing transmission history. The scheme also utilizes an adaptive duty cycle assignment to control the energy consumption of the Machine-to-Machine (M2M) devices within the cellular M2M communication cycle. The results show that the proposed EELA-RACH scheme achieves better performance compared to the Prioritized Learning Automata Q-Learning RACH (PLA-QL-RACH) and an Enhanced Learning Automata QL-RACH (ELA-QL-RACH) schemes with 9.41% and 65.72% decrease in energy consumption and increase in device lifetime, respectively.**

Keywords*— *Machine-to-Machine, Access Barring, LTE Network, RACH congestion, ALOHA protocol, Q-learning, Q-value, Quality of Service, Energy Efficiency, Lifetime, Learning Automata.

I. INTRODUCTION

Cellular-based Machine-to-Machine (M2M) communications are largely described by a massive number of concurrent active M2M devices, small payload size, and highly diverse Quality of Service (QoS) requirements [1]. Due to the concurrent activation of a large number of M2M devices to contend for Random Access (RA) resources from the cellular network through the Random Access Channel (RACH); massive M2M traffic occurs. In cellular networks such as the 4th Generation (4G) Long Term Evolution (LTE), a User Equipment (UE) communicates a uniformly selected preamble sequence on the RACH to alert the eNode B (eNB) of its connection request. If a massive number of M2M devices is simultaneously operating, preamble collisions on all available Random Access (RA) channels may occur. A device that suffers a collided preamble typically fails in its RA procedure, resulting to severe RACH collisions which block all RACH resources leading to congestion, overload and energy wastages [2, 3]. Therefore, the massive number of M2 leads to RACH collapse due to inefficient access and inherent stability issues of the existing protocol used to access the RACH [2, 4]. Furthermore, the M2M devices which are battery driven consumes energy while in active mode, and since the batteries are hardly replaced, their lifetime is generally low [5]. In line with the 5th Generation (5G) vision by Nokia [3], low energy efficiency for cellular M2M devices needs to be enhanced in order to extend battery lifetimes. Thus, prolonged battery lifetime is crucial for normal and sustained operations of the devices especially when deployed in remote or hazardous environments. Energy efficient RACH access schemes are considered as key requirements for extending the lifetime of cellular M2M devices [3, 6]. Several approaches have been proposed to provide energy efficient RACH access scheme [2, 7-12]. Some of the approaches employed clustering algorithm [2, 7] and others adopted duty cycle concepts [9, 10, 12-14]. However, the existing scheme which introduced Q-Learning RACH (QL-RACH) [11] result to low energy efficiency due to the fixed duty cycle context, and the Q-value (Q-learning value function) update and storage.

In this paper, an Energy Efficient Learning Automata QL-RACH (EELA-RACH) access scheme is proposed to enhance energy efficiency. The proposed scheme utilizes a Distributed Learning Automata (DLA) technique according to the Learning Automata (LA) feedback to minimise the energy consumed during Q-value update and transmission history storage. It also employs an adaptive duty cycle assignment technique to control the energy consumption of the M2M devices within the cellular M2M communication cycle thereby extending their lifetime. The proposed EELA-RACH scheme is evaluated with existing schemes using simulation.

The rest of the paper is structured as follows: Section 2 presents the details of the QL-RACH and the proposed EELA-RACH access schemes. Section 3 provides a performance evaluation of the proposed scheme and the paper is concluded in Section 4.

II. QL-RACH AND PROPOSED EELA-RACH ACCESS SCHEMES

As presented in [11, 13], QL-RACH employs Q-Learning to control M2M devices while coexisting with Human-to-Human Communications (H2H) devices in sharing the RACH channel. These devices were categorised into two groups, learning M2M and non-learning H2H. The learning M2M devices operate according to the QL-RACH access scheme while the non-learning H2H devices maintain the conventional Slotted Aloha RACH (SA-RACH) access scheme and coexisting together in a combined RACH access scheme. The learning was implemented by designing a virtual frame for M2M called the M2M-frame with size (in time slots) equal to the number of M2M devices on the assumption that slots are acquired when the mean RACH request rate is less than the M2M-frame time. The performance of the QL-RACH scheme is hindered by the random effect of the H2H traffic as it approaches the Slotted

Aloha (s-Aloha) capacity. As the H2H traffic load grows towards the s-Aloha capacity, the probability of collisions amongst the H2H and M2M devices rise which leads to RACH-throughput collapse due to the random effect of the H2H traffic coupled with the failure to prioritize cellular M2M traffic and the excessive reward and punishment. This results in poor access delay performance thereby forcing the M2M devices into another Q-learning cycle.

This challenge was first addressed by introducing a Frame-based Back-off QL-RACH (FB-QL-RACH) in [14] to enhance the performance of QL-RACH using a dynamic frame size adaptation method to the H2H back-off frame. The scheme reduces the probability of collision between H2H and M2M devices when sharing the same frame for both the initial access and the back-off and improves RACH throughput. However, the performance of the scheme is limited to only the M2M devices with periodic traffic and it also fails to consider the effect of the penalty factor in QL-RACH and its impact on the QoS performance of cellular M2M. Full details of the QL-RACH scheme can be found in [11, 13]. The challenges presented by [13] were first addressed by [8], where a Priority-based adaptive access barring for M2M Communications in LTE networks using Learning Automata (LA) to support different Quality of Service (QoS) priority classes of M2M devices in resource allocation procedure was proposed. The scheme dynamically allocates random access resources to different classes of M2M devices according to their priorities and demands. Furthermore, the scheme adjusts the barring factor for each class to control the possible overload. This scheme reduces access delay and resource wastage but leads to poor QoS when considering both H2H and M2M. A hybrid of Q-Learning and Learning Automata RACH Access scheme was also introduced first as a Prioritized Learning Automata QL-RACH (PLA-QL-RACH) access scheme to improve the performance of QL-RACH which was eventually optimised into an Enhanced Learning Automata QL-RACH (ELA-QL-RACH). The schemes initially employ a PLA technique and dynamic resource allocation strategy to enhance the RACH throughput performance by minimizing the level of interaction and collision of M2M with H2H devices thereby improving the QoS. Additionally, the scheme eliminates the excessive punishment suffered by the M2M devices by controlling the administration of penalty factor in QL-RACH and also introduces a Q-value update technique to reduce idle slots according to the probability of collision. Full details of the Learning Automata scheme can be found in [8]. The ELA-QL-RACH scheme, which is an enhancement of the PLA-QL-RACH, employs a dynamic resource allocation to enhance resource utilization according to the LA feedback. The scheme also introduces Q-value update integration technique to reduce the chances of producing idle slots within an M2M-Frame based on the probability of collision. In the frame, all devices maintain a Q-value per slot with each user transmitting in the slot with the highest Q-value as illustrated in Fig. 1.

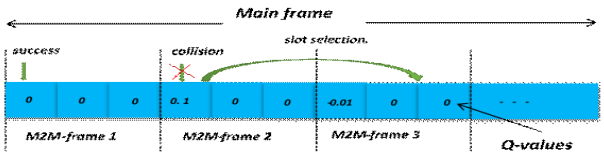


Fig. 1: QL-RACH Slot Selection adopted from [11, 13]

In Fig. 1, the Q-value is updated in a reinforcement manner using (1) after each transmission attempt and the transmission history is recorded on the individual slots [13].

$$Q' = (1-\gamma)Q + \gamma R \tag{1}$$

where Q is the current Q-value, γ is the learning rate, R is the reward (+1) or punishment (-1).

Furthermore, the transmission attempts are based on a fixed duty cycling context that predicts only wake up times of the M2M devices. The devices go through a repeated transition from sleep to wake up mode even without demanding for any resources. These behaviors consume energy and hence render the scheme energy inefficient.

To address the aforementioned problem, an EELA-RACH access scheme is proposed. Firstly, the proposed EELA-RACH introduces a DLA technique based on LA feedback to minimise the energy consumed during updating Q-value and storing transmission history [15-17]. The DLA technique utilises a network of LAs in a cooperative manner by making one automata active at a time such that when LA in state q (LA_q) selects action y_c^q, LA in state c (LA_c) is activated. The action probability vector for LAq is denoted by (2) [17].

$$P^q = \{p_1^q, p_2^q, \ldots\ldots\ldots\ldots, p_{nq}^q\} \tag{2}$$

where p_c^q is the probability related to action y_c^q, and n_q is the number of actions LA_q is able to execute during a cellular M2M communication cycle. In the cycle, the M2M device is assumed to go through five main stages and in each stage an action y is performed in the form of receiving (*rc*), transmitting (*tr*), listening (*lt*), sampling (*sm*) or sleeping (*sl*); $y \in \{rc, tr, lt, sm, sl\}$. Recall that each M2M device is equipped with an LA to form a network of LAs with each having a set of actions to choose from and at any instance. The probability that a preamble is successfully utilized by a user is derived in (3) [8].

$$p_x^{succ}(t) = \binom{N_x(t)}{1} p_x^m(t)\left(1 - p_x^m(t)\right)^{N_x(t)-1} \tag{3}$$

while the probability that this preamble suffers collision is derived in (4) [8].

$$p_x^{coll}(t) = 1 - N_x(t) p_x^m(t)\left(1 - p_x^m(t)\right)^{N_x(t)-1} - \left(1 - p_x^m(t)\right)^{N_x(t)} \tag{4}$$

and the LA feedback ($r_x(t)$) is related to $p_x^{succ}(t)$ as in (5) [8].

$$r_x(t) = \begin{cases} 0 & \text{if } p_x^{succ}(t) < S \\ 1 & \text{if } p_x^{succ}(t) \geq S \end{cases} \tag{5}$$

such that when

$$p_x^{succ}(t) = S \tag{6}$$

then

$$p_x^{coll}(t) = 1 - S \tag{7}$$

where

$$S = 2e^{-1} \tag{8}$$

The S in (8) refers to the convergence of a preamble in the successful state which is compared with the maximum throughput achieved through s-Aloha of $2e^{-1}$.

In addition, the proposed scheme utilizes an adaptive duty cycle assignment technique to control the energy consumption of the devices within the cellular M2M communication cycle as they go through the set of actions $y \in \{rc, tr, lt, sm, sl\}$. These actions when summed up together, yields the overall energy consumption of the device and hence its lifetime. The total energy consumed by the devices can be derived as follows:

First, the energy expended in receiving data is derived by (9) [15].

$$E_{rc} = t_{rx} C_{rxb} V \tag{9}$$

and t_{rx}, the time (s) to switch the receiver, is simplified in (10) [15].

$$t_{rx} \leq h\theta \times (L_{preamble} + L_{packet}) t_{rxb} \tag{10}$$

where:

- h is the neigbourhood size of the devices which is unity for cellular M2M with converge cast traffic;
- θ is the sample rate (packets/s);
- $L_{preamble}$ is the preamble length (bytes);
- L_{packet} is the packet length (bytes);
- t_{rxb} is the time (s) to receive one byte;
- C_{rxb} is the current required to receive one byte of data; and
- V is the supply voltage.

Second, the transmit energy (E_{tr}) expended is also related to the preamble length and the packet length as presented in (11) [15].

$$E_{tr} = t_{tx} C_{txb} V \tag{11}$$

where C_{txb} is the current required to transmit one byte, while t_{tx}, is the transmitter switching time in (s) which is simplified in (12) [15];

$$t_{tx} = \theta \times (L_{preamble} + L_{packet}) t_{txb} \tag{12}$$

and t_{txb} is the time (s) to transmit one byte.

In order to reliably receive packets, the Low Power Listening (LPL) check interval, t_i, must be less than the time of the preamble, thus (13) holds [15];

$$L_{preamble} \geq \left[\frac{t_i}{t_{txb}}\right] \tag{13}$$

The power consumption of a single LPL radio sample is considered as $17.3 \mu J$ [18]. Hence, the total energy spent listening to the channel is the energy of a single channel sample times the channel sampling frequency. Therefore, the sampling energy is assumed to be $E_{sm} = 17.3\ \mu J$. And the time spent listening is expressed in (14) [15].

$$t_{lt} = (t_{rinit} + t_{ron} + t_{\frac{tx}{rx}} + t_{sr}) \times \frac{1}{t_i} \tag{14}$$

where:

- t_{rinit} is the initialized radio time;
- t_{ron} is the turn in radio time;
- $t_{rx/tx}$ is switch to rx/tx time, and
- t_{sr} is the time to sample radio.

Third, the energy spent listening for messages on the radio channel is expressed in (15) [15].

$$E_{lt} \leq E_{sm} \times \frac{1}{t_i} \tag{15}$$

Furthermore, the energy associated with sampling data is expressed in (16) [15].

$$E_{sm} = t_d C_{data} V \tag{16}$$

where C_{data} is the current of sample devices (in mA) and the sampling data is expressed in (17) [15];

$$t_d = t_{data} \times \theta \tag{17}$$

with t_{data} being the sample devices. For simplicity, we evaulate the M2M device sleep time (t_{sl}) by (18) [15].

$$t_{sl} = 1 - t_{rx} - t_{tx} - t_d - t_{lt} \tag{18}$$

Fourth, the energy expended while in sleep mode is expressed by (19) [15].

$$E_{sl} = t_{sl} C_{sl} V \tag{19}$$

where C_d is the current consumed while in sleep mode.

Finally, from (9) to (19), the total energy consumption of an M2M device after passing through atleast one of the stages can be determined and expressed in (20).

$$E = E_{rc} + E_{tr} + E_{lt} + E_{sm} + E_{sl} \tag{20}$$

The lifetime of the device (L) which depends on the capacity of the battery (C_{batt}) and the total energy consumed by the battery (E) is presented by (21) [15].

$$L = \frac{C_{batt} \times V}{E} \tag{21}$$

However, in terms of duty cycle cycle (d) and the transmission energy (E_{tx}), the lifetime of the device is derived in (22) [15].

$$L = \frac{(C_{batt} \times 3600)}{E_{tx} \times d} \tag{22}$$

As the duty cycle of the devices increases, the lifetime of their battery decreases. The final duty cycle employed by the scheme is adopted from [19], which is expressed in terms of listen/sleep periods as presented in (23).

$$d = \frac{L_p}{L_p + S_p} \tag{23}$$

where L_p is the listen period and S_p is the sleep period. The listen/sleep approach will facilitate the adaptive duty cycle assignment according to the LA feedback expressed in (5) inorder to achieve the adjustment of the S_p and L_p in the following algorithm.

Algorithm 1: Proposed EELA-RACH Algorithm

```
if p_x^succ(t) ≥ S
        r_x(t) ← 1
                y_sl^tr ← LA_tr
                LA_tr ← LA_sl
                LA_sl ← (S_p ↑, Δd, L ↑)
else if p_x^succ(t) < S
        if p_x^coll(t) ≤ 1
          r_x(t) ← 0
                y_lt^rc ← LA_rc
                LA_rc ← LA_lt
                LA_lt ← (ΔL_p, Δd)
```

```
    end
  end
end
```

The technique operates according to the algorithm 1 without the conventional Q-value updates or transmission history storage by rewarding or penalizing its chosen action according to the LA feedback expressed in (5). Accordingly, the LA of each device chooses one of its actions randomly based on probability of actions outlined in (2) and its duty cycle is determined for the next stage in line with (23).

III. PERFORMANCE EVALUATION

This section presents performance evaluation of the proposed EELA-RACH scheme and other two RACH Access schemes: PLA-QL-RACH, and ELA-QL-RACH. The schemes are evaluated in terms of energy consumption and device lifetime by means of Matlab simulation as described in the following sections.

A. Simulation Scenario

The simulation assumes an ON/OFF traffic arrival rate such that when the device is active (ON), the distribution of the traffic generation follows a Poisson distribution and it is OFF when inactive. We also assume that one RA slot occurs per cycle having fifty preambles and the M2M devices contend within a cycle based on a stochastic arrival process of RA requests. Furthermore, 5000 M2M devices are assumed to be spread within an eNB coverage area in a single cell of LTE network each having applications with different QoS requirements. The simulation metrics used are energy consumption and device lifetime.

B. Simulation Parameters

Table 1 present details of the parameters used in this simulation based on LTE standard, IEEE 802.15.4 standard and from the specifications of the CC2520 radio transceiver typically used in M2M networks.

TABLE VII. SIMULATION PARAMETERS

Parameter	Value
PRACH configuration index	12
RA-slot period	1ms, 1 cycle
1 RA-slot	50 preambles
Preamble format duration	1ms
Back-off period / AC-Barring Time	28ms
Number of allowed retransmit	7
Learning Rate	0.01
Data rate	250kbps
Transmit power	100.8mW
Receive power	66.9mW
Idle listen power	66.9mW
Sleep power	90nW
Standby power	525µW
No of M2M devices	5000

C. Simulation Results and Discussion

In this section, we evaluate the performance of the proposed scheme and compare it against the previous schemes in terms of energy consumption and device lifetime. Fig. 2 shows the energy consumption against duty cycle for the three schemes (PLA-QL-RACH, ELA-QL-RACH and the Proposed EELA-RACH). The Fig. 2 illustrates that the energy consumption decreases linearly with lower duty cycle until it reaches a turning point. Beyond the turning point, the energy consumption increases linearly with the rise in duty cycle due to the increase in the L_n. This behavior is confirmed by Fig. 3, which shows the relationship between energy consumption and the number of M2M devices where the energy consumption grows linearly with the increase in the number of devices. All the three schemes exhibited similar behavior; however, the proposed EELA-RACH scheme reaches its duty cycle turning point of 0.015% with lower energy consumption compared to the existing schemes.

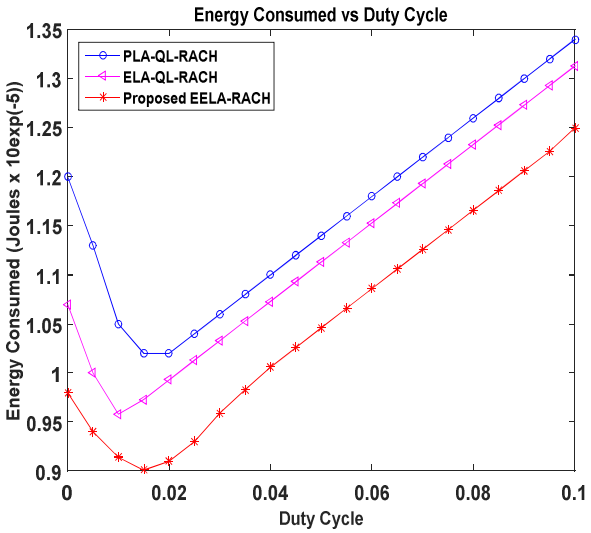
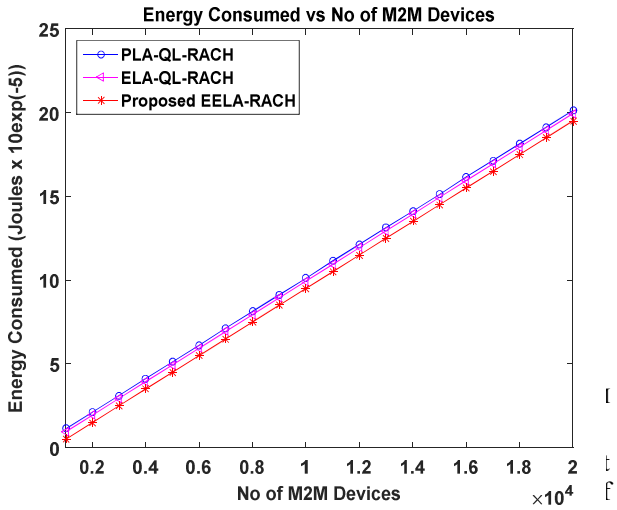


Fig. 2: Energy consumption against duty cycle - comparison of the three schemes.



the device is prolonged. When the duty cycle values are above the turning point, the energy consumption of the proposed EELA-RACH scheme rises as the other compared schemes but with lower values. The proposed scheme exhibited this behavior to alleviate traffic congestion by increasing the listening period. Consequently, the proposed EELA-RACH is shown to have 9.41% and 6.11% lower energy consumption than PLA-QL-RACH and ELA-QL-RACH schemes respectively.

Furthermore, Fig. 4 presents the lifetime against duty cycle comparison of the three schemes.

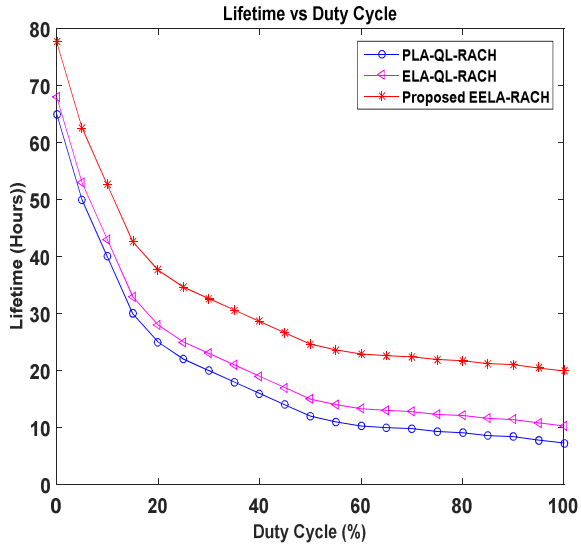


Fig. 4: Lifetime against duty cycle - comparison of the three schemes.

Fig. 4 shows that the proposed EELA-RACH has 65.72% and 43.28% higher average lifetime than PLA-QL-RACH and ELA-QL-RACH, respectively. The higher lifetime indicates lower energy consumption of the devices. The improvement is due to the impact of DLA technique with adaptive duty cycle assignment which results to reduction in energy consumed during Q-value update and storage.

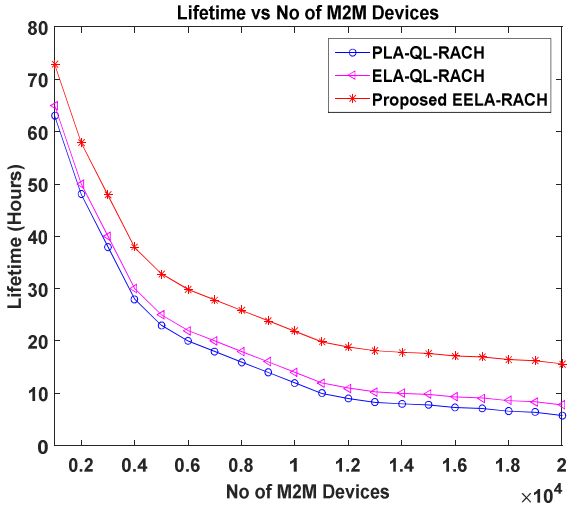


Fig. 5: Relationship between lifetime and the number of M2M devices.

However, Fig. 5, which shows the relationship between lifetime and the number of M2M devices, indicates that the lifetime of the devices per scheme decreases with the rise in their number.

IV. CONCLUSION

In this paper, an EELA-RACH scheme is proposed to improve energy efficiency. The scheme employs a DLA technique based on LA feedback to minimise the energy consumed during updating Q-value and storing transmission history. It also utilizes an adaptive duty cycle assignment technique to control the energy consumption of the M2M devices within the cellular M2M communication cycle as they go through set of actions. Simulation is used to evaluate the performance of the proposed EELA-RACH scheme against the compared schemes. The results show that the proposed EELA-RACH scheme outperforms the existing schemes in terms of energy consumption and device lifetime. The proposed EELA-RACH scheme achieves 9.41% and 6.11% lower energy consumption than the PLA-QL-RACH and ELA-QL-RACH schemes, respectively and consequently, has 65.72% and 43.28% higher average lifetime than the other schemes.

REFERENCES

[28] 3GPP, Service Requirements for Machine-Type Communications. TS 22.368 V13.1.0, 2014.

[29] 3GPP, Study on RAN improvements for machine-type communications.3GPP TR 37.868 V11.0.0, 2011.

[30] Nokia Networks., "Looking Ahead to 5G: Building a Virtual zero Latency Gigabit Experience. Nokia Netw.", Espoo, Finland, Tech. Rep. Retrieved from http://www.5gamericas.org/les/3614/3898/6583/Nokia_White_Paper_-_Looking_ahead_to_5G.pdf, 2014

[31] 3GPP, Medium access control (MAC) protocol specification. 3GPP TS 36.321 V11.0.0, 2012.

[32] Neeraj, K., Sudhanshu, T., Der-Jiunn, D., "LAEEHSC: Learning Automata based Energy Efficient Heterogeneous Selective Clustering for Wireless Sensor Networks", Journal of Network and Computer Applications, http://dx.doi.org/10.1016/j.jnca.2014.07.015, 2014

[33] Laya, A.,Alonso, L., and Alonso-Zarate, J., "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives", IEEE Commun.Surv.Tutor.16.pp 4–16, 2014.

[34] Hussain, F., Anpalagan, A., Khwaja, A. S., and Naeem, M., "Resource allocation and congestion control in clustered M2M communication using Q-learning", Trans. Emerging Tel. Tech., doi: 10.1002/ett.3039, 2016

[35] Morvari, F., and Ghasemi, A., "Priority-based Adaptive Access Barring for M2M Communications in LTE Networks using Learning Automata" Int J Commun Syst. 2017; e3325. https://doi.org/10.1002/dac.3325,

[36] Alavikia, Z., and Ghasemi, A., "A Multiple Power Level Random Access Method for M2M Communications in LTE-A Network", Trans. Emerg. Telecommun. Technol. vol. 28, no. 6, pp.e3137, 2017.

[37] Kim, K.,Jaeho, L.,Lee, J., "Energy Efficient and Reliable ARQ Scheme (E2R-ACK) for Mission Critical M2M/IoT Services", Wireless Pers Commun.pp.78:1917–1933, 2014.

[38] Bello, L. M., Mitchell, P. and Grace, D, "Intelligent RACH access techniques to support M2M traffic in cellular networks", IEEE Transactions on Vehicular Technology (Early Access). ISSN 0018-9545, 2018

[39] Zhou, Z. et al., "Energy-efficient Game-theoretical Random Access for M2M Communications in overlapped Cellular Networks. Comput. Netw"Retrieved from http://www.sciencedirect.com/science/article/pii/S1389128617302633 , 2018

[40] Bello, L. M., Mitchell, P., and Grace, D., "Application of Q-Learning for RACH Access to Support M2M Traffic over a Cellular Network", European Wireless Conference Proceedings, pp1-6, 2014.

[41] Bello, L. M., Mitchell, P., and Grace, D., "Frame based back-off for Q-learning RACH access in LTE networks", Telecommunication Networks and Applications Conference (ATNAC), Australasian, pp176-181, 2014.

[42] Saraswat, J., and Bhattacharya, P. P., "Effect of Duty Cycle on Energy Consumption in Wireless Sensor Networks", International Journal of Computer Networks & Communications (IJCNC), Vol.5, No.1, pp125-140, 2013.

[43] Barati, M.,and Sajadi, J., "A Learning Automata Based and Multicast Routing Energy Efficiency Algorithm Using Minimum Spanning Tree for Wireless Sensor Networks", Australian Journal of Basic and Applied Sciences. pp. 239-248, 2011.

[44] Beigy, H., and Meybodi, M. R., "Utilizing Distributed Learning Automata to Solve Stochastic Shortest Path Problems", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 14, pp591-615, 2006.

[45] Zhang, Y., Feng, C.-H., Demirkol, I., and Heinzelman, W. B, "Energy-Efficient Duty Cycle Assignment for Receiver-Based Convergecast in Wireless Sensor Networks", IEEE GLOBECOM. pp1–5, 2010.

[46] Moghiss, V., Meybodil, M. R., and Esnaashari, M., "An Intelligent Protocol to Channel Assignment in Wireless Sensor Networks: Learning Automata Approach", International Conference on Information, Networking and Automation (ICINA). pp. V1-338-V1-343, 2010.

hauWE: Hausa Words Embedding for Natural Language Processing

Idris Abdulmumin
Department of Computer Science
Ahmadu Bello University, Zaria

Bashir Shehu Galadanci
Department of Software Engineering
Bayero University, Kano

***Abstract* – Words embedding (distributed word vector representations) have become an essential component of many natural language processing (NLP) tasks such as machine translation, sentiment analysis, word analogy, named entity recognition and word similarity. Despite this, the only work that provides word vectors for Hausa language is that of [1] trained using fastText, consisting of only a few words vectors. This work presents words embedding models using Word2Vec's Continuous Bag of Words (CBoW) and Skip Gram (SG) models. The models, hauWE (Hausa Words Embedding), are bigger and better than the only previous model, making them more useful in NLP tasks. To compare the models, they were used to predict the 10 most similar words to 30 randomly selected Hausa words. hauWE CBoW's 88.7% and hauWE SG's 79.3% prediction accuracy greatly outperformed [1]'s 22.3%.**

Keywords—Hausa, words embedding, natural language processing.

I. INTRODUCTION

Words embedding (distributed word vector representations) have become an essential component of many natural language processing (NLP) tasks such as machine translation [2, 3], sentiment analysis [4] and text classification [5].
To improve the accuracy of low resource machine translation, [6] used words vectors to improve the alignments between words. The authors argued that words of similar meanings should have similar translations and that not all words that are substitutable are synonyms.

Natural Language Processing has in recent years receive a lot of attention from researchers. Despite this, only a few works considered low resource Hausa language especially in the field of word vector representations. Most of the attention is on high resource languages such as English, French, Arabic, Chinese, etc. To the best of our knowledge, the only publicly available dataset for machine translation is the Tanzil dataset, consisting of a meagre 127k parallel sentences. The dataset is also made of less than 10k unique Hausa sentences. This dataset also does not represent the various dialects of Hausa language, which according to [7], include Eastern Hausa (e.g. Kano), Western Hausa (e.g. Sokoto) and dialects from Niger (e.g. Aderanci).

The only work that provides word vectors for Hausa language is that of [1]. The work provides trained models for 294 languages including Hausa. The vectors were trained using fastText [8] and the Hausa vectors model consists of a vocabulary of 4347 words only with more than 1500 English words, representing about 40%.

Learning word representations require a lot of data. The performance of the CBoW and SG varies with the amount of data available. Mikolov et al. [9] have found that with more dataset, CBoW model outperforms SG model. While both models perform well with abundant quantities of training data, the SG model provides good representation for rare words in low resource settings.

In this paper, the work presented describes the processes involved in generating hauWE, an open-source distributed words vector representation (words embedding) consisting of a larger vocabulary to provide the research community with a better model for Hausa NLP tasks. The model is built based on text resources crawled from Hausa news sites on the World Wide Web and is trained based on the gensim's Word2Vec implementation of the CBoW and SG models. The model is created using 692k text documents (sentences) consisting of 18.1 million words – 77 times bigger than the old model – and a vocabulary size of 90k words – 20 times bigger than the old model – among which are about 8k English words – mostly names of persons, cities, etc. – representing just 9% of the entire corpus.

	Bojanowski	hauWE
Total no. of words	234, 779	18, 182, 511
No. of English words	50, 255 (21%)	1, 278, 069 (7%)
Vocabulary size	4, 347	90, 451
English words	1, 661 (38%)	8, 584 (9%)

Table 1. Words distribution

Data generation, data cleaning and preprocessing and choice of hyper-parameters are further explained in the paper. The models are compared with that of [1] –

labelled Bojanowski. They were used to predict 10 most similar words to each word in a given set of 30 common words and people names. The two newly created models greatly outperformed the previous model.

Words Embedding

Words embedding is a way of representing words as vectors in $\mathbb{R}^n$ where words that share a degree of semantic and syntactic similarity are represented by similar (or closer) vectors. A word can be represented in hundreds of dimensions. The idea of words embedding is building a neural language model [9] that can predict the next word given a set of context vectors [10].

The widely used methods to learn word vector models are [9]'s Continuous Bag of Words (CBoW) and Continuous Skip-gram (SG) models. The CBoW model predicts the most probable word given a context vector while the SG model, on the other hand, determines the set of context words given an input word.

The CBoW and SG models consider words as basic units ignoring rich subword information, thereby, significantly limiting the performance of the models [11]. This shortcoming was addressed in [1] by extending the SG model using subword information. Each word is represented as a bag of character n-grams and the word vector is derived as the summation of the n-gram vectors.

Global Vectors, GloVe [12], is another model for word representation that was proposed to address the drawbacks suffered by CBoW and SG models. The models train on separate local context windows instead of global word-word co-occurrence counts, thereby, capturing and making efficient use of global statistics. It was shown to outperform word2vec [9] on word analogy, named-entity-recognition and word similarity tasks.

Related work. Many works have been carried out on generating words embedding for a number of languages. Example of such works includes that of [13] that provides Arabic words embedding for NLP. [14] Provides word vectors for 100 languages to be used for multilingual NLP. [15] Also provided words embedding for 157 rich and under-resourced languages. None of these works includes Hausa language. The only publicly available work on word vectors representation that includes Hausa language among other 293 languages is that of [1]. The Hausa words embedding consists of about 4k words in Hausa, English and other languages.

The rest of the paper is organized as follows: in section II we describe the methodology which includes data generation, data cleaning and other preprocessing processes, and model generation. In section III, the models are evaluated qualitatively and finally, in section IV, we conclude the paper and discuss future work.

II. METHODOLOGY

A. Data Generation and Preprocessing

The models are trained using datasets composing of the Tanzil dataset and crawled news articles from the web.

Tanzil: The Tanzil dataset is a translation of the Quran in various languages including Hausa. The monolingual Hausa dataset consists of 127k sentences. The model will be trained on the Hausa monolingual dataset in the Tanzil corpus.

Web Crawl: news data collected from Hausa news sites in the world wide web using crawlers created for each news site to maximize the amount of data collected. The crawled news data consists of data in different domains such as sports, literature, finance, education, culture, security, etc.

The statistics of all data used in training the model is showed in Table 1.

The following are the data cleaning and preprocessing processes carried out on the raw data before training

1. Removing punctuations
2. Removing sentences made only of characters, not words
3. Removing numbers and non-*Boko* scripts
4. Minimum words count in a sentence is set to 3
5. Tokenization was done using a python script created for this task
6. Duplicate sentences were removed.

To determine the number of English words in the entire corpus both in the old and new models, each word in the models' vocabularies was checked against the Wordnet dataset. The Wordnet dataset is an English dictionary contained in the Natural Language Toolkit (NLTK) [16]. Some of the words appear both as Hausa

and English words. We, therefore, removed some of them which we were sure they are used in the documents as Hausa words. These words are: *a, da, ta, ya, na, ba, yi, su, yi, ne, ce, shi, ga, za, sai, yan, aka, wa, kan, nan, ko, ka, hau, mu, masu, kasa, kai, dan, ake, sa, amma, yana, yin, tare, bai, ita, ni, baya, ana, masa, din, tun, mun, kafa, dama, akan, ji, zaman, fi, tana, zo, abu, kama, mana, sha, kula, zan, jin, kayan, boko, ki, dole, babu, dace, gare, dauke, damar, kansa, kashi, rana, dari*.

Determining the number of English words is to ascertain the usability or otherwise of the model as Hausa words embedding. An embedding that contains the majority of its words vectors as English words would not be appropriate to be used for Hausa NLP tasks.

B. The Models

The models were trained using gensim, a tool created by [14] for various NLP tasks such as topic modelling and word vectors generation.

Two models were created: hauWE CBOW and hauWE SG based on the implementation by [1].

The models were trained for 5 epochs setting the minimum word count to 1 and model size as 300. All other parameters were left as default as set in gensim's Word2Vec model implementation. This means an initial learning rate of 0.025 which will, as the training progresses, drop to 0.0001. The default window size is set to 5. The training is done using negative sampling with 5 negatives or "noise words" to be drawn.

The trained models are available for research at https://abumafrim.github.io/main/2019/06/14/hauWE.html.

III. EVALUATIONS

For evaluating the models, we use the model to generate similar Hausa words. The two models perform better than Bojanowski.

The Hausa words used are: *miji, mata, makaranta, gida, tafiya, kyau, ido, waya, kira, hadisi, kara, godiya, kuka, ibrahim, so, kallo, unguwa, dariya, kaga, sai, kuma, rawa, kida, waka, habiba, zainab, kalma, musa, abdullahi* and *littafi*. These include common day-to-day spoken words and names of individuals.

words	similar words predicted correctly		
	Bojanowski	hauWE CBOW	hauWE SG
miji	0	10	10
mata	1	10	9
makaranta	8	10	10
gida	3	7	6
tafiya	2	8	9
kyau	1	9	8
ido	0	8	10
waya	0	10	10
kira	5	10	9
hadisi	3	10	10
kara	2	8	7
godiya	0	10	10
kuka	1	9	4
ibrahim	5	10	10
so	0	9	5
kallo	1	7	7
unguwa	4	10	8
dariya	0	9	8
kaga	1	7	6
sai	0	10	0
kuma	0	8	0
rawa	1	3	4
kida	0	10	10
waka	0	10	10
habiba	1	10	10
zainab	2	10	10
kalma	0	4	7
musa	8	10	10
abdullahi	9	10	10
littafi	9	10	9
	22.3%	**88.7%**	79.3%

Table 2 Similar words predicted correctly by the models given an input word. The words predicted are shown in Appendix A

The model was used to predict 10 most similar words, given an input word. 30 common words and names were used as input and the statistic of the prediction is showed in Table 2. hauWE CBOW produced the best result, predicting 88.7% words accurately whereas Bojanowski performed the worst with just 22.3% accuracy. The huge accuracy margin can be attributed to the difference in the size of vocabulary and the amount of training data.

Both models – hauWE CBOW and hauWE SG – were accurate in predicting only female named entities when a female name was given as input and also, predicted male named entities when a similar word is given as input.

The accuracy of the two models over Bojanowski can also be attributed to the ratio of English to Hausa words in the training corpus. The hauWE models and Bojanowski contain 1:14 and 1:5 English to Hausa words respectively. 38% of the vocabulary in the Bojanowski model are English words which translate to 21% of the entire corpus. The huge amount of English words in Bojanowski influenced the accuracy of the model to learn good vectors for Hausa words. This can be seen in Table 3 where the similar words predicted by Bojanowski are mostly not Hausa words.

IV. CONCLUSION

In this work, we created word vector models using the CBoW and SG for Hausa NLP research. The set of models, hauWE, consists of a considerably larger amount of words in the vocabulary and has been shown to perform considerably better than the previous publicly available model. hauWE CBoW, in particular, outperforms the two models. This confirms the hypothesis by [8] that for a large dataset, CBoW model outperforms SG model. In future work, we aim to create comprehensive test sets, especially analogy datasets, to be able to measure the accuracy of Hausa words embedding. More text resources will also be generated to improve the performance of the models. Lastly, some tasks in NLP require in-domain text resources. In the future, in-domain data will be used to create words embedding for domain-specific tasks such as in medicine.

References

[1] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *CoRR*, vol. abs/1607.0, 2016.

[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv Prepr. arXiv1409.0473*, 2014.

[3] M. Artetxe and H. Schwenk, "Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond," *arXiv:1812.10464v1*, 2018.

[4] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, 2017.

[5] P. Wang, B. Xu, J. Xu, G. Tian, C. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, pp. 806–814, 2016.

[6] N. Pourdamghani, M. Ghazvininejad, and K. Knight, "Using Word Vectors to Improve Word Alignments for Low Resource Machine Translation," in *Proceedings of NAACL-HLT*, 2018, pp. 524–528.

[7] P. J. Jaggar, "Hausa," *Elsevier Ltd*, pp. 222–225, 2006.

[8] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "FastText.zip: Compressing text classification models," *arXiv Prepr. arXiv1612.03651*, 2016.

[9] T. Mikolov, G. Corrado, K. Chen, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv:1301.3781v3*, pp. 1–12, 2013.

[10] G. Berardi, A. Esuli, and D. Marcheggiani, "Word Embeddings Go to Italy : a Comparison of Models and Training Datasets," in *Proceedings of the Italian Information Retrieval Workshop*, 2015.

[11] W. Chen and W. Sheng, "A Hybrid Learning Scheme for Chinese Word Embedding," in *Proceedings of the 3rd Workshop on Representation Learning for NLP*, 2018, pp. 84–90.

[12] J. Pennington, R. Socher, and C. D. Manning, "GloVe : Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Emperical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[13] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," in *3rd International Conference on Arabic Computational Linguistics*, 2017, vol. 117, pp. 256–265.

[14] R. Al-Rfou, B. Perozzi, and S. Skiena, "Polyglot: Distributed Word Representations for Multilingual NLP," *arXiv:1307.1662v2*, 2014.

[15] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning Word Vectors for 157 Languages," *arXiv:1802.06893v2*, pp. 1–5, 2018.

[16] S. Bird, L. Edward, and K. Ewan, *Natural Language Processing with Python*. O'Reilly Media Inc., 2009.

APPENDIX A

words	models	similar words
miji	hauWE CBOW	*namiji, mijinta, mijin, kishiya, maigida, matarsa, mahaifi, aure, mahaifiya, macen*
	hauWE SG	*namiji, mijinta, mijin, mijinsu, mijinki, auro, manemin, mace, saketa*
	Bojanowski	*daidaito, yunwa, daidai, zagaye, firinji, gwargwadon, yiwa, alheri, kura, dawowa*
mata	hauWE CBOW	*matan, matansu, marayu, mace, zawarawa, yayansu, mãtã, yara, mazaje, iyaye*
	hauWE SG	*matan, maza, damata, yaranta, yanmata, iyaliina, yara, samari, mace, yammata*
	Bojanowski	*kamata, fata, matar, huta, zata, sata, cutarwa, nata, fita, ambata*
makaranta	hauWE CBOW	*jamia, makarantar, makarantarmu, makarantu, firamare, makarantan, makarantun, makaratar, kwaleji, sakandare*
	hauWE SG	*makarantar, makarantan, makarata, makarantarmu, jamia, islamiya, islamiyyar, makaratar, islamiyya, makarantarsu*
	Bojanowski	*makarantun, karanta, makarantu, makarantar, karantarwa, karantar, karatunsa, karatun, makkah, burujerdi*
gida	hauWE CBOW	*daki, gidanta, gidansu, hayyacinsu, hayyacinta, gidana, shago, hayyacinsa, gidajensu, gidanmu*
	hauWE SG	*hayyacinsa, falo, daki, hayyacin, siton, hayyacina, waje, gidansu, kayayyakina, dakina*
	Bojanowski	*gidaje, gudu, biyu, gidajen, hudu, uku, biyo, watsi, gidan, gishiri*
tafiya	hauWE CBOW	*juyawa, tafiyar, wucewa, zagayawa, yawo, hutawa, taruwa, lalacewa, dawowa, canzawa*
	hauWE SG	*tafiyar, tafiyarsu, tafi, tafiyarsa, dakatawa, jiransa, sassarfa, firgice, karasawa, balaguro*
	Bojanowski	*tafiyarki, mafiya, lafiya, miliyoyin, gaskiya, kimiya, tasowa, talauci, tsinci, burbushin*
kyau	hauWE CBOW	*kyawu, inganci, tsabta, nagarta, maana, muhimmanci, kayatarwa, sauki, tsafta, mahimmanci,*
	hauWE SG	*kyawu, inganci, mahimmaci, bullewa, gautsi, araha, kayatarwa, kayu, muhimmanci, rudarwa,*
	Bojanowski	*kyauta, gungun, magunguna, abinda, fahimta, kyautata, lalata, koyi, tabbata, taimako*
ido	hauWE CBOW	*idanu, idon, klwallaye, idanunsu, idonta, gugansa, hijabi, saido, hanu, idonsu,*
	hauWE SG	*idanu, idon, idonta, idanuwansu, idanunsu, saido, idonsu, idanun, rintse, hazohazo,*
	Bojanowski	*canji, fahimta, canje, rayuwarsu, koyi, gungun, taimako, magunguna, zuba, tsoro*
waya	hauWE CBOW	*wayarsa, lambarsa, tes, wayata, wayarta, wayan, wayarka, wayoyinsu, imel,* blackberry,
	hauWE SG	*wayarsa, lambarsa, salula, wayarta, tes, tarho, wayata, gsm, salular, bugo,*
	Bojanowski	*mota,* nobel, *wasanni, kwayoyin, dinga, fasahar, wasan, wasa, shayar, samuwar*
kira	hauWE CBOW	*kiran, kirana, kira, gargadi, kiranka, shawarci, shawartar, roko, bukaci, godiya,*
	hauWE SG	*kiran, kira, tattausar, kirana, bukaci, shawartar, shawararka, kiranka, kiranki, shawarci,*
	Bojanowski	*zurfi, karshen, kiransa, kirari, hira, misira, girmansa, suyi, kiransu, sulhu*
hadisi	hauWE CBOW	*hadisin,* ibnu, bukhari, *sahihul,* tirmizi, *ibn, hibban, mujalladi, khaldun,* muslim,
	hauWE SG	*hadisin, ibnu,* tirmizi, *sahihul,* bukhari, *isnadi, mujalladi, hudud, hibban,* nasai,
	Bojanowski	*hawza, hakuri, hajji, hatsi,* hajj, *agha, fikihu, masallatai,* madina, *ruhu*
kara	hauWE CBOW	*ƙara, qara, dada, rage, karawa, nunar, karin, saukaka, sake, karfafa,*
	hauWE SG	*dada, ƙara, qara, karawa, kudurinsu, yakara, karada, akara, karin, shigarmu,*
	Bojanowski	*karami, karo, karfi, kokari, ƙara, kari, kare, sankara, shawara, karya*
godiya	hauWE CBOW	*godiyar, godiyarsa, godiyata, jinjina, godewa, gode, godiyarmu, godiyarsu, taaziyya, mubayaa,*
	hauWE SG	*godiyata, godiyarmu, godiyar, godiyarsa, jinjina, godiyarta, gode, godewa, godiyarsu, bangajiya,*
	Bojanowski	*italiya, libya, najeriya, kambodiya,* kenya, *gabascin, girka, kudanci,* aljeriya, *koriya*
kuka	hauWE CBOW	*kun, ku, kika, muka, kunã, kuke, aka, lãrabci, kanku, zã,*
	hauWE SG	*kun, ƙaryatãwa, anani, jinku, ku, dakuka, karkiyata, aryatawa, shirinku, tambayarku,*
	Bojanowski	*gaggawa, kusurwa, yiwa, kowacce, iyaye, rinka, duka,* jigawa, *jurewa, kaucewa*
ibrahim	hauWE CBOW	aliyu, sulaiman, isah, adamu, bashir, salihu, yahaya, ibrahin, hassan, muhammad,
	hauWE SG	ibrahin, ibarahim, ibarhim, muhamamd, abdulhamid, mikailu, galadanci, mahammad, idris, abubukar,
	Bojanowski	*alaihi,* jawad, *umarni,* islam, ahmadu, *ummu,* zainab, abdullahi, ahmadi, *alkur*
so	hauWE CBOW	*son, sonsa, iyawa, nema, kaunar, tunanin, shaawa, bukata, wulakanta, alfahari,*
	hauWE SG	*son, sonsu, sonmu, shugabanka, kudurcewa, shizan, shiina, kaunace, dangantuwa, galibinmu,*
	Bojanowski	*can, re, st/acre,* arabic, fine, generated, done, compare, most, related
kallo	hauWE CBOW	*kallon, kallonsa, dugwidugwi, kallonsu, sara, kalo, wasansa, sauraro, ihu, tauri,*
	hauWE SG	*kallon, kallonsa, kallona, kallonsu, kalo, finafinaina, dariyar, downloading, lokeshin, finafinanmu,*
	Bojanowski	*kalli, kawai, bahaushe, kadai, koyon, kala, kawo, naka, allura,* jigawa
unguwa	hauWE CBOW	*anguwa, unguwar, kauye, dakinsa, gunduma, anguwar, unguwan, kauyen, mazaba, rumfa,*
	hauWE SG	*anguwa, dagaci, unguwar, magume, unguwarsu, unguwarmu, gunduma, festac, rugar, bukka,*
	Bojanowski	*soji, fito, kauri, ɓangare, ura, sarakuna, gona, gagara,* sahara, *gidansa*
dariya	hauWE CBOW	*murmushi, ihu, shaawa, mamaki, haushi, shiru, faxa, burge, tsorata, surutu,*
	hauWE SG	*dariyar, zolaya, murmushi, kwakwalwata, kyalkyace, zagina, kallona, kalleka, kyalkyale, mufeeda,*
	Bojanowski	*tudu, hudu, kwango, uku, wato, goma, najeriya,* pakistan, *takwas, tura*
kaga	hauWE CBOW	*kace, naga, ance, nasan, gashi, akace, suce, wallahi, ai, kam,*
	hauWE SG	*zakaga, anzo, suce, zakaji, baayi, akeso, inka, nakeyi, zaice, tinanin,*
	Bojanowski	*kenan, zakaga, kalla, habaka,* zidane, *kaka, awannan, yanzu, wannan, akalla*
sai	hauWE CBOW	*in, idan, tunda, ashe, bari, gara, shiru, saannan, to, zarar,*

	hauWE SG	*wartsake, sia, matsu, hayyacina, hakura, loton, jakata, basai, sannusannu, yanzuyanzu,*
	Bojanowski	*sauka, saidai, sako, salla, sace, ubangiji, matuka, zuba, kuwa, haihuwa*
kuma	hauWE CBOW	*saboda, domin, da, nufin, don, kuwa, amma, babancin, tare, dai,*
	hauWE SG	*jiinci, akenutsarwa, mã, daallah, laifidaga, fakuwa, roi, nasĩha, iblĩs, bayyanannen,*
	Bojanowski	*miyagun, kulawa, kula, hukunta, kunun, yazama, tsayawa, kasancewa, kanta, kun*
rawa	hauWE CBOW	*tsantsan, rawar, tsan, rawarsu, mahimmiyar, birki, muhimmiyar, burki, takawa, kahankadaran,*
	hauWE SG	*taka, muhimmiyar, rawar, mahimmiyar, muhimmiyyar, takawa, muhimiyar, bazarsu, mihimmiyar, rawan,*
	Bojanowski	*kulawa, talakawa, awa, sunayen, kwana, kwashe, tsayawa, nawa, kwanan, awannan*
kida	hauWE CBOW	*kidan, wakar, rera, waka, waƙa, kidi, wakokin, kadekade, wakokinsa, wakewake,*
	hauWE SG	*kidan, kidi, kidansa, kalangu, lela, waka, gurmi, wakokinsa, bakandamiya, wakar,*
	Bojanowski	*kiwo, tushe, aji, kanfanoni, gaji, katako, bude, kishirwa, kirkiro, gane*
waka	hauWE CBOW	*wakar, wakokin, kida, wakoki, waƙa, wakata, rubutu, rubucerubuce, shata, kidan,*
	hauWE SG	*wakar, wakata, wakokin, wakarsa, wakoki, waƙa, wakokinsa, kidan, bakandamiya, kida,*
	Bojanowski	*wakanda, kaka, kaucewa, zaka, kotunan, habaka, matuka, wucewa, mummunan, wacce*
habiba	hauWE CBOW	bilkisu, mariya, jamila, ummi, hajara, halima, maimuna, hassana, salma, maryam,
	hauWE SG	zarah, nusaiba, uwale, faiza, zuwaira, siyama, bilkisu, kanwarta, salma, delu,
	Bojanowski	ataturk, *nemo, tsoro, gobe, hakuri, bukatun, waxanda, budurwa, gaggawa, oi*
zainab	hauWE CBOW	fatima, amina, hadiza, aishatu, saadatu, hafsat, jummai, rukayya, maryam, jamila,
	hauWE SG	fatima, kande, hajiya, babaji, jummai, hafsat, talatu, hajia, bilkisu, aishatu,
	Bojanowski	*marmaro,* mirza*, chadi, ppp,* kebbi*, almasihu,* nuruddeen*,* islam, ummu, jawad
kalma	hauWE CBOW	*siffa, manhaja, itaciya, muujiza, taska, masarrafa, kalmar, dabia, maadana, mukala,*
	hauWE SG	*maanarta, kalmar, muujiza, taska, siffa, decryption, jimla, kalmomin, kalmomi, asalinta,*
	Bojanowski	*kida, gano, matsaloli, kiwo, kirkiro, kanfanoni, gane, kunshi, nasu, akafi*
musa	hauWE CBOW	sulaiman, shuaibu, tijjani, rabiu, abdulkadir, jamilu, saidu, muazu, salisu, iliyasu,
	hauWE SG	rabiu, isyaku, ismaila, sulaiman, sambawa, auwalu, msa, shamsudden, dangwani, abullahi,
	Bojanowski	mustafa, ahmadi, ahmad, khamene, umaru, ahmadu, musulme, jawad, iran, mali
abdullahi	hauWE CBOW	umar, sulaiman, jamilu, yahaya, aliyu, muhammad, kabir, haruna, saidu, shuaibu,
	hauWE SG	abdulahi, umar, abudllahi, abdulalhi, abudullahi, abdulhamid, ismail, ayagi, abullahi, abdillahi,
	Bojanowski	husaini, abdu, sayyid, ahmad, dr, ibrahim, ayatullah, khamene, malam, jawad
littafi	hauWE CBOW	*littafin, littafinsa, alurani, rubutu, littattafai, littãfi, attaura, littafina, makala, littattafan,*
	hauWE SG	*littafin, juzui, littattafan, littafinsa, littafina, littattafai, littattafansa, tatsuniyoyi, tatsuniyoyin, galatiyawa,*
	Bojanowski	*littafin, littafai, littafinsa, littafan, littãfi, littatafai, littattafai, littattafan, sani, li*

Table 3 Similar words predicted by the three models

Comparison of Various Instantaneous Power Methods for the Estimation of Time-Parameters of Poly Phase Coded Radar Signal

Ashraf Adamu Ahmad[1,2*], Mohammed Ajiya[2], Zainab Yunusa Yusuf[2], Lawal Mohammed Bello[2]

[1]Dept of Electrical and Electronic Engineering,
Faculty of Engineering and Technology,
Nigerian Defence Academy (NDA),
Kaduna, Nigeria.
[*]aaashraf@nda.edu.ng

[2]Dept. of Electrical Engineering,
Faculty of Engineering and Technology,
Bayero University, Kano (BUK),
Kano, Nigeria

***Abstract*— Key feature of the electronic intelligence (ELINT) setup as part of the electronic warfare is the interception and analysis of radar signals. This paper as part of ELINT provides a time parameter estimation algorithm associated with poly-phase pulsed radar signal of low probability of intercept (LPI) in a non-cooperative environment in order to identify its other characteristics. Three different methods based on power obtained at instants of time are used. The first two are obtained from the time-aspect approximation of a time-frequency distribution (TFD) while the third one is the traditional method of using its conjugate form. The main TFD used in this paper is Wigner-Ville distribution (WVD) of separable kernel and its brief heuristic approach was presented. Thereafter, an algorithm based on measuring the time-spectrum width is developed so as to determine two basic time parameter of the radar signal; either the pulse width (PW) or the pulse repetition period (PRP). The quality of the algorithm developed is tested at various signal-to-noise ratios (SNRs) in the presence of noise modeled by pseudorandom numbers of Gaussian nature. The Result obtained shows that for a fixed threshold of 37.5%, power at instants of time obtained from the TFD performs best at minimum SNR of -15 dB. The traditional method of getting power at instants of time performs very well if a variable threshold is applied at minimum SNR of -11 dB. The results shows that the proposed methods may be deployed in the practical ELINT field for analysis of radar signal**

***Index Terms*— Electronic Intelligence (ELINT), Low Probability of Intercept (LPI), Pulse Width (PW), Pulse Repetition Period (PRP), Time-Frequency Distribution (TFD), Wigner-Ville distribution (WVD).**

I. INTRODUCTION

Electronic intelligence (ELINT) primarily involves two sections; the inter-pulse analysis and the intra-pulse analysis. The inter-pulse analysis deals with the determination of pulse-to-pulse characteristics of radar signals such as pulse width (PW) and pulse repetition period (PRP) [1]. This analysis is the focus of this paper. The intra-pulse analysis deals with the determination of parameters within the pulse duration of a radar signal such as internal frequency and phase modulations characteristics [1]. PW is the time in which the radar system radiates each pulse while PRP or the interpulse period is the time between the beginning of one pulse and the start of the next pulse. Other sections of ELINT involves emitter location, direction finding, transmitter power estimation [1]. Inter-pulse analysis is considered in this paper in order to distinguish pulsed radar signals from the continuous wave (CW) ones.

PW and PRP are used conventionally to determine range resolution and unambiguous range respectively [2], they can also be used to estimate angle of target [3]. The PRP can undergo various changes to achieve specific functions such as staggering for blind speeds elimination in moving target indicator (MTI) radar [4] and jittering for specific type of jamming reduction [5]. The PRP can also undergo sliding for constant altitude coverage during radar elevation scanning among many variety of PRP schemes [1], [6].

Recently in the field of ELINT, a comparison of the spectrogram and the scalogram for the characterization of low probability of intercept (LPI) frequency hopping signal of 4-components and 8-components type was presented [7]. Modified B-distribution and particle filters for the estimation and extraction of four cases of radar signal models were considered [8]. The blind estimation of the instantaneous frequency hopping (FH) spectrum without the knowledge of hopping patterns using the joint time-frequency domain was also considered [9]. In a more robust research, an automatic waveform recognition method on eight LPI-based signal sets/classifications in a high noise environment using the Choi-Williams distribution (CWD) and Elman neural

network (ENN) supervised classification system was proposed [10], while a novel methodology of combining Cross-Wigner Ville distribution (XWVD) and Hough transform (HT) was used for detection and parameter extraction of frequency modulated continuous waveforms (FMCW) signals [11]. Most recently on poly phase coded radar signal, generalized Pareto distribution (GPD) was used for recognition of phase shift keying class of radar signals [12]. A brief theory and concise review of the use of time-frequency analysis for radar applications in the year of its publication can be found in [13].

The objective of this paper in line with previous papers is therefore to develop a modified Wigner-Ville distribution (mWVD), a joint time-frequency distribution (TFD) using non-adaptive windows for analysis of incoming polyphase radar signal. Thereafter the instantaneous power (IP) is derived via either the peaks of the TFD (IPm) or the time-marginal of the TFD (IPtm) or directly using its conjugate (IPd) for PW and PRP estimation at any chosen threshold. Finally methods presented are validated in the presence of noise at various signal-to-noise ratios (SNR).

II. REVIEW OF BARKER CODED POLY PHASE RADAR SIGNAL

The key aspect of the LPI design is the use of wideband CW radar waveform to achieve pulse compression. Therefore an intercept ELINT receiver must focus on counteracting the goal of the LPI radar design by coming up with an algorithm to analyse the intercepted signal appropriately. Phase modulating LPI radar signals obtained from phase shift keying (PSK) techniques and generally classified into two; binary phase modulation and polyphase modulation are required for pulse compression LPI radar. This is to have lower side lobes in order to prevent side lobe jamming and Doppler range tolerance due to its CW nature [14]. An analytical modulated digital phase signal can simply be represented by (1).

$$s(t) = Ae^{j(2\pi f_c t + \emptyset_k)} \quad (1)$$

where A is the amplitude of the signal, $\emptyset_k$ is the phase modulation function that is shifted in time t and, f_c is the carrier information contain in the frequency. The polyphase modulation of the radar signal considered in this paper uses polyphase codes similar to binary phase codes in terms of finite sentence length and discrete time complex sequences with constant magnitude. However it differs from the binary phase codes by having a variable phase $\emptyset_k$. Increasing the phase changes in the code sequence allows for construction of higher sequences. It also allows for counteracting the limitation of binary phase by allowing for lesser sidelobes, less sensitivity to Doppler shift and hence better pulse compression ratio [15]. The first among these polyphase codes are the ones constructed based on barker sequences and are of more than two value phase codes unlike Barker binary phase shift keying (BPSK) codes. The polyphase barker codes are numerous in number and no specific formula has been found for all possible formats but is normally design based on specific requirements. These requirements are based on increasing the number of alphabet size from two of BPSK (0 or 180^0) to higher sizes (i.e. $k > 2$) [14]. For example, a polyphase sequence of four length code types (±1,±j) based on the Barker codes sequence that translates to various phase changes is shown in Table 1.

TABLE VIII. SEVEN ALTERNATING BARKER CODES FOR POLYPHASE MODULATION

Code Length	Code elements
2	+1, +j
3	+1, +j, +1
4	+1, +j, -1, +j
5	+1, +j, -1, +j, +1
7	+1, +j, -1, +j, -1, +j, +1
11	+1, +j, -1, +j, -1, -j, -1, +j, -1, +j, -1
13	+1, +j, -1, -j, +1, -j, +1, -j, +1, -j, -1, +j,+1

Table 1 show alternating real and imaginary terms and it is seen that the odd length polyphase barker codes are palindromic and shows some degree of symmetry as compared to the binary representation. For a pictorial depiction of poly phase shift keying (PPSK), the time and frequency representation of the polyphase barker for code length of seven based on the sequence given in table 1 is shown in Fig. 1.

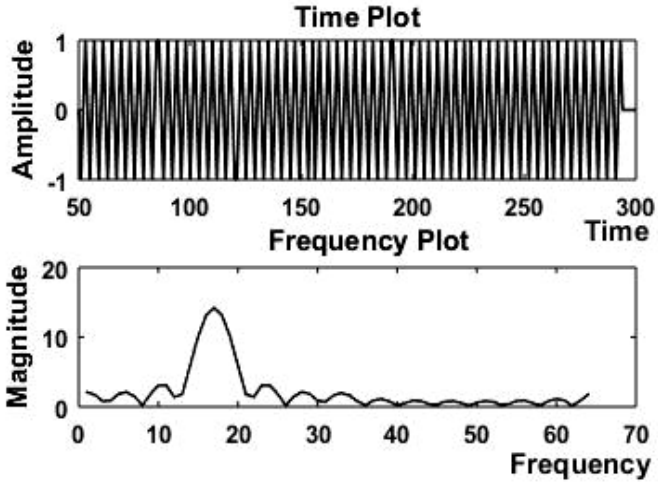


Fig. 10. Time and frequency representation of 7-length polyphase Barker Binary signal

Careful examination of the time representation of Fig. 1 shows that there is a constant phase change per sub-code (just like all the other polyphase barker codes) and in accordance with the 7-length sequence of +1+j-1+j-1+j+1. Also the frequency representation obtained from the fast Fourier transform (FFT) of the signal shows its constant frequency modulation characteristic. However, it does not show the phase changing characteristics of the signal.

III. METHODOLOGY

The quadratic TFDs (QTFDs) were developed to have better signal energy concentration distribution and good temporal resolutions of signal [16]. There are various ways of determining class of QTFDs. However, a major and common way considered in this paper is the through the WVD. The WVD, an extension of the Wigner distribution is a key QTFD that uses a form of quadratic function to concentrate the signal energy along the signal's instantaneous frequencies (IFs) [17, 18]. (2) and (3) are used to describe WVD.

$$W_z(t,f) = \underset{\tau \to f}{F}\{K_z(t,\tau)\} \quad (2)$$

$$W_z(t,f) = \int_{-\infty}^{\infty} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \quad (3)$$

where $F_{\tau \to f}$ denotes taking a Fourier transform with respect to time-lag, τ, $z(t)$ is the analytical or the complex form associate of the signal $s(t)$, and * (superscript) denotes the complex conjugate of the signal of interest. However, the WVD suffers major drawbacks of non-negative energy distribution, inner terms and cross-terms production in the presence of noise and other components; hence the need for modification as done by numerous researches including this one. One group of this modification is the development of other QTFDs known as reduced interference distributions (RIDs) whereby a unique kernel filter is designed to eliminate these limitations. The kernel filter may be non-separable such as the popular CWD (also known as the exponential distribution) that uses an exponential-like kernel function in the Doppler-lag domain as shown in (4) [19].

$$g(v,\tau) = e^{\left(\frac{-v^2\tau^2}{\sigma}\right)} \quad (4)$$

where v is the Doppler frequency, τ is the time-lag and σ is the variance Its corresponding time-lag equivalent is shown in (5) obtained from the Fourier transform relationship of the diamond diagram [16] (6) and the resulting QTFD with its full integral is also shown in (7).

$$G(t,\tau) = \int g(v,\tau) e^{-j2\pi vt} dv = \frac{\sqrt{\pi\sigma}}{|\tau|} e^{\frac{-\pi^2\sigma t^2}{\tau^2}} \quad (5)$$

From the diamond diagram

$$g(v,\tau) = \underset{t \to v}{F}\{G(t,\tau)\} \quad (6)$$

$$\rho_z(t,f) = \int\int \frac{\sqrt{\pi\sigma}}{|\tau|} e^{\frac{-\pi^2\sigma(t-u)^2}{\tau^2}} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} du d\tau \quad (7)$$

The CWD was recently used for LPI radar signal detection based on an optimized support vector machine classifier [20]. Alternatively, the kernel filter may be separable which is used in this paper. This scenario in the Doppler-lag domain is defined mathematically in (8).

$$g(v,\tau) = G_1(v) g_2(\tau) \quad (8)$$

Such that;

$$G_1(v) = \underset{t \to v}{F}\{g_1(t)\} \ , \ G_2(f) = \underset{\tau \to f}{F}\{g_2(\tau)\} \quad (9)$$

Therefore, mWVD used in this paper could be defined in (11) from the instantaneous autocorrelation function (IAF) of (10).

$$R_z(t,\tau) = g_2(\tau) K_z(t,\tau) \underset{t}{*} g_1(t) \quad (10)$$

$$\rho_z(t,f) = g_1(t) \underset{t}{*} W_z(t,f) \underset{f}{*} G_2(f) \quad (11)$$

where $R_z(t,\tau)$ is the generalized IAF of the value $K_z(t,\tau)$ for a normal WVD with no modifications. The Hamming window is used as the time-lag kernel ($g_2(\tau)$) for its low side-lobe level, while the Kaiser window is used as the time-smoothing kernel ($g_1(t)$) for its good ripple factor control. The mWVD can now be re-expressed by taking into cognizance this choice of kernel filters in (12).

$$\rho_{z,m}(t,f) = \int_{-\infty}^{\infty} \frac{I_0\left\{\beta\sqrt{1-\left(\frac{t}{\tau}\right)^2}\right\}}{I_0\{\beta\}} * 0.54 - 0.46\cos\left(\frac{2\pi\tau}{T}\right) z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \quad (12)$$

where a default value of $\beta = \frac{1}{2}$ is used in order to appropriately capture the radar signal, $I_0\{.\}$ is the modified zero-order Bessel function, T is the window length and $*$ (normal script) denotes convolution operations. After the development of the mWVD, the next set of objectives involving basic time signal parameters (PW and PRP) estimation were brought into view through the instantaneous power(IP,$P_i(t)$). The various ways of getting the IP is given in (13) – (15).

$$IPm = P_{i,m}(t) = \max\left(\rho_{z,m}(t,f)\right) \quad (13)$$

$$IPtm = P_{i,tm}(t) = \int_{-\infty}^{\infty} \rho_{z,m}(t,f) df \quad (14)$$

$$IPd = P_{i,d}(t) = |z(t)|^2 = z(t) * z^*(t) \quad (15)$$

where $z^*(t)$ is the complex conjugate of $z(t)$. The first way IPm is the approximate method of tracing the maximum power based on the frequency along the time axis while the second way IPtm is the summation (or integration) of power along the frequency and storage in time axis (hence its name time marginal).The last way IPd is the traditional obtainment of power at instants using its conjugate. Furthermore, the smooth versions of the IPs are obtained via convolution using a hamming window to smoothen out the rough edges due to noise and intra-phase modulation interactions. Its relation is given in (16).

$$P_{i,smoothed}(t) = w(t) * P_i(t) = \int_{-\infty}^{\infty} w(\lambda) P_i(t-\lambda) d\lambda \quad (16)$$

where $w(t)$ is the window and λ is the smoothing control operator. The ending aspect of the interpulse analysis deals with the design of a counting algorithm, capable of estimating all the PW and PRP present in the test signal from the IP selected based on specified threshold.

IV. RESULTS AND DISCUSSION

The performance validation section is obtained through iterative estimation of PW and PRP suitable range of SNR in the presence of noise modelled by pseudo random numbers of Gaussian in-line with practical radar scenario [21].The test barker PPSK radar signal of four standard pulses contains a PW of 2 μs to adequately cater for the poly phase changes and PRP of 100 μs for medium range radar [22]. It also

contain medium choice phase change of seven length barker codes, sampling frequency of 40 MHz based on current radar technology and center frequency of 10 MHz to adequately avoid aliasing. The PW length is divided equally into seven in order to represent the sub-pulse phase change. Three thresholds are considered for the performance analysis; 25%, 50%, and 37.5%; i.e. traditional threshold for LPI signals, non-LPI signals, medium threshold between these two respectively [1]. Suffices of –s and –ns are added to legend of the results in order to indicate the difference between the smooth and non-smooth versions of IP method respectively as given in (16) . The result obtained for the Barker coded PPSK radar signal at the 25% threshold for the PW estimations based on the probability of correct estimation (PCE) is given in Fig. 2. PCE is the ratio of correction estimations to total estimations.

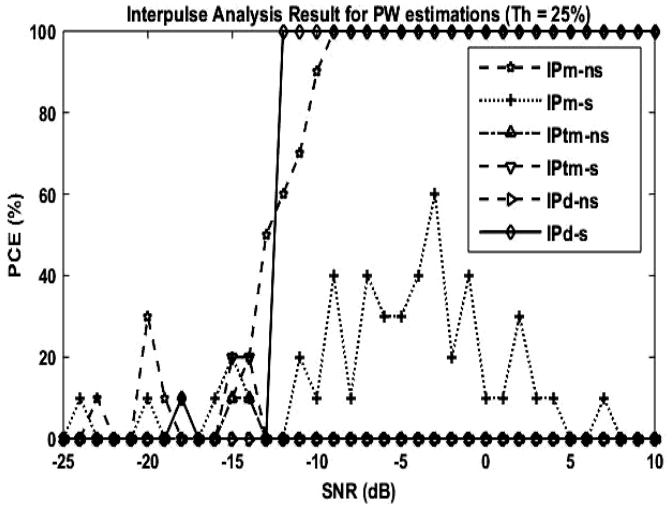


Fig. 11. PW estimation performance results for Barker PPSK radar signal at threshold = 25%

It is seen from Fig. 2 that some methods achieve 100% PCE at certain SNR while other do not achieve 100% PCE. The non-smooth version of IPm and smooth version of IPd achieve 100% PCE at SNR of -12 dB and -9 dB respectively. This indicates that IPm-ns is the best method for PW estimation of the test signal at threshold of 25%. The worst method from Fig. 2 is the non-smooth version of the IPd as no single value of PCE is achieved throughout the test SNR range. This is attributed to the internal phase changes of the test signal being affected by noise which the method cannot overcome. The remaining three methods, IPm-ns, IPtm-ns and IPtm-ns never achieve 100% PCE at any certain SNR but achieve certain values of PCE at specific SNR. This is attributed to interaction between the test signal and the random noise. The result obtained for the Barker coded PPSK radar signal for PW estimation at threshold of 37.5% is presented in Fig 3.

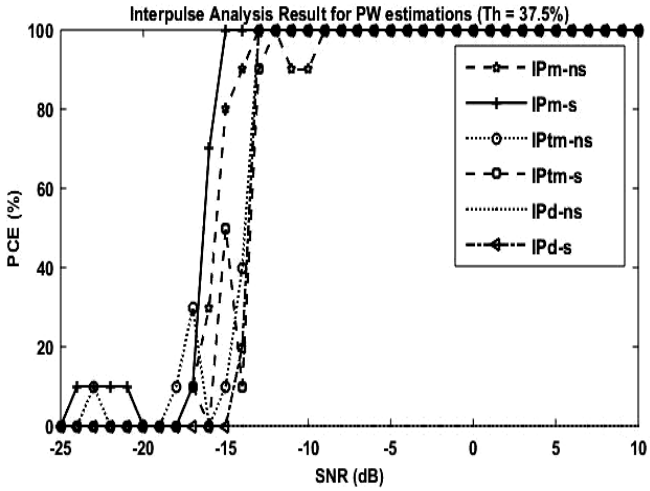


Fig. 12. PW estimation performance results for Barker PPSK radar signal at threshold = 37.5%

Fig. 3 shows that all but one method achieves PCE of 100% at certain SNR eventually with smooth version of IPm coming best having 100% PCE at minimum SNR of -15 dB. Similar to result in Fig. 2, the non-smooth version of IPd never achieves 100% PCE and same reason of internal phase changes can also be identified. The improvement in this result of Fig. 3 as compared to the one presented in Fig. 2 can be attributed to the change in threshold. The 37.5% threshold is able to capture the PW duration better than 25% threshold as the latter threshold proves to be quite low to achieve this ability. The result obtained for the PW estimation for the main test signal at threshold of 50% is presented in Fig. 4.

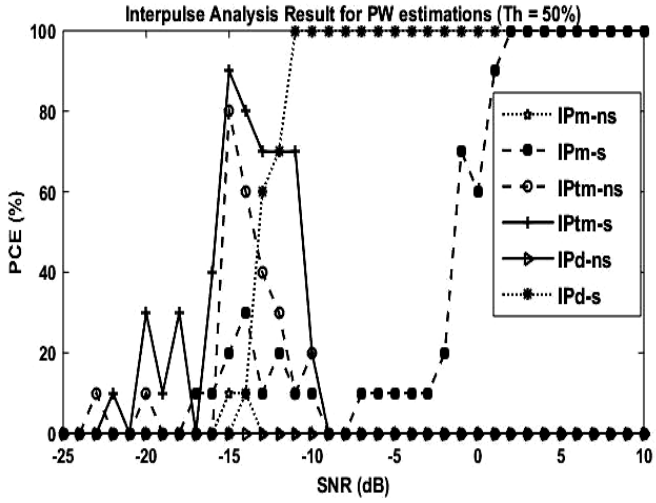


Fig. 13. PW estimation performance results for Barker PPSK radar signal at threshold = 50%

A reversal in improvement of result obtained is seen in Fig. 4 as compared to Fig. 3. This is because only two methods achieve 100% constant PCE at specific SNR as compared to five methods in the result of Fig. 3. As such, the result of Fig. 4 is similar to that of Fig. 2 in terms of number of methods that achieve 100% PCE. However these two methods of Fig.

4 are the smooth versions the IPm and IPd with 100% PCE achieved at SNR of 2 dB and -11 dB respectively. Similarly to previous results of Fig. 2 and Fig. 3 presented, IPd- ns never achieve a single PCE for the whole duration of the SNR range. Finally, the three remaining methods, IPm-ns, IPtm-ns and IPtm- s achieve certain value of PCE at specific SNR but never 100%.

It is also observed firstly that the random nature of the noise is responsible for the zigzag nature noticed in the estimation performance result obtained and hence justifying its selection in line with practical noise scenario. This deduction is similar to other papers highly related in nature to this paper [23, 24]. Secondly, some methods never achieve 100% PCE at specific threshold due to the effect of noise. From the careful study of the Fig. 2 to Fig. 4, the most versatile, worst and best method is the IPd-s, IPd-ns and IPm-s respectively. Smooth versions for most part in this estimation perform better as compared to the non-smooth versions. Another point to note is the appearance of positive SNR of 2 dB for 100% PCE estimation when IPm-s was used at a threshold of 50%, meaning that in this case, the signal power has to be greater than the noise power before a correct estimation is achieved. This can be attributed to the combination of a high threshold used and the method being an approximation from mWVD as compared to being obtained directly. The results obtained for the PRP estimation for the test radar signal at 25% threshold is given in Fig. 5.

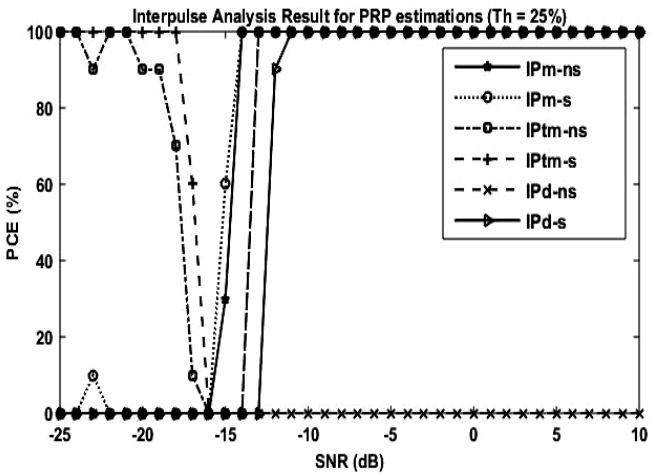


Fig. 14. PRP estimation performance results for Barker PPSK radar signal at threshold = 25%

All except the IPd-ns. method is seen to achieve 100% constant PCE at certain SNR in Fig. 5. This inability of IPd-ns to achieve 100% PCE can be attributed to the PW estimation despite the absence of modulation during the PRP duration. This is because the algorithm written for PRP estimation partially depends on the PW estimation to a certain level of degree. It can also be observed that IPm-ns, IPm-s and IPtm-ns all achieve 100% PCE at the same minimum SNR of -14 dB with different path followed to achieve this. The smooth version of IPtm and IPd achieves 100% PCE with difference of 1 dB and 3 dB from the minimum best SNR of -14 dB respectively. The results obtained for the PRP estimation for the test signal at 37.5% threshold is given in Fig. 6.

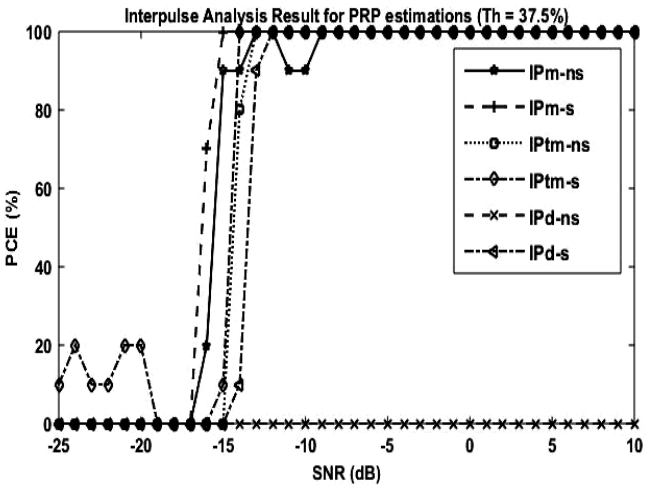


Fig. 15. PRP estimation performance results for Barker PPSK radar signal at threshold = 37.5%

The result obtained in Fig. 6 is similar to that of Fig. 5 in terms of five out the six methods obtained 100% PCE at various minimum SNRs. The only method that didn't achieve this feat still remains the IPd-ns due to the same reason presented in the discussion of Fig. 5. The best minimum SNR at this threshold for 100% PCE is at -15 dB obtained by the smooth version of IPm method. Therefore the result indicates a slight improvement of 1 dB at this threshold of 37.5% when compared to that of 25% threshold. The other remaining methods achieve 100% at other various higher SNRs with a maximum difference of 6 dB observed. Incidentally, the method to achieve 100% PCE at this maximum difference is the non-smooth version of IPm. This is henceforth demonstrating the superiority of the smooth version over the non-smooth version and hence providing justification for comparison and usage. The results obtained for the PRP estimation for the test signal at 50% threshold is given in Fig. 7.

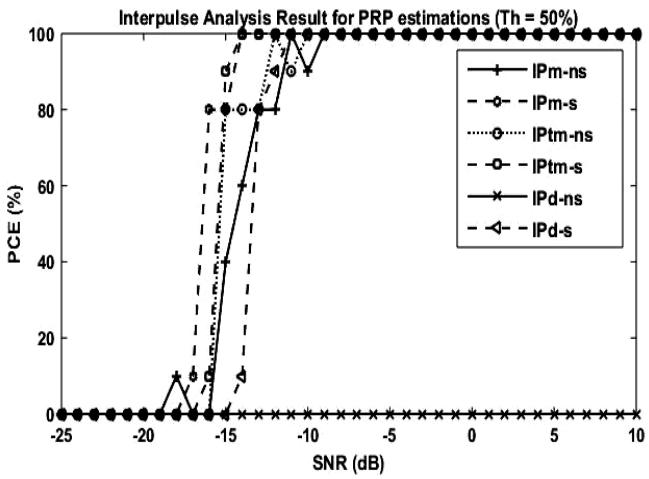


Fig. 16. PRP estimation performance results for Barker PPSK radar signal at threshold = 50%

The result obtained in Fig. 7 is also in tandem with the previous PRP estimation results of the test signal presented in Fig. 5 and Fig 6. This is because the IPd-ns never achieve 100% PCE at any SNR while the other five remaining methods do so at different SNR. Best minimum SNR of 100% PCE is similar to that of 25% threshold at SNR of -14 dB shared by the smooth versions of the approximate IPs from the main TFD, i.e. IPm-s and IPtm-s. The other three methods obtain 100% PCE at SNR difference of 3 dB, 4 dB and 5 dB. It is also observed that the path to achieving 100% PCE is slightly smoother in Fig. 7 than those of Fig. 5 and Fig 6. This indicates the threshold of 50% being a better choice than the other two thresholds in PRP estimation of the Barker coded PPSK signal.

Similar to PRP estimations of other papers of same objective, performance seems to be better than that of the PW due to high number of PRP samples. Also the IP-ns remain the worst method while IPm-s remains best method when the SNR for 100% PCE is examined for the last three figures. Versatility is also shared by all but the IPd–ns method; similarly and the general view of smooth version of the IP method being better than non-smooth versions also applies here. Another point noted and worthy of mentioning, is that smooth versions have similar results, i.e. Similar or very close SNR obtained for 100% PCE irrespective of the threshold selected with a difference of only 1dB noticed where applicable. This can be attributed to the absence of modulation during majority of PRP duration and as such effect of threshold selected being highly minimized. Furthermore, table 4 presents a comparison of similar papers based on time-parameter estimation where the minimum SNR of 100% probability of estimation is presented.

TABLE I. COMPARISON OF VARIOUS TIME PARAMETER ESTIMATION METHODS

Method	Minimum SNR
mWVD instantaneous power maxima (this paper)	2dB
Filters and Fast Fourier transform [23]	4dB
Smoothed Instantaneous Energy [24]	5dB
Short time Fourier transform [25]	9dB
Auto-Convolution/Peak Estimates [26]	20dB

It is seen in table 4 that this paper performs similarly to previous where minimum positive SNR or signal power being higher than the noise power is required for accurate time parameter estimation. However it outperforms previous paper due to this method requiring lower SNR as compared to others.

V. CONCLUSION

This paper presents an introduction into field of ELINT objectives, recent literatures and the characteristics of the main test signal of view, the polyphase radar signal. Thereafter, the approach used in getting the main tools of analysis, mWVD was presented and various ways of getting instantaneous power from its approximation were shown. Results obtained showed that smooth versions of the method used obtained through convolution operations/filtering generally outperforms the non-smooth ones. Different PW and PRP estimations results are obtained despite the usage of constant value emphasizing the effect of random noise on the methods examined. Generally the smooth version of IPd can be considered to be the best method due its capability to estimate any time parameter at any threshold if a choice of threshold is not specific. However, If the midpoint threshold of 37.5% between the other two standard ones of 25% and 50% is chosen the smooth version of the IPm performs best and justifies the use of main tool analysis of mWVD which can also be used to estimate other parameters. Finally, these parameter estimation methods may be used as a first step classifier capable of separating LPI signals from non-LPI signals before further analysis is carried out.

REFERENCES

[1] Wiley, R. G., *ELINT: The Interception and Analysis of Radar Signals*. Boston: Artech House, 2006.

[2] Skolnik, M., "Introduction to Radar Systems," *IEEE Aerospace and Electronic Systems Magazine,* 2001 16, (10), pp. 19-19.

[3] Weihong, L., Yongshun, Z., Guo, Z., and Hongbin, L., "A Method for Angle Estimation using Pulse Width of Target Echo," in *International Conference on Wireless Communications & Signal Processing, 2009. WCSP 2009.* , 2009, pp. 1-5.

[4] Sedivy, P., "Radar PRF Staggering and Agility Control Maximizing Overall Blind Speed," in *2013 Conference on Microwave Techniques (COMITE)*, 2013, pp. 197-200.

[5] Nishiguchi, K. i. and Kobayashi, M., "Improved Algorithm for Estimating Pulse Repetition Intervals," *IEEE Transactions on Aerospace and Electronic Systems,* 2000 36, (2), pp. 407-421.

[6] Matuszewski, J., "The Radar Signature in Recognition System Database," in *2012 19th International Conference on Microwave Radar and Wireless Communications (MIKON)*, 2012, pp. 617-622.

[7] Stevens, D. L. and Schuckers, S. A., "Low probability of Intercept Frequency Hopping Signal Characterization Comparison using the Spectrogram and the Scalogram," *Global Journal of Research In Engineering,* 2016 16, (2), pp. 13-23.

[8] Mikluc, D., Bujakovic, D., Andric, M., and Simic, S., "Estimation and Extraction of Radar Signal Features using Modified B Distribution and Particle Filters," *Frequenz,* 2016 70, (9-10), pp. 417-427.

[9] Liu, S., Zhang, Y. D., Shan, T., Qin, S., and Amin, M. G., "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum

estimation," in *Compressive Sensing V: From Diverse Modalities to Big Data Analytics*, 2016, p. 98570N.
[10] Zhang, M., Liu, L., and Diao, M., "LPI Radar Waveform Recognition based on Time-Frequency Distribution," *Sensors,* 2016 16, (10), pp. 1682-1702.
[11] Erdogan, A. Y., Gulum, T. O., Durak-Ata, L. t., Yildirim, T. l., and Pace, P. E., "FMCW Signal Detection and Parameter Extraction by Cross Wigner-Hough Transform," *IEEE Transactions on Aerospace and Electronic Systems,* 2017 53, (1), pp. 334-344.
[12] Li, Y., Guobing, H., Xiaoyang, X., and Xuelong, H., "A Recognition Algorithm for BPSK/QPSK Signals based on Generalized Pareto Distribution," in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, 2017, pp. 396-400.
[13] Ahmad, A. A., Abubakar, S. A., and Muhammad, A. L., "Recent Developments in the Use of Time-Frequency Analysis for Radar-Based Applications," in *AFRICON, 2015*, Addis Ababa, Ethiopia, 2015, pp. 1-5.
[14] Pace, P. E., *Detecting and Classifying Low Probability of Intercept Radar*. Norwood: Artech House, 2009.
[15] Blunt, S. D. and Mokole, E. L., "Overview of Radar Waveform Diversity," *IEEE Aerospace and Electronic Systems Magazine,* 2016 31, (11), pp. 2-42.
[16] Boashash, B., *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*. London: Academic Press, 2016.
[17] Wigner, E., "On the Quantum Correction for Thermodynamic Equilibrium " *Physical Review,* 1932 40, (5), pp. 749-755.
[18] Ville, J., "Theory and Applications of the Notion of Complex Signal, translated by I," *Seline in RAND Tech. Rpt, Santa Monica, CA,* 1958 92,
[19] Choi, H.-I. and Williams, W. J., "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 1989 37, (6), pp. 862-871.
[20] Liu, L., Wang, S., and Zhao, Z., "Radar Waveform Recognition Based on Time-Frequency Analysis and Artificial Bee Colony-Support Vector Machine," *Electronics,* 2018 7, (5), p. 59.
[21] Ziemer, R. E. and Peterson, R. L., *Introduction to Digital Communication* vol. 2: Prentice Hall Upper Saddle River, NJ, 2001.
[22] Stimson, G., *Introduction to Airborne Radar*. New Jersey: SciTech Pub, 1998.
[23] Pei, W. and Bin, T., "Detection and estimation of non-cooperative uniform pulse position modulated radar signals at low SNR," in *2013 International Conference on Communications, Circuits and Systems (ICCCAS)*, 2013, pp. 214-217.
[24] Adam, A., Adegboye, B., and Ademoh, I., "Inter-Pulse Analysis of Airborne Radar Signals using Smoothed Instantaneous Energy," *International Journal of Signal Processing Systems (IJSPS),* 2016 4, (2), pp. 139-143.
[25] Ahmad, A. A., Daniyan, A., and Gabriel, D. O., "Selection of Window for Inter-pulse Analysis of Simple Pulsed Radar Signal using the Short Time Fourier

Transform," *International Journal of Engineering & Technology,* 2015 4, (4), pp. 531-537.
[26] Chan, Y., Lee, B., Inkol, R., and Chan, F., "Estimation of Pulse Parameters by Convolution," in *Canadian Conference on Electrical and Computer Engineering, 2006. CCECE'06.*, 2006, pp. 17-20.

Cultured Artificial Fish Swarm Algorithm: An Experimental Evaluation

[1]M. L. Imam, [2]G. A. Olarinoye, [3]B. H. Adebiyi, [3]M. O. Momoh & [3]H. B. Salau

[1] Department of Electrical and Electronics Engineering, University of Jos, Nigeria
[2] Department of Electrical Engineering, Ahmadu Bello University Zaria, Nigeria
[3] Department of Computer Engineering, Ahmadu Bello University Zaria, Nigeria

ABSTRACT

In this work, a Weighted Cultural Artificial Fish Swarm Algorithm (wCAFSA) which is an amendment of standard artificial fish swarm algorithm (AFSA) is proposed. This algorithm can adaptively select its parameters at every generation in order to reduce the ease at which standard AFSA falls into local optimal. We first introduce inertial weight to adaptively determine visual distance and step size of AFSA thereafter, the Situational and Normative knowledge inherent in cultural algorithm are used to develop new variants of weighted cultural AFSA (wCAFSA Ns, wCAFSA sd, wCAFSA Ns+Sd and wCAFSA Ns+Nd). A collection of sixteen (16) optimization benchmark functions are used to test the performance of the algorithms. The simulation results disclosed that all the new variants of the wCAFSA outclassed the AFSA.

Keywords: Artificial Fish Swarm Algorithm; Inertia Weight; Culture Algorithm; Test Functions

I. INTRODUCTION

Artificial Fish Swarm Algorithm (AFSA) belongs to the class of swarm optimization algorithm. They are inspired by the intelligent behaviours of school of fish. AFSA was proposed in 2002 by Li Xio Lei [1]. In a pool of water, the fishes can locate the areas with more foods. This is because of its searching abilities, hence the location with more foods is generally the rich location [2]. The AFSA possess some distinctive features such as good convergence frequency, insensitivity to initial values, good fault tolerance and flexibility which enhances optimization efficiency. However, like most swarm algorithms AFSA sometimes losses its explorative mechanism so soon. Hence, the algorithm fails to attain a global solution especially when dealing with complex, multimodal problems [3]. To overcome these challenges, several research works have been carried out towards alteration to the original AFSA. The work presented in [4] proposed a modification to the AFSA by modifying the visual and step size. A leaping behavior was introduced to reduce the constant effect of visual distance while and adaptive technique was introduced such that the algorithm can iteratively select its step size. The method was implemented for job scheduling in grid computing. In their results, the modified method had a better performance than the original algorithm. In [5], a new AFSA for dynamic optimization problems was proposed. The algorithm was made dynamic with respect to the environmental nature of the optimization hyperspace. The performance of the developed algorithm was validated with ten benchmark algorithms and results disclosed it supremacy over the others. The work in [3] proposed an improved AFSA using cultural algorithm. The situational and normative knowledge were used to implement four new variation of AFSA. The performance of all the variants showed a significant improvement over the original AFSA. However, the visual and step size were still assigned a constant value. In [6], the idea of similar fragment in VRP was used to proposed a hybrid modification to AFSA. This was specific to solve the problem of poor convergence and improve precision. Modification AFSA using Levy Flight and Firefly. The praying and chasing behaviors of AFSA was modified using the movement strategies of firefly algorithms. Thereafter, the swarming behavior was modified using Levi flight and a dynamic parameter selection was introduced. The performance of the new methods was validated using some benchmark functions and results showed an improved performance. It has been established from the literatures that, the improvement of the convergence speed and optimization accuracy has been given a significant research attention. However, developing a new algorithm that will outperform other optimization algorithms in terms of attainment of global optimization result is still a serious subject of concern for researchers. It is on this gap that this paper was inspired.

II. METHODS

This section presents the methods used to developed the proposed modified algorithm.

Inertial Weight: Initially Particle Swarm Optimization (PSO), has no actual parameter(s) for the control of velocity. To address this shortcoming, a method called the inertial weight was introduced [8]. In this paper, an adaptive linear value was introduced to iterate the given weight in equation (1).

This is to ensure that every iteration has its suitable visual distance and step size.

$$w_t = \left(\frac{itt_{\max} - itt}{itt_{\max}}\right)^m \quad (1)$$

where, $itt_{\max}$ is the maximum number of iterations, itt is the current iteration and the size of the swarm is taken as m.

The visual and step size of the original AFSA becomes from equation (1) becomes:

$$visual_{w_t} = w_t \times visual \quad (2)$$

$$Step_{w_t} = w_t \times Step \quad (3)$$

The weighting factor w_t in equation (1) is used to control the step size and visual distance of AFSA population in an iterative manner.

2.1 wCAFSA Influence Function

The configuration of knowledge (situational and normative) influence the AFSA population. Henceforth four (4) variants of weighted cultural artificial fish swarm algorithm (wCAFSA Ns, wCAFSA Sd, wCAFSA Ns+Sd and wCAFSA Ns+Nd) is developed.

2.1.1 wCAFSA Ns

This variant utilizes only normative knowledge to guide the preying activities of wAFSA.

A) Preying Ns

A selection operator was first introduced to initialize the wCAFSA Ns. This operator is given as follows [3].

$$X_{ii}^{t+1} = X_{ii}^{t} + rand(0,1) \times size(I_{ii}) \quad (4)$$

where $size(I_{ii}) = u_{ii} - l_{ii}$ is the magnitude of the belief space interval for parameter ii, which is decided by the normative knowledge for the variable [15, 16] and $rand(0,1)$ is a normally distributed random variable with a mean of zero and a standard deviation of one (1). The mathematical expression describing the preying behaviour of wCAFSA Ns is given as follows:

$$x_{iik}^{t+1} = x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|} \quad (5)$$

where $size(I_{ik}) = u_k - l_k$ is taken as the size of the belief space interval for parameter k which is decided by the normative knowledge for the k^{th} variable. Equation (5) is the influence function when $ii \le try\_num$.

When $ii > try\_num$, the preying behaviour of wCAFSA_Ns is influenced as follows

$$x_{iik}^{t+1} = x_{iik}^{t} + size(I_k) \times rand(0,1) \quad (6)$$

B) Swarming_Ns

The swarming activities of wCAFSA Ns can be represented by the expression:

$$x_{iik}^{t+1} = x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|} \quad (7)$$

where $X_c = \sum_{j=1}^{mf} X_j / mf$ the centre position and mf is the number of its associates within the visual distance, x_{ck} is the centre position connected with the upper bound for variable j.

C) Chasing Ns

The chasing activities when only normative knowledge is used can be represented by this expression:

$$x_{iik}^{t+1} = x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t} - x_{iik}^{t})}{\|x_{ii}^{t} - x_{ii}^{t}\|} \quad (8)$$

where X_n stands for the best artificial fish individual within X_{ii}'s visual distance and nf is the number of X_i's fellows within the visual distance.

2.1.2 wCAFSA Sd

In this variant, the situational knowledge is used to guide the direction of evolution of AFSA.

A) Preying Sd

The description of the preying behaviour of wCAFSA Sd is as follows:

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left|\frac{rand(0,1) \times size_{iik} \times (x_{iik}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|}\right| & if\ x_{iik}^{t} < S_k^t \\ x_{iik}^{t} - \left|\frac{rand(0,1) \times size_{iik} \times (x_{iik}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|}\right| & if\ x_{iik}^{t} > S_k^t \\ x_{iik}^{t} + \frac{rand(0,1) \times size_{iik} \times (x_{iik}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|} & otherwise \end{cases} \quad (9)$$

Equation (9) is the influence function for preying wCAFSA_Sd when $ii \le try\_num$. When $ii > try\_num$, the preying behaviour of wCAFSA_Sd is influenced as follows:

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + |rand(0,1) \times step_{iter}| & if\ x_{iik}^{t} < S_k^t \\ x_{iik}^{t} - |rand(0,1) \times step_{iter}| & if\ x_{iik}^{t} > S_k^t \\ x_{iik}^{t} + rand(0,1) \times step_{iter} & otherwise \end{cases} \quad (10)$$

B) Swarming Sd

The explanation of swarming behaviour of the wCAFSA Sd is as follows:

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left|\frac{rand(0,1) \times size_{iik} \times (x_{ck}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|}\right| & if\ x_{iik}^{t} < S_k^t \\ x_{iik}^{t} - \left|\frac{rand(0,1) \times size_{iik} \times (x_{ck}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|}\right| & if\ x_{iik}^{t} > S_k^t \\ x_{iik}^{t} + \frac{rand(0,1) \times size_{iik} \times (x_{ck}^{t+1} - x_{iik}^{t})}{\|x_{ii}^{t+1} - x_{ii}^{t}\|} & otherwise \end{cases}$$

(11)

C) Chasing Sd

The chasing behaviour of the wCAFSA Sd is described as follows:

$$x_{ik}^{t+1} = \begin{cases} x_{ik}^{t} + \left| \frac{rand(0,1) \times step_{kv} \times (x_{mk}^{t+1} - x_{ik}^{t})}{\| x_{m}^{t+1} - x_{i}^{t} \|} \right| & if\ x_{ik}^{t} < S_{k}^{t} \\ x_{ik}^{t} - \left| \frac{rand(0,1) \times step_{kv} \times (x_{mk}^{t+1} - x_{ik}^{t})}{\| x_{m}^{t+1} - x_{i}^{t} \|} \right| & if\ x_{ik}^{t} > S_{k}^{t} \\ x_{ik}^{t} + \frac{rand(0,1) \times step_{kv} \times (x_{mk}^{t+1} - x_{ik}^{t})}{\| x_{m}^{t+1} - x_{i}^{t} \|} & otherwise \end{cases}$$

(12)

2.1.3 wCAFSA Ns+Sd

In this variant, both the normative and situational knowledge are integrated to guide the step size and direction of evolution respectively.

A) Preying wCAFSA Ns+Sd

The preying behaviour of wCAFSA Ns+Sd is designated as follows:

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\| X_{ii}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} < S_{k}^{t} \\ x_{iik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\| X_{ii}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} > S_{k}^{t} \\ x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{ik}^{t+1} - x_{ik}^{t})}{\| X_{ii}^{t+1} - X_{ii}^{t} \|} & otherwise \end{cases}$$

(13)

B) Swarming wCAFSA Ns+Sd

The report of chasing activities of wCAFSA Ns+Sd is as follows:

$$x_{ik}^{t+1} = \begin{cases} x_{ik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{ik}^{t})}{\| X_{c}^{t+1} - X_{i}^{t} \|} \right| & if\ x_{ik}^{t} < S_{k}^{t} \\ x_{ik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{ik}^{t})}{\| X_{c}^{t+1} - X_{i}^{t} \|} \right| & if\ x_{ik}^{t} > S_{k}^{t} \\ x_{ik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{ik}^{t})}{\| X_{c}^{t+1} - X_{i}^{t} \|} & otherwise \end{cases}$$

(14)

C) Chasing Ns+Sd

The account of Chasing conduct of wCAFSA Ns+Sd is as follows.

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t} - x_{iik}^{t})}{\| X_{n}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} < S_{k}^{t} \\ x_{iik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t} - x_{ik}^{t})}{\| X_{n}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} > S_{k}^{t} \\ x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t} - x_{ik}^{t})}{\| X_{m}^{t+1} - X_{i}^{t} \|} & otherwise \end{cases}$$

(15)

2.1.4 wCAFSA Ns+Nd

In this variant, only normative knowledge is used to guide the step size and the direction of evolution of the visual distance. The influence functions for these variants are expressed in below.

A) Preying wCAFSA Ns+Nd

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\| X_{i}^{t+1} - X_{i}^{t} \|} \right| & if\ x_{iik}^{t} < l_{k}^{t} \\ x_{iik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\| X_{ii}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} > u_{k}^{t} \\ x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{iik}^{t+1} - x_{iik}^{t})}{\| X_{ii}^{t+1} - X_{ii}^{t} \|} & otherwise \end{cases}$$

(16)

B) Swarming wCAFSA Ns+Nd

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{iik}^{t})}{\| X_{c}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} < l_{k}^{t} \\ x_{iik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{ik}^{t})}{\| X_{c}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} > u_{k}^{t} \\ x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{ck}^{t+1} - x_{ik}^{t})}{\| X_{c}^{t+1} - X_{ii}^{t} \|} & otherwise \end{cases}$$

(17)

C) Chasing wCAFSA Ns+Nd

$$x_{iik}^{t+1} = \begin{cases} x_{iik}^{t} + \left| \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t+1} - x_{iik}^{t})}{\| X_{n}^{t+1} - X_{ii}^{t} \|} \right| & if\ x_{iik}^{t} < l_{k}^{t} \\ x_{iik}^{t} - \left| \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t+1} - x_{iik}^{t})}{\| X_{n}^{t+1} - XI_{i}^{t} \|} \right| & if\ x_{iik}^{t} > u_{k}^{t} \\ x_{iik}^{t} + \frac{size(I_k) \times rand(0,1) \times (x_{mk}^{t+1} - x_{iik}^{t})}{\| X_{n}^{t+1} - X_{ii}^{t} \|} & otherwise \end{cases}$$

(18)

where S_k^t is the best paradigm parameter value of the situational knowledge for variable j in the belief space. Note that, if an individual's parameter value is smaller than that of the current best, probably it would be better to go toward the current best. The current best itself is mutated using a random direction, which is otherwise the case in equation (18) [10]. Details of these mathematical expressions can be found in [17,18,19,20].

The Flowchart of the proposed algorithm is given as follows:

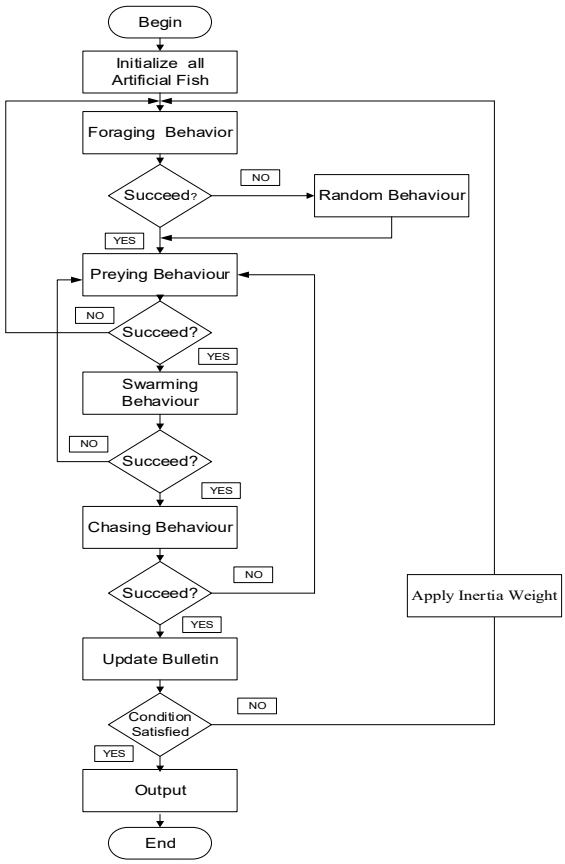


Fig 1: Flowchart of Modified Artificial Fish Swarm Algorithm

III. SIMULATIONS, RESULTS AND ANALYSIS

The evaluation of the performance of the proposed algorithms using sixteen (16) high dimensional applied mathematical optimization test functions and the average of 10 runs performed on Compaq Presario CQ56, with 64bit windows 8pro operating system and 3GB RAM were recorded.

- **wAFSA:** It is obvious from Table 1, that, wAFSA outperformed the AFSA in almost all the optimization test function except in test case 5, 6 and 9. However, AFSA did well for this class of functions (low-dimensional and relatively easy functions), but did poorly on others (high-dimensional and more complex functions). The superiority of wAFSA over AFSA is expected since inertial weight was introduced which guided its evolution towards the global optimal.
- **wCAFSA Ns:** In test case nine (9), eleven (11) and thirteen (13), the AFSA produced a better result when compared with wCAFSA Ns and in other test cases except 4 and 5, the wAFSA produced the best results. The introduction of knowledge in the belief space ensured a significant improvement on the precision accuracy of AFSA. In test case four (4) and five (5), all the algorithms produced the same result, which is an indication of unimodal nature of these functions.
- **wCAFSA Sd:** The situational knowledge based wCAFSA performed much better when compared with the AFSA in most of the optimization test functions.
- **wCAFSA Ns+Sd:** This variant produced the best result in ten (10) out of sixteen (16) cases. However, in test cases four (4) and five (5) the population-only AFSA and three of the self-adaption wCAFAC produced the same global result.
- **wCAFSA Ns+Nd:** The variant that employed normative knowledge for step size, visual distance and direction of evolution (wCAFSA Ns+Nd) has similar performance when compared with wCAFSA NsSd.

In summary, wAFSA and all the knowledged introduced variants outperformed the AFSA. This confirms the significance of the inertial weight and knowledge in the belief space. It can be observed that wCAFSA Ns+Sd is the most successful of all the variants, although it barely edged out wCAFSA Ns+Nd with significant margin in most of the test cases.

The unimodal and separable functions are relatively easier to optimize. On these functions, the performance of AFSA, wAFSA, and wCAFSA variants are mostly similar.

Table 1: Performance Evaluation of the proposed Algorithm

S/N	Test Functions	Global minimal	AFSA	wAFSA	wCAFAC Ns	wCAFAC Sd	wCAFAC Ns+Sd	wCAFAC Ns+Nd
1	Ackley	0.0000E+00	1.6650E-01	1.0590E-01	8.8818E-13	7.1200E-02	**3.0933E-16**	8.8818E-016
2	CM	-3.0000E+00	-2.7819E+00	2.8396E+00	-2.9380E+00	2.7769E+00	**-2.9370E+00**	-2.8650E+00
3	DejongF4	0.0000E+00	1.1540E-04	1.1918E-00	**0.22532E-05**	1.4889E-04	2.9161E-04	3.3725E-04
4	ExpFun	1.0000E+00	1.0000E+00	1.0000E+00	1.0000E+00	1.0000E+02	1.0000E+00	1.0000E+00
5	Grienwangk	0.0000E+00	0.0000E+00	1.4950E-260	0.0000E+00	1.7260E-189	0.0000E+00	0.0000E+00
6	Hepereliptic	0.0000E+00	4.9310E-01	7.2590E01	5.0055E-46	1.4510E-01	**5.0342E-58**	7.8235E-58
7	LM1	0.0000E+00	5.7000E-03	3.5000E-03	3.3000E-03	8.7000E-03	**1.2374E-50**	5.0929E-50
8	LM2	'0.0000E+00	1.2490E-01	7.7700E-02	9.0000E-03	1.2520E-02	**3.9114E-49**	4.5400E-50
9	Neunaier3	4.9300E+00	8.2196E+02	2.6576E+02	5.6179E+02	9.3558E+02	**-1.0126E+03**	-7.7103E+02
10	Rastrigin	'0.0000E+00	3.9448E+00	2.9670E+00	**1.9320E+00**	2.4909E+00	4.6819	4.0191
11	Rosenbrock	0.0000E+00	3.4313E+00	3.2519E+00	4.0827E+00	**3.2104E+00**	4.3120E+00	1.8244E+01
12	Sal	0.0000E+00	1.1350E+03	1.2000E-00	**9.0000E-04**	9.9900E+00	0.09999	3.0007E-01
13	Schwcfel	0.0000E+00	4.1500E+03	4.1359E-03	4.1504E+03	4.1505E+00	**3.5538E+03**	4.0521E+03
14	Schaffer	0.0000E+00	4.4977E+00	4.3354E+00	4.5235E+00	4.6513E+00	**0.2004**	6.5510E-01
15	Sphere	0.0000E+00	1.8300E-02	1.7300E-02	2.5000E+00	1.7500E-02	**2.9811E-12**	6.4954E-09
16	Bukin	0.0000E+00	3.6973E-03	1.3883E-03	2.9825E-16	3.7531E-18	4.3593E-19	**3.7531E-19**

Therefore, it can be deducted that, all the new variants of AFSA, have better performance than the standard algorithm. Also, it can be seen that the normative knowledge seems to be the dominant knowledge over the situational knowledge.

CONCLUSION

In this paper, an improved artificial fish swarm algorithm called the weighted cultural artificial fish swarm algorithm (wCAFSA) has been proposed. An adaptive parameter selection approach was introduced such that AFSA can select a suitable visual distance and step size. The situational and normative knowledge inherent in cultural algorithm were then used to produce four (4) variants of wAFSA called the wCAFSA Ns, wCAFSA Sd, wCAFSA Ns+Sd and wCAFSA NsNd. Simulation results indicate that the proposed algorithm outclassed the standard AFSA with wCAFSA Ns+Sd having a superior performance over AFSA, wAFSA, wCAFSA Sd and wCAFSA Ns+Nd. This paper focused on evolving and improving the optimization capability of AFSA using situational and normative knowledge, which are intrinsic in cultural algorithm. In our future work, we will be exploring other forms of cultural knowledge and more configurations of cultural knowledge will be considered. The research will be extended to engineering applications like PID controller tuning, Data Clustering, Wireless Sensor networking, Image Processing etc.

References

[1] X. L. LI, J. Shao and J. X. Qian, "An optimizing method based on autonomous animats: fish-swarm algorithm," vol. 22, no. 11, pp. 32-38, 2002.p

[2] Y. Wang and L. Li, "An improved intelligent algorithm based on the group search algorithm and the artificial fish swarm algorithm," 2015.

[3] Y. Wu, X.-Z. Gao, and K. Zenger, "Knowledge-based artificial fish-swarm algorithm," vol. 44, no. 1, pp. 14705-14710, 2011.

[4] S. Farzi, "Efficient job scheduling in grid computing with modified artificial fish swarm algorithm," vol. 1, no. 1, p. 13, 2009.

[5] D. Yazdani, M. R. Akbarzadeh-Totonchi, B. Nasiri, and M. R. Meybodi, "A new artificial fish swarm algorithm for dynamic optimization problems," in *2012 IEEE Congress on Evolutionary Computation*, 2012, pp. 1-8: IEEE.

[6] J. Che, K. Zhou, K. Zhang, X. Tong, X. Zhou, S. Jia and Y. Zhen "Application of hybrid artificial fish swarm algorithm based on similar fragments *in VRP"* (Tenth

International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2017)). SPIE, 2018.

[7] D. Yazdani, B. Nasiri, A. Sepas-Moghaddam, M. Meybodi, M. J. S. Akbarzadeh-Totonchi, and E. Computation, "mNAFSA: A novel approach for optimization in dynamic environments with global changes," vol. 18, pp. 38-53, 2014.

[8] C. Ye, "WSNs Performance Evaluation Using Non-equilibrium Statistical Mechanics Method" 2014. International Journal of Future Generation Communication and Networking, 7(6), pp.155-164.

[9] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, 2014. "Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications," vol. 42, no. 4, pp. 965-997.

[10] W. Shen, X. Guo, C. Wu, and W. Desheng "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm" 2011. vol. 24, no. 3, pp. 378-385

[11] C. Zhang, F.M. Zhang, F. Li, and H. S Wu, "Improved artificial fish swarm algorithm," in 2014 9th IEEE Conference on Industrial Electronics and Applications, 2014, pp. 748-753: IEEE.

[12] N. H. Awad, M. Z. Ali, P. N. Suganthan, and R. G. Reynolds, "CADE: a hybridization of cultural algorithm and differential evolution for numerical optimization" 2017 vol. 378, pp. 215-241.

[13] P. Azad and N. J Navimipour Computing "An energy-aware task scheduling in the cloud computing using a hybrid cultural and ant colony optimization algorithm" 2017 vol. 7, no. 4, pp. 20-40.

[14] C. A. C. Coello and R. L. Becerra, "Evolutionary multiobjective optimization using a cultural algorithm," in *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS'03 (Cat. No. 03EX706)*, 2003, pp. 6-13: IEEE.

[15] R. G. Reynolds and C. Chung, "Fuzzy approaches to acquiring experimental knowledge in cultural algorithms," in *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*, 1997, pp. 260-267: IEEE.

[16] M. Mohammadhosseini, A. T. Haghighat, and E. J. T. J. o. S. Mahdipour, "An efficient energy-aware method for virtual machine placement in cloud data centers using the cultural algorithm," pp. 1-30, 2019.

[17] A. T. Salawudeen, A. O. Abdulrahman, B. O. Sadiq and Z. A. Mukhtar, An Optimized Wireless Sensor Network Deployment Using weighted Artificial Fish Swarm (wAFSA) Optimization Algorithm. 2016. International Conference on Information and Communication Technology and its Applications. p. 203-207

[18] A. T. Salawudeen, Development of an Improved Cultural Artificial Fish Swarm Algorithm with Crossover. M.Sc Thesis. Department of Electrical and Computer Engineering, Ahmadu Bello University, Zaria,Nigeria. p. 158. 2015.

[19] M. B. Mu'azu, A. T. Salawudeen, T. H. Sikiru, A. I. Abdul and A. Mohammed. “Weighted artificial fish swarm algorithm with adaptive behaviour based linear controller design for nonlinear inverted pendulum,”. Journal for Engineering Researcch, Faculty of Engineering, University of Lagos. Vol 20(1), 2015.

[20] B. H. Adebiyi, M. B. Mu’azu, A. M. S. Tekanyi, A. T. Salawudeen, and R. F. Adebiyi. "Tuning of PID Controller Using Cultural Artificial Bee Colony Algorithm". Aride Zone Journal of Engineering, Technology and Environment. Faculty of Engineering, University of Maiduguri, Nigeria. 2018. Vol. 14(3): 491-496.

Speed Control of a Three Phase Induction Motor using a PI Controller

G. A. Olarinoye, C. Akinropo
Department of Electrical Engineering, Ahmadu Bello University
Zaria, Nigeria
oyeguzeh@yahoo.com,
akinropocharles@gmail.com

G. J. Atuman
Department of Computer Engineering, Ahmadu Bello University, Zaria Nigeria
atumanjg@yahoo.co.uk

Z. M. Abdullahi
Department of Communications Engineering, Ahmadu Bello University
Zaria, Nigeria
zmabdullahi@abu.edu.ng

***Abstract*—Three-phase Induction motors are typically used in an increasing variety of applications such as fans, milling machines, transportation, etc. This paper discusses the speed control of the three-phase induction motor and an associated PI controller. A model of the three-phase induction motor is presented, analyzed and used to demonstrate the effectiveness of the speed controller. The performance of the motor system is carefully examined and compared with and without the PI controller. The unit step response of the speed control loop is characterized by rise time, peak overshoot, settling time and steady state error of 0.0815s, 6.97%, 0.82 and 0.11% respectively. Simulation results show that the speed of the uncontrolled motor changed whereas that of the controlled motor returned quickly to its initial value after the motor is subjected to load changes from 1 pu to 0.5 pu and subsequently to 0.25 pu.**

Keywords — controller, induction motor, model, steady state and speed control.

I. INTRODUCTION

The induction motor is widely acclaimed to be the work horse of the industry owing to its robustness, low cost and ruggedness and the ease of speed control [1]. These motors are designed to operate efficiently under rated conditions. Rated condition is one in which the motor is operating at rated speed and carrying rated load.

Unfortunately, the induction motors do not operate at rated conditions most of the time. This is because they are often made to carry non-rated loads which includes low, partial and overloads. Non-rated loads cause a deviation from the rated speed; a situation that leads to inefficient operation of the motor. It is desirable to maintain a constant rated speed of operation regardless of the load applied to the motor in order to keep efficiency levels high. It is for this reason that motor speed control becomes necessary. Speed controllers meet this need and therefore are extremely valuable in applications where constant speed is required.

One good example of speed controllers is the classical PI controller. It still plays an extremely important role in Engineering practice and industries today. Despite advances in the development of control techniques, there is still nothing like the classical proportional-integral (PI) and PID controllers [2], [3]. One of the reasons for this, perhaps, is because these controllers do not rely on precise mathematical models and can eliminate steady state errors through the integral action [4]. The choice of parameters, of course, determines the effectiveness of the controller which can be a complicated assignment. It is complicated because of the difficulty in obtaining a closed form formula for calculating the controller gains. The approach employed in a lot of industrial processes depend mostly on experience and/or experiments [2].

In recent times, the classical PI controller has been effectively utilized to achieve very good system performances. A PI controller was used to control the speed of a doubly-fed three phase induction motor in [5]. The speed response shows a faithful tracking of the speed command in steady state. In [1] the use of Atmega 328/p with PI controller for speed control of a three-phase inverter-fed induction motor was presented. Simulation results in MATLAB/Simulink showed different rise time and percentage speed overshoot values but very minimal steady state speed error for three different pulse width modulation techniques. Reference [6] proposed an advanced speed control for an inverter-driven dual three-phase induction motor system in which a PI controller was used to establish a stable control loop regardless of changing load conditions. In [7] and [8], a PI controller was designed to ensure the global synchronization of some directed networks and to increase the robustness of the whole system respectively

This paper provides a method of obtaining an explicit formula for calculating the proportional and integral gains of the PI controller and tests the design on a three-phase induction motor model. This paper will show that the closed loop speed control of the induction motor is stabilized by the calculated controller gains. The rest of the paper is organized into sections. Section I gives the introduction while section II describes the model of the three-phase induction motor and formulation of the controller gains. Section III provides the results and discussion. Section IV concludes the paper.

II. MATHEMATICAL MODEL

The schematic diagram of a line-operated three-phase induction motor is illustrated in Figure 1. Its equivalent in d – q coordinates is illustrated in Figure 2. The model of the three-phase induction motor with the d-axis aligned with the rotor flux linkage is provided in equations (1) – (10) [9], [10].

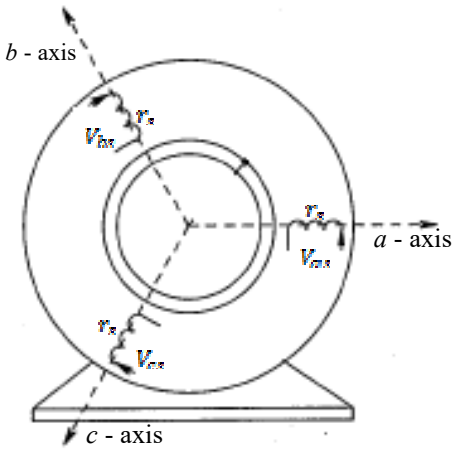


Figure 1. Three Phase Induction Motor

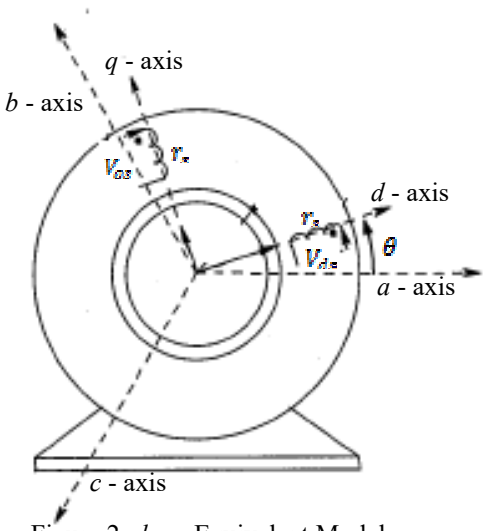


Figure 2. $d-q$ Equivalent Model

$$\begin{bmatrix} V_{qs} \\ V_{ds} \\ V_{0s} \end{bmatrix} = K_s \begin{bmatrix} V_{as} \\ V_{bs} \\ V_{cs} \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{0s} \end{bmatrix} = K_s \begin{bmatrix} i_{as} \\ i_{bs} \\ i_{cs} \end{bmatrix} \quad (2)$$

$$K_s = \begin{bmatrix} \cos\theta & \cos(\theta - 2\pi/3) & \cos(\theta + 2\pi/3) \\ \sin\theta & \sin(\theta - 2\pi/3) & \sin(\theta + 2\pi/3) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} V_{qs} \\ V_{ds} \end{bmatrix} = r_s \begin{bmatrix} i_{qs} \\ i_{ds} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_{qs} \\ \lambda_{ds} \end{bmatrix} + \omega_s \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{qs} \\ \lambda_{ds} \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} V'_{qr} \\ V'_{dr} \end{bmatrix} = r_r' \begin{bmatrix} i'_{qr} \\ i'_{dr} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} 0 \\ \lambda'_r \end{bmatrix} + (\omega_s - \omega_r) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \lambda'_r \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} \lambda_{qs} \\ \lambda_{ds} \\ 0 \\ \lambda'_r \end{bmatrix} = \begin{bmatrix} L_s & 0 & L_{ms} & 0 \\ 0 & L_s & 0 & L_{ms} \\ L_{ms} & 0 & L'_r & 0 \\ 0 & L_{ms} & 0 & L'_r \end{bmatrix} \begin{bmatrix} i_{qs} \\ i_{ds} \\ i'_{qr} \\ i'_{qr} \end{bmatrix} \quad (6)$$

$$\frac{d}{dt}\lambda'_r + \frac{\lambda'_r}{\tau} = \frac{L_{ms}}{\tau} i_{ds} \quad (7)$$

$$\omega_{sl} = \omega_s - \omega_r = \frac{L_{ms}}{\tau \lambda_r} i_{qs} \quad (8)$$

$$\lambda'_r = L_{ms} i_{ds} \quad (9)$$

$$T_{em} = \frac{P}{2} \frac{\lambda'_r L_{ms} i_{qs}}{L'_r} \quad (10)$$

The equation that describes the mechanical dynamics of the machine is given in (11).

$$T_{em} = J_i \frac{d\omega_m}{dt} + T_L \quad (11)$$

where V_{qs}, V_{ds}, V'_{qr} V'_{dr} are the dc stator and rotor voltages in the q-d synchronous reference frame. Subscript 's' represents stator quantities while subscript 'r' represents rotor quantities. i_{qs}, i_{ds}, i'_{qr} i'_{dr} are the dc stator and rotor currents. λ_{qs}, λ_{ds} and are the flux linkages in the stator while λ'_r is the rotor flux magnitude. L_{ms} is the magnetizing inductance. L_s is self-inductance of the stator windings while L'_r is the self-inductance of the rotor windings. Primed quantities are quantities of the rotor referred to the stator. ω_m and ω_r are the mechanical and electrical rotor angular speed respectively. J_i is the inertia of the rotor and the connected load while T_L is the load torque. T_{em} is the electromagnetic torque produced in the motor. τ is the rotor time constant. ω_s and ω_{sl} are the supply and slip angular frequency respectively. θ is the reference frame angle. The solution of the induction motor model is provided in the block diagram of Figure 3. In this Figure, three phase voltages V_{as}, V_{bs} and V_{cs} are fed to the windings of the motor. Three phase ac currents are the inputs to the system while speed is the output. The load torque T_L, represents the disturbance in the system which may mean a sudden increase or decrease in the load applied to the motor. The behavior of the motor speed, in steady state, when the load on the motor is suddenly reduced from its rated value is examined and discussed in section III.

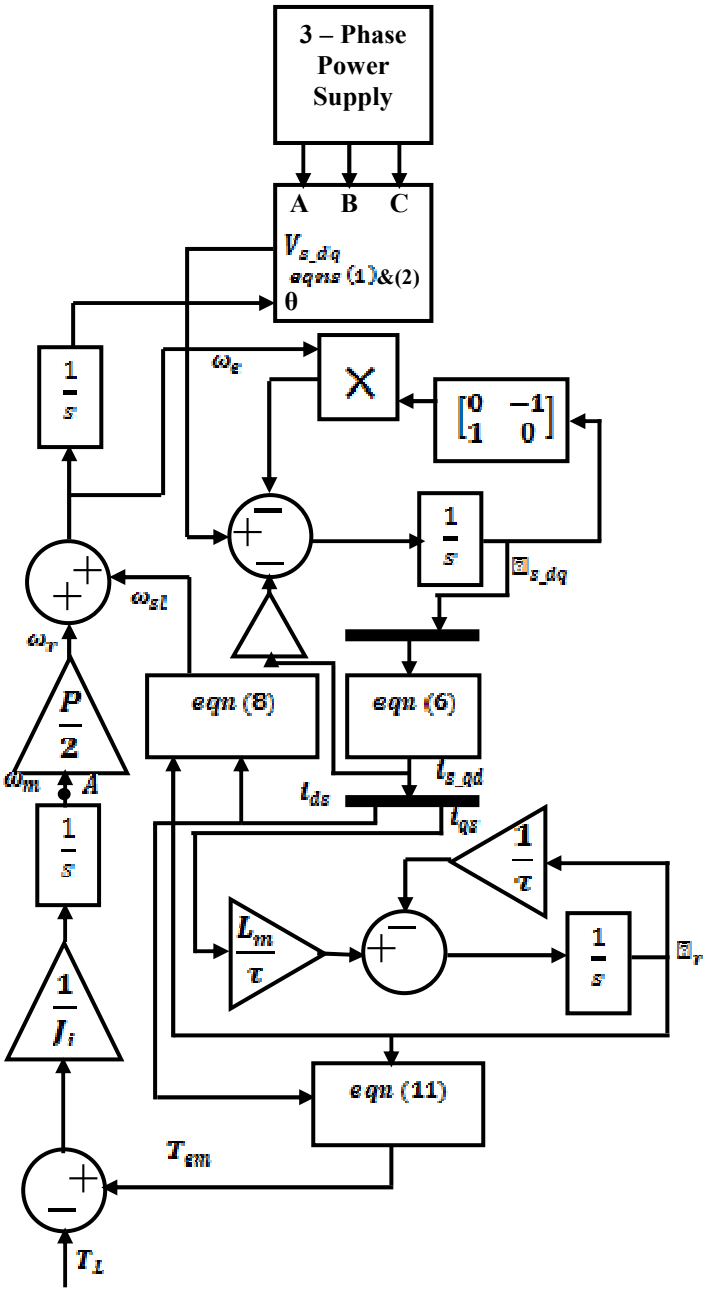


Figure 3. Block diagram representation of the induction model solution

A. Motor Control

The block diagram of the control system of the induction motor is shown in Figure 4. A current regulated converter is employed to deliver desired current to the three phase-induction motor. The objective of the control design is to make the motor speed constant at rated value (or initial value) regardless of any load disturbances. A simplified equivalent block diagram of the speed control loop is provided in Figure 5. The output of the PI block is the dc *q*-axis stator current. The relationship is obtained from (10) – (11) with the assumption that $T_L = 0$. This current is then transformed to sinusoidal reference motor currents i^*_{abc} using (2). The reference currents are translated to the motor as command inputs by the converter.

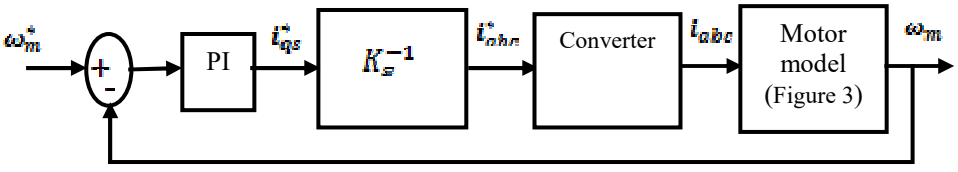


Figure 4. Speed control loop block Diagram

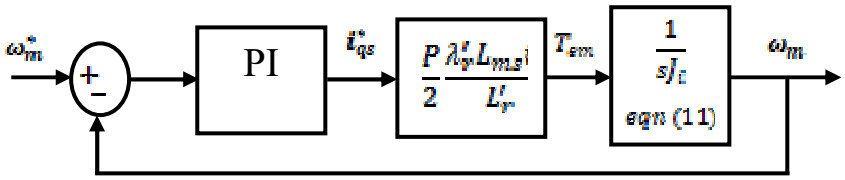


Figure 5. Simplified block diagram of the speed control loop

B. Controller Design

The control system block diagram of Figure 5 is reduced as shown in Figure 6

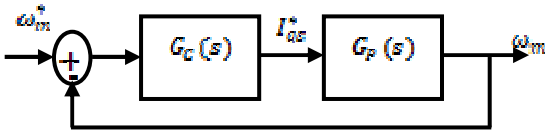


Figure 6. Reduced speed control loop block Diagram

$$G_C(s) = k_p + \frac{k_i}{s} \tag{12}$$

$$G_p(s) = \frac{c}{sJ_i} \tag{13}$$

$G_C(s)$ is the transfer function of the PI controller while $G_p(s)$ is the transfer function of the motor and the connected load.

$$c = \frac{P}{2}\frac{\lambda_r' L_{ms}}{L_r'} \tag{14}$$

$$G_0(s) = G_C(s)\,G_p(s) = \frac{k_p\left(s+\frac{k_i}{k_p}\right)c}{s^2 J_i} \tag{15}$$

$G_0(s)$ is the open loop gain of the speed control loop. The closed loop gain $G_{CL}(s)$ can be obtained from the following;

$$G_{CL}(s) = \frac{G_0(s)}{1+G_0(s)} \tag{16}$$

In this design, closed loop bandwidth is chosen equal to the crossover frequency and is selected arbitrarily as 25 rad/s (Mohan, 2014) A phase margin of 60^0 is chosen for a stable closed loop operation. Considering that the open loop gain is unity at crossover frequency, the following equation is valid;

$$\left|\frac{k_p\left(j\omega_c+\frac{k_i}{k_p}\right)c}{-\omega_c^2 J_i}\right| = 1 \tag{17}$$

Also valid is the following equation considering that the phase delay caused by the open loop gain must be less than 180^0 for a stable closed loop speed control.

$$\angle\left(\frac{k_p\left(j\omega_c+\frac{k_i}{k_p}\right)c}{-\omega_c^2 J_i}\right) + 180^0 = Phase\ margin \tag{18}$$

The parameters k_i and k_ω are calculated by solving equations (17) and (18) simultaneously. They are obtained as follows;

$$k_i = \frac{J_i\omega_c^2}{k\sqrt{1+\left(\frac{\omega_c k_p}{k}\right)^2}} \tag{19}$$

$$k_p = \frac{\tan(phase\ margin)\,k_i}{\omega_c} \tag{20}$$

III. RESULT AND DISCUSSION

The parameters of the selected three-phase induction motor are provided in Table 1. The induction motor was operated in steady state. The load on the motor was suddenly reduced by half from rated value of 1 pu to 0.5 pu and remained at that value for 1.5s. The load was again reduced by half to 0.25 pu and this load was maintained till end of simulation time. These disturbances were triggered at simulation time of 0.5s and 2s. The electromagnetic torque and speed performances are plotted in Figure 7.

TABLE 1. PARAMETERS OF THE THREE-PHASE INDUCTION MOTOR

Parameters	Values
Rated Power	2.4kW
Line Voltage	460V
Line frequency	60Hz
Rated Speed	1770 rpm
Stator resistance	1.77 Ω
Rotor resistance	1.34 Ω
Stator leakage inductance	0.01392 H
Rotor inductance	0.0126 H
Mutual inductance	0.369 H

It can be observed in Figure 7, that the torque produced in the motor falls first from a rated value of 1 pu to a steady value of 0.5 pu and second from 0.5 pu to 0.25 pu. Oscillations can be observed from the time of the disturbance to the time when the torque steadies at the respective values. The speed of the motor, on the other hand, rises from its rated value of 1770 rev/min to 1785 rev/min after the change in load from 1 pu to 0.5 pu and then from 1785 rev/min to 1793 rev/min after the sudden change of load from 0.5 pu to 0.25 pu.

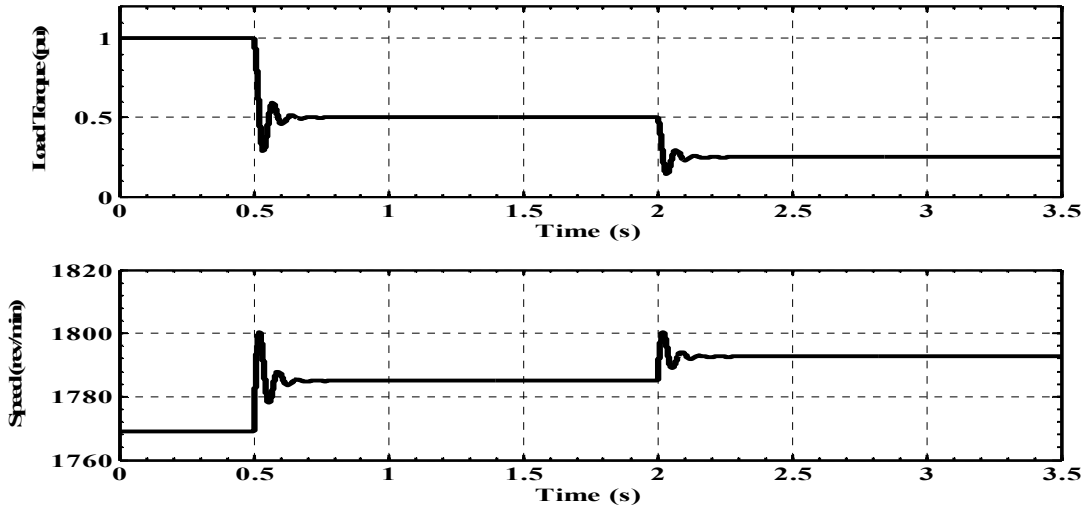


Figure 7. Electromagnetic torque and speed performances of the uncontrolled three-phase induction motor

Oscillations can be observed in the speed as well during these changes. In an un-controlled system such as this one, speed changes after load disturbances lead to inefficient motor operation. In high performance applications such as robotics and factory automation, speed changes can lead to loss of precision. Neither of these conditions is desirable. Hence the need for speed control.

The unit step response of the system of Figure 4 is plotted in Figure 8. It can be seen that the response tracks its reference faithfully. Associated system performance indices were calculated using the control system tool box in MATLAB and are provided in Table 2.

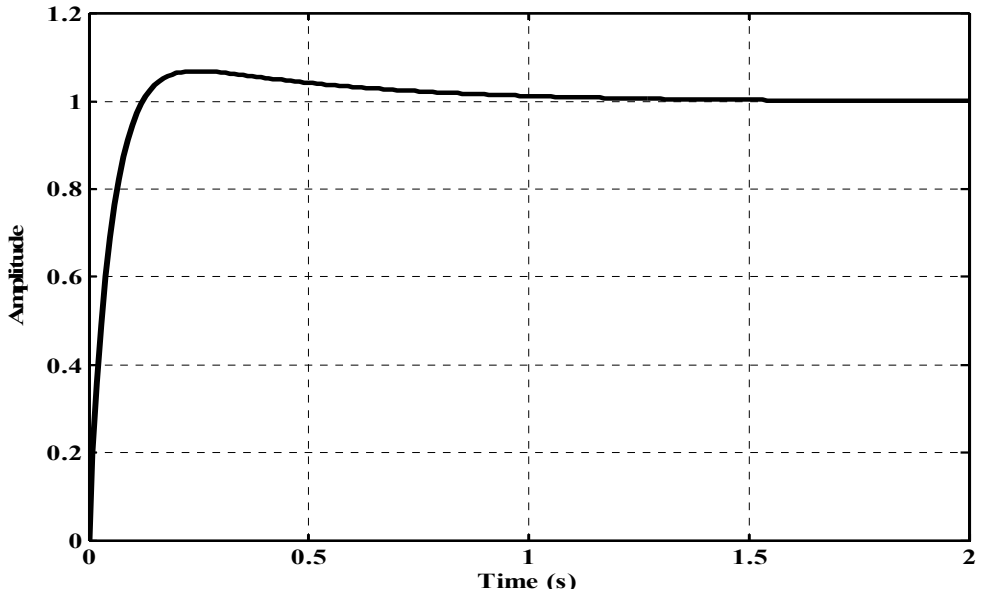


Figure 8. Unit step response of the speed control loop

TABLE 2. COMMAND-TRACKING PERFORMANCE OF THE SPEED CONTROL LOOP

Performance Index	Value
Rise Time	0.0815s
Maximum Overshoot	6.9664%
Peak Time	0.25s
Settling Time	0.82s
Steady State Error	0.11%
Closed Loop Poles	-2.4412, -19.2137

The speed control loop poles have negative real values indicating that the closed loop is stable. The controlled system of Figure 4 was also implemented in MATLAB/Simulink with controller parameters k_i and k_p calculated as 3.53 and 0.24 respectively. The electromagnetic torque and speed performances of this controlled Induction motor are plotted in Figure 9.

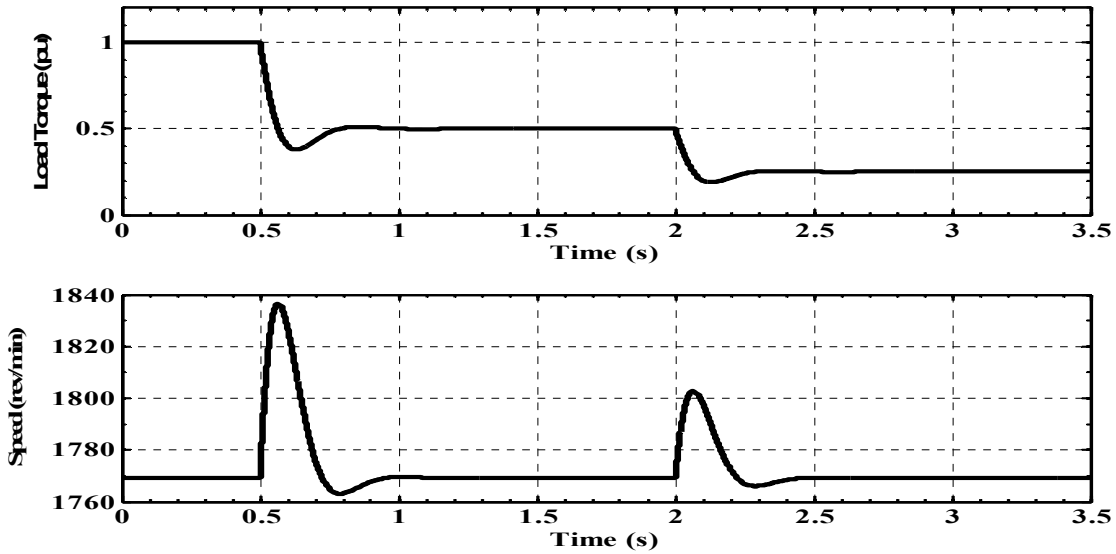


Figure 9. Electromagnetic torque and speed performances of the controlled three-phase induction motor

Given same load disturbances at the same simulation times as before, it can be observed that the motor speed rises sharply at the instant of disturbances but returns to its steady state rated value of 1770 rev/mins and maintains this value till end of simulation time despite the changes in load applied on the motor. Oscillations in load torque and speed have reduced significantly. There is, however, an overshoot in the value of motor speed following the disturbances. These results validate the prediction from the step response plot of Figure 8.

IV. CONCLUSION

The speed response of an un-controlled and controlled three phase induction motor was examined and discussed. The model of the induction motor was presented and its solution was clearly depicted in a way that is not common in literature. A PI controller was designed in the light of classical control theory and analytical expressions were obtained for controller parameters k_i and k_p. A time domain

closed loop response of the controlled motor was obtained and characterized. The speed of the uncontrolled motor changed when subjected to load disturbances whereas that of the controlled motor returned quickly to its initial value after similar disturbances. The results demonstrate the effectiveness and applicability of the controller design suggested in this paper. The controlled speed response will invariably lead to high efficiency and good precision in applications where it is deployed.

REFERENCES

[1] El-Zohri E. H. and Mosbah M. A. (2019) "Comparative Evaluation of PWM Techniques Used at Mega 328/p with PI Control for Inverter-Fed Induction Motor," *In Proceedings of the International Conference on Innovative Trends in Computer Engineering (ITCE)*, Aswan, Egypt, pp. 219-224.

[2] Zhang J. and Guo L., (2019) "Theory and Design of PID Controller for Nonlinear Uncertain Systems," in *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 643-648.

[3] Samad T, (2017) "A survey on industry impact and challenges thereof [technical activities]," IEEE Control Systems Magazine, vol. 37, no. 1, pp. 17–18.

[4] Ribeiro J. M. S., Santos M. F., Carmo M. J. and Silva M. F. (2017), "Comparison of PID controller tuning methods: analytical/classical techniques versus optimization algorithms," In Proceedings of the *18th International Carpathian Control Conference (ICCC)*, Sinaia, 2017, pp. 533-538.

[5] Bodson M. (2019) "Speed Control for Doubly Fed Induction Motors with and without Current Feedback," in *IEEE Transactions on Control Systems Technology*. Pp. 1-10

[6] Lim Y., Lee J. andLee K. (2019) "Advanced Speed Control for a Five-Leg Inverter Driving a Dual-Induction Motor System," in *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 707-716.

[7] Liu P., Gu H., Lü J. and Kang Y. (2018), "Global Pinning Synchronization with PI Controller in General Complex Directed Networks," In proceedings of the *37th Chinese Control Conference (CCC)*, Wuhan, China, pp. 24-29.

[8] Lei W., Li C. and Chen M. Z. Q. (2018), "Robust Adaptive Tracking Control for Quadrotors by Combining PI and Self-Tuning Regulator," in *IEEE Transactions on Control Systems Technology*. Pp. 1-9.

[9] Krause, P. C., Wasynczuk, O., Sudhoff, S. D. and Pekarek, S. (2013). *Analysis of electric machinery and drive systems*. U.S.A, John Wiley & Sons.

[10] Mohan, N. (2014). *Advanced electric drives: analysis, control, and modeling using /Simulink*. Hoboken, New Jersey, John wiley & sons

Usability and Accessibility Evaluation of Nigerian Mobile Network Operators' Websites

Solomon Adelowo Adepoju
Computer Science Departmemt
Federal University of Technology
Minna, Nigeria
solo.adepoju@futminna.edu.ng.

Adeniyi Oluwaseun Ojerinde
Computer Science Departmemt
Federal University of Technology
Minna, Nigeria
o.ojerinde@futminna.edu.ng

Ayobami Ekundayo
Computer Science Departmemt
Federal University of Technology
Minna, Nigeria
a.ekundayo@futminna.edu.ng

Rofiat Ahmed
Computer Science Departmemt
Federal University of Technology
Minna, Nigeria
rofiat.ahmed@gmail.com

***Abstract*— Usability and accessibility of websites have been a subject of great importance to human computer interaction researchers. Hence, this paper carries out the usability and accessibility evaluation of the four main Nigeria mobile network operators' websites. The websites tested are: www.mtnonline.com, www.gloworld.com, www.ng.airtel.com, and www.9mobile.com.ng. Four online automated tools were used which are: Mobile SEO, WAVE, TAW and Achecker. User testing was also conducted in a controlled environment with forty participants who were given five tasks to perform on the websites. The activities of the participants were recorded using Camtasia studio to get the total time taken to complete the task. Questionnaire for Interaction Satisfaction (QUIS) was used to get users feedback on the tested websites. The QUIS was designed based on five categories with each category dealing with an aspect of usability. The results show that the overall usability level of MTN was the best among the four. This was followed by GLO, 9mobile and Airtel respectively. From the results obtained, it is recommended that the web designers of the mobile network operators compare their website design with web content accessibility guideline (WCAG 2.0) to make sure the different categories of subscribers are satisfied with the features and services provided by the websites.**

Keywords— Accessibility, Usability, Websites, mobile network operator,

I. Introduction

Since the introduction of internet technologies, a new communication channel has been opened for citizens, business organisation and government. This has subsequently paved way for effective and efficient way of access to information and services [1]. Among emerging business organisation, telecommunication sector with focus on mobile services have emerged with a formidable and conspicuous online platform via websites. Through this platform, a lot of services are made available to various subscribers. This unparallel uptake of mobile services in Nigeria and Africa as a whole has no doubt had sizeable direct and indirect impact on the local economies [2].The importance of websites has been increasing with leaps and bounds over the years. Hence, it has penetrated every aspect of life and many business organisation have embraced its usage in dissemination of timely, accurate and up to date information to users.

This is to ensure seamless communication between subscribers and mobile network operators (MNO). In addition to this they intimate subscribers and website users with the avalanche of available products, offers, promos, callers tune, data plan, Frequently Asked Questions (FAQ), customer's feedback among others. So, a great level of interaction is expected between the users and the MNOs. This implies that the websites should not only be accessible but must also be usable.

For websites to be usable, the twin's quality of usability and accessibility must be met. It has been proved over the years that these elements are very crucial to the successful deployment of any website by any organization. Hence, tremendous research efforts have been channeled into evaluation websites based on usability and accessibility [3]–[5].

In Nigeria there are four major MNO which are MTN Nigeria communication Limited (MTN), Globacom limited (GLO), Airtel Nigeria (Airtel), and EMTS Limited (now 9mobile but formerly known as Etisalat). All these MNO use GSM and HSDPA technology. As at December, 2018 MTN has 67,133,009 million subscribers, GLO has 45,255,297million subscribers, Airtel has 44,180,484 million subscribers and 9mobile has 15,917,015 million subscribers [6]. So, these amount to over 172 million mobile subscribers as at 2018. Expectedly, some of these subscribers will have one reason or the other to carry out some tasks on the MNO websites. Hence, there is need for good usability and accessibility.

These MNO all have an online platform to provide services to their subscribers at any particular point in time. Thus, it provides easy means of accessing their services to their subscriber's. Presently the mobile network operator websites are being underutilized because most subscribers prefer to call customer care than to go online to solve their problem and not all their subscribers have access to the internet. MNO in Nigeria have websites so that their subscribers to get update of their services online, but they still make use of short message services (SMS) to send messages for their tariff plans updates, caller tunes among other services. The reason according to some subscribers is that the websites are not effective, efficiency and satisfactory enough. More so, they are not up to standard, most of the feature are not functional, and they are not well structured. Since, various services are being provided

by these MNO on the websites which users interact with them always. Hence the need arises to investigate the usability and accessibility levels of these websites.

There have been previous studies conducted in evaluating the usability as well as accessibility of different genre of websites like educational websites [7]-[8], e-government websites [9]–[11], e-commerce websites [12]-[13], e-learning websites [14], airline websites [15], hotel websites [16] and so on to mention a few. Till date there is no known previous study conducted to evaluate the usability and accessibility of Mobile network operators' websites. Hence, this paper tends to fill that gap with a view of showing the status of these important websites.

II. LITERATURE REVIEW

Usability has been described as a quality attribute which evaluates the ease with which user interface interfaces are used by different people. It refers to procedure to improve ease-of-use during the process of design [17]. The recent International Organization for Standardization (ISO) 9241-11 [18] defines usability as "usability can be defined as "the extent to which a product, service or system can be used by specified users to achieve a specified goals with effectiveness, efficiency, and satisfaction in a specified context of use"". Usability can also be seen as a capability to communicate, learn and use a service or product to accomplish the goal that it is meant to facilitate [19].

Web accessibility is concerned with the ability to design websites that can be used by different categories of people irrespective of their disability. It implies that people with disability can easily observe, comprehend, navigate and interact with websites and tools [20]. Accessibility involves a wide range of disabilities, including visual, auditory, physical, speech, cognitive, language, learning, and neurological disabilities. It is based on Web Content Accessibility Guidelines (WCAG) which is an internationally accepted standard that consists of fourteen guidelines that with provision to specify on how accessible website can be developed. Web Accessibility Initiative (WAI) which was founded by the World Wide Web Consortium (W3C) to promote the accessibility of the Web further defines web accessibility as a means by which people with disabilities can use the Web [19]. Both attributes are very important for websites to be effective.

Different attributes of usability have been viewed from various disciplines. Booth [21] listed four usability aspects namely usefulness, effectiveness, learnability, and attitude. Shackel [22] identified four usability evaluation criteria which focus on how users accomplish their tasks while using the system. These are learnability, flexibility, effectiveness, and user attitude. Jacob and Thomas [23] further proposed five attributes of usability as learnability, efficiency, memorability, low error rate (easy error recovery), and subjective satisfaction [19]. Brinck, Darren and Wood [24] definition of usability includes functionally correct, efficient to use, easy to learn and remember, error tolerant, and subjectivity. Jackob Nielsen, a famous usability expert, defined usability as comprises five components of learnability, efficiency, memorability, minimization of errors, and satisfaction [25].

According to Rinder [26] usability tests typically involves having a user performing a task in order to test the ease and efficiency of task completion. In addition to this, the user's successive satisfaction with their performance or the product is observed.

Ways by which usability can be conducted include the use of expert evaluators. These experts review a system based on a set of usability principles otherwise known as heuristics. This is done in order to check the conformance level with the heuristics and to identify potential usability problems [27].

According to Garcia and Diaz [28] usability and accessibility have become important to enable satisfactory online brand communication. Web navigation will facilitate the usability and accessibility of websites which improves the company's image and will favour loyalty towards the brand. Petrie and Kheir [29] are of the opinion that accessibility and usability are well confirmed concepts for user interfaces and websites. It is believed that accessibility and usability problems can be seen as two over lapping sets which can be classified into three. These are pure accessibility problems, pure usability problems and universal usability problems.

Some related studies in the field of usability and accessibility of websites are reviewed as follows. Mentes and Turan [30] assessed the usability of the website of university in Turkey based on attractiveness, controllability, helpfulness, efficiency and learnability. The study discovered that there exists a positive relationship between attractiveness, helpfulness, learnability, efficiency and usability perception of website but a negative relationship with controllability.

Sixteen e-commerce websites in Pakistan were evaluated based on usability and accessibility by [31].The study was based on conformance of these websites with Nielsen's guidelines principle as well as WCAG. Adepoju and Shehu [7] conducted a study on usability evaluation of academic websites in Nigeria using automated tools which focused on accessibility. Three Automated Tools were used namely Web Accessibility Checker, HERA and WAVE were used in the study. The automated tools inspected the conformity of the website with the WCAG standard. Results obtained from the study shows that there are lot of accessibility errors with the websites used in the study. Hence, they are not in total compliance with WCAG standard. Adepoju, Shehu and Bake [9] also carried out a study to know the level of accessibility conformance of e-Government websites in Nigeria, two online Automated tools were used to analyze the e-Governments websites. The results obtained further shows that there is non-conformity with WCAG 2.0 standard by the websites used in the study.

Another study was carried out in Malaysia which involves the use of three automated tools of website optimization, Axandra, and EvalAcess 2.0 by [32] . It was reported at the end of the study that there are numerous issues in Malaysia e-government websites usability and accessibility. Further study by Awlad and Yavuz[1] focused on Libya e-government websites evaluation based on usability and accessibility. The accessibility testing was done using two automated tools and it was reported that none of them passed the accessibility test. Thirty-two

OPERATOR	JUN 2018 (Q1.Q2)	SEP 2018 (Q2.Q3)	DEC 2018 (Q2.Q3)
MTN	66,448,706	64,160,404	67,133,009
GLO	40,108,508	40,856,649	45,255,297
AIRTEL	39,898,448	41,313,633	44,180,484
9MOBILE	15,811,684	15,355,061	15,917,015
TOTAL	***162,307,346***	***161,685,747***	***172,485,805***

Table 1: Quarterly subscriber operator data (source; NCC)

evaluators were involved in the usability testing and about one hundred and sixty-eight usability problems were discovered by them.

From the review above, various methods and approaches are been used by researchers to carry out website evaluation. While some conducted both accessibility and usability in their studies, others only focused on either of these.

III. MOBILE NETWORK OPERATORS

As stated earlier, in Nigeria there are four main mobile network operators which are MTN, GLO, Airtel and 9mobile.

MTN Group as a multinational mobile telecommunications company was founded in 1994. Its operation is in more than twenty-five African, European and Middle Eastern countries. MTN is from South Africa and its Headquarter is in Johannesburg. MTN have the Market Share of about 38%. It is the largest mobile network operator in Nigeria. Its website is www.mtnonline.com.

Globacom Limited (or GLO) was founded in 2003. It is a Nigerian multinational telecommunications company. GLO operates in four countries in West Africa: Nigeria, Republic of Benin, Ghana and Côte d'Ivoire. GLO network is from Nigeria and its headquarters is in Lagos. GLO have the Market Share of about 26%. Presently, it is the second largest network operator in Nigeria. Its website is www.gloworld.com

Airtel is an Indian multinational telecommunications services company. Its operation is present in twenty countries across South Asia & Africa. Airtel is from India and it headquarter is in New Delhi. Airtel has the market Share of 25.5% and the third largest mobile network in Nigeria. The website is www.airtel.ng.com.

9mobile (formerly Etisalat) is a UAE based telecommunications services provider, currently operating in fifteen countries across Asia, the Middle East and Africa. Etisalat is from United Arabs Emirates and their headquarters is in Abu Dhabi. Etisalat have the market Share 9%. The website is www.9mobile,ng.com.

Table 1 shows the quarterly data on mobile subscriber information for three quarters in 2018.

Various services are being provided by these mobile network operators on the websites and users interact with them always. Hence the need arises to investigate the usability and accessibility levels of these websites.

IV. METHODOLOGY

This research adopts three steps to carry out the usability and accessibility evaluation of Nigeria MNO websites. These methods which are stated below are further described in the sections following.

i. Online automated evaluation tools
ii. Usability testing
iii. Questionnaire-based method

D. Automated tools

This involves the use of online automated tools to test for the conformance of the selected websites for WCAG conformance. Once the user's login into websites, the URL of the websites will be entered into the available box and then the test will be activated by pressing the enter key or search (or any available) button. Four online automated tools were used for the evaluation; Achecker, WAVE, TAW, and mobile SEO.

Achecker: Accessibility Checker is an open source accessibility evaluation tool developed in 2009 by the Inclusive Design Research Centre (formerly known as the Adaptive Technology Resource Centre) of the University of Toronto Canada. Users can submit a web page via this tool, or can upload its HTML file. Various guidelines which users can select for evaluation are the HTML Validator, BITV, Section 508, Stanca Act, WCAG 1.0 and WCAG 2.0.

WAVE: This stands for Web Accessibility and Versatile Evaluator, and it is an automated tool developed by Web Accessibility in Mind (WebAIM). It is available online as well as a Firefox add-on. The accessibility violations report is presented by annotating a copy of the evaluated page and subsequently provides recommendations on how to repair them. Instead of providing a complex technical report, WAVE discloses the original Web page with embedded icons and indicators that reveal the accessibility information within the page. WAVE carries out evaluation based on WCAG.

TAW: Test de Accesibilidad Web is an automatic online tool for accessibility analysis. It was developed by the CTIC Centro Tecnólogico, TAW clearly marks the accessibility violations that it discovers by providing an annotated version of the web site as well as recommendations on how to resolve them. TAW categories the result of valuation into perceivable, operable, understandable and robust, other attributes includes warning, problems and not reviewed. It is available both online and as a desktop application as well as a Firefox add-on [33]

Mobile SEO: mobile SEO means "mobile search engine optimization" of websites combined with ability to view it flawless on mobile devices such as smartphones and tablet.an open source for accessibility evaluation tool which checks for Mobile friendly, Mobile speed, Google access and Page redirects.

E. Usability Testing

Usability testing involves carrying out a test with the real users in order to know the ease of use of the system, product or service been tested. The first part involves users being asked to complete representative tasks on the

websites. Typically, they are being observed by a moderator so as to know where problems are encountered and confusion experienced. Consequently, recommendations will be made to overcome these usability issues if many people encounter similar problems. However, it has been suggested that to get a comprehensive evaluation, automated tools should be combined with usability testing.

In the study forty users were used and the participants were asked to perform the following tasks on each of the websites.

1. Check the websites for latest device
2. Look for the cheapest tariff plan and data plan
3. Report for loss of SIM card
4. Chat with the online customer care service
5. Check for roaming inquiry

F. Questionnaire-based method

The last part involves the use of questionnaire to get feedback from the users based on their interaction with websites. Questionnaire is a direct way to getting feedback from end user, hence questionnaire must be simple to understand, straight forward and carry a good meaning to the evaluation purpose. This method involves the process of recruitment of users, distribute questionnaires to the participant and perform few tasks on the websites. The questionnaire adopted for the research is divided into three sections; section A demographic information of the participant, section B Questions for the usability studies, and the section C was used for general observation of the website. This research project adopts Questionnaire for Interaction Satisfaction (QUIS).

QUIS was developed by the University of Maryland designed in order to evaluate satisfaction of users' subjectively when interacting with some aspects of the human-computer interface. The current version QUIS 7.0 which is an update of previously validated QUIS 5.5. It contains a demographic section, and another section which measure the overall system satisfaction along six scales. Four measures of specific interface factors: screen factors, terminology and system feedback, learning factors, system capabilities.

The overall satisfaction with facet of the interface is measure by each area. Also, the factors that make up that facet based on a 9-point scale is measured. The questionnaire can be configured according to the needs of each interface analysis by including only the sections that are of interest to the user.

G. Test Environment

A computer laboratory was used in conducting the test sessions with four different laptops system. Camtasia studio software was used to record the session during the test.

i. HP Compaq CQ57 with intel® dual core™ CPU @ 2.12GHz processor 4.00GB RAM 500GB hard disk and windows 10 pro.
ii. Lenovo with CPU @ 2.16GHz processor 4.00GB RAM 500GB hard disk and windows 10.
iii. HP 15 r011dx with intel® inside™ Pentium™ CPU 2.14GHz processor 4GB RAM 750GB hard disk 64 Bits and windows 10.
iv. DELL Inspiron 5447 with Intel® core™ i5 CPU @ 2.4GHz processor 8GB RAM 64 Bits 1TB hard disk windows 8.1 touch screen and fast access facial recognition.

V. RESULTS AND DISCUSSION

H. Online automated tools results

Typically, Achecker could check for a whole website, single webpage online, offline html file or html code.

Table 2 shows results obtained from Achecker with MTN recording the largest numbers of problems while Airtel records the least. Some of the common problems are img element missing alt attribute, image used as anchor is missing, valid Alt text, input element type of "text", missing an associated label and input element, type of "text" has no text in label, label text is empty.

Table 3 gives the results obtained from TAW. MTN website recorded the highest of problems, while GLO website recorded highest numbers of warnings. The least number of problems and warning was recorded by Airtel website.

Table 4. shows the mobile SEO results all the websites perform well on mobile friendliness but GLO websites is the most mobile friendly. Also, the speed is for all the websites are not encouraging.

Table 5 shows WAVE results. Airtel returns no values as at the time the study was conducted. Some of the errors include empty link contains no text, a button is empty/, has no value text, missing form label. The alerts include heading level is skipped, non-script element is present, redundant link and missing first level heading.

Table 2: Results of Achecker

MNO	Known problems	Likely problems	Potential problems
MTN	114	0	550
GLO	13	1	640
Airtel	3	0	15
9mobile	50	0	477

Table 3: Results of TAW

	Attributes	MTN	GLO	Airtel	9mobile
Warnings	Perceivable	53	35	0	36
	Operable	67	14	2	43
	Understandable	12	6	0	12
	Robust	369	844	0	301
	Total	**501**	**899**	**2**	**392**
Problems	Perceivable	24	3	1	29
	Operable	27	10	1	17
	Understandable	3	2	1	2
	Robust	47	9	0	14
	Total	**101**	**24**	**3**	**62**
Not reviewed	Perceivable	4	4	4	4
	Operable	7	8	7	6
	Understandable	5	4	5	5

Item	Value
Gender:	
Male	29
Female	11
Age:	
16-20 years	8
21-25 years	27
26-30 years	4
41-45 years	1
Internet experience:	
Less than 1 year	2
2years- 4 years	6
Above 5 years	32

Table 6: Demographic data of the participants

Websites	Mean	Std. Deviation
MTN	6.42	2.05
GLO	6.17	2.84
Airtel	5.02	2.87
9mobile	6.13	2.23

Table 7: Overall website performance

Items	MTN	GLO	Airtel	9mobile
Errors	55	19	-	29	37
Alerts	4	27	-	12
Features	19	12	-	2
Robust	0	0	1	0
Total	**16**	**16**	**17**	**15**

Table 4: Results of Mobile SEO

Table 5: Results of WAVE

Fig 1: Users' overall reaction to the website

I. User Testing results

The usability test was conducted at a laboratory in School of Information and Communication Technology. A total of 40 participants were recruited randomly for the study with ten users per each mobile network website. The demographic data of the participants are shown in Table 6. experience of participants based on data collected, are as follows: 2 participants (5%) with less than 1 year, 9 participants (22.5%) with 2years-4years, 6 participants (15%) with 5years and 23 participants (57.5%) with above 5 years.

Result on the overall reactions of the users to the websites is shown in fig 1. This is based on user's reaction to how wonderful, easy, satisfying, stimulating and flexible the website is to them in the course of interacting with the websites. users find MTN websites most interesting in this regard and this is closely followed by GLO and 9Mobile websites. Airtel website did not provide good user satisfaction based on the results obtained.

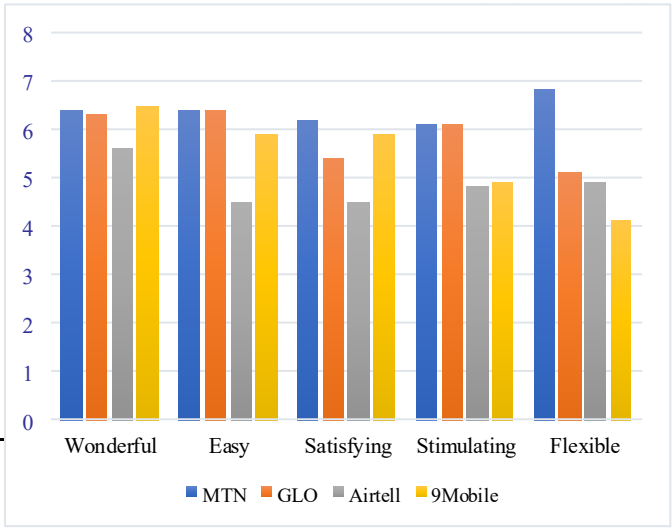
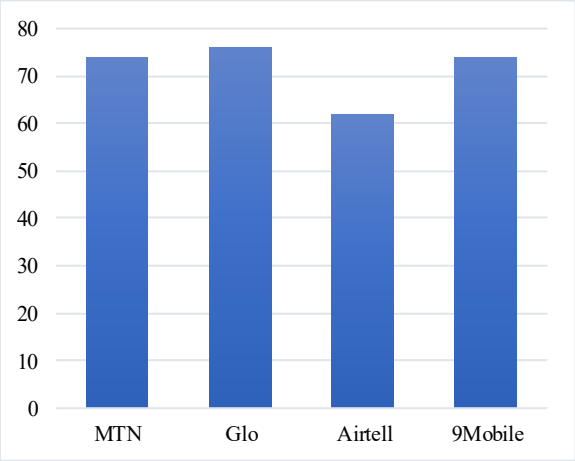


Fig 2 shows the task completion rate in percentage. This is based on the number of tasks each participant was able to perform across the websites. In all there are a total of five tasks. But an average of three tasks were able to be accomplished by most participants. GLO websites also take the lead in this regard with 76%. task completion rate. MTN and 9mobile have 74 % task completion rate while Airtel task completion rate is 62%.

Table 7 show the overall performance across the websites based on the analysis of the data collected from the questionnaire. Airtel has the lowest mean value (M = 5.02, SD = 2.87) and MTN has the highest mean value (M = 6.42, SD = 2.05). 9. GLO and 9mobile websites have an average mean and SD score of 6.17, 2.84 and 6.13, 2.23 respectively. The maximum available score is 9. So, none of the websites perform extremely excellent.

In conclusion, from the questionnaire result, the order of performance is MTN, GLO, 9mobile and Airtel.

Fig 2: Task completion rate



J. *Further Analysis Questionnaire Report*

According to some participants, there were complaints that Airtel website is not easily accessible, not user friendly, not flexible, not easy to navigate and difficult to access. Few others said it was easy use and accessed.

On 9mobile website, some users are of the opinion that it is user friendly and easy to operate while some said it was a bit confusing, no latest device (not available), not well organized, no comparison chat for cheapest tariff plan and data plan and roaming inquiries are not available.

As regards MTN website, a lot of participants said it was user friendly and that the Graphical User Interface (GUI) is fine.

Some users of GLO website are of the opinion that it is not user friendly and the content is not well organized while some believe otherwise

VI CONCLUSION AND RECOMMENDATIONS

Nigerian mobile network operators' websites are becoming more increasingly important for millions of subscribers daily. Hence, the need for the operators of these websites to develop not only usable but accessible websites which will make information easily available to people irrespective of their age, sex, ability and experience. Thus, this study has taken an in-depth usability and accessibility evaluation of the four MNO websites in Nigeria.

The evaluation methods used are: online automated tools, user testing and questionnaire-based method. The results obtained from user testing and questionnaire-based method showed that the overall usability level of MTN was the best among the four. Further results show that none of the website satisfied completely the WCAG 2.0 guidelines. Also, the overall satisfaction with the websites put MTN ahead of others, with Airtel website coming last. All the participant cannot perform all the selected five tasks on the websites and as a result the best task completion rate was achieved by GLO website with 76% satisfaction.

This research is very relevant to the MNO management as well as to the subscribers of various MNO that are accessing their websites. It is suggested that the MNO operators should improve their websites look and feel. This is necessary in order to provide subscribers with services which are both usable and accessible. This will further allow subscribers to have easy access to all information they need online instead of been limited to using only short message services (SMS). More so, it will also improve their relationship and enhance communication between the subscribers and the MNO.

It is therefore recommended that MNO website's designer should follow and incorporate the guidelines as stipulated in World Wide Web (WWW) and Web Content Accessibility Guidelines (WCAG) 2.0. This will ensure a design that is usable, accessible, maintainable, user friendly and enjoyable to all users. The websites should also be updated regularly for information like latest/cheapest tariff plan, latest device and cheapest data plan.

REFERENCES

[1] N. Awlad and K. Yavuz, "Usability and accessibility evaluation of Libyan government websites," *Univers. Access Inf. Soc.*, vol. 18, no. 1, pp. 207–216, 2019.

[2] E. Mutafungwa *et al.*, "A Collaborative Approach for Local Training on Contemporary Mobile Technologies in Eritrea," in *IEEE AFRICON*, 2017, pp. 811–816.

[3] E. Park, Y. Lim, and H. Lim, "A Study on Mobile Accessibility for Public Institution Mobile Websites in Korea and USA," *Adv. Sci. Technol. Lett.*, vol. 94, pp. 1–5, 2015.

[4] M. Blake, K. Majewicz, A. Tickner, and J. Lam, "Usability Analysis of the Big Ten Academic Alliance Geoportal : Findings and Recommendations for Improvement of the User Experience," *Code4Lib J.*, vol. 38, 2017.

[5] G. Rodríguez, J. Pérez, S. Cueva, and R. Torres, "A framework for improving web accessibility and usability of Open Course Ware sites," *Comput. Educ.*, vol. 109, pp. 197–215, 2017.

[6] Nigerian Communication Commision, "Subcriber statistics," *Subcriber statistics*, 2019. [Online]. Available: https://www.ncc.gov.ng/stakeholder/statistics-reports/subscriber-data#quarterly-subscriber-operator-data.

[7] S. A. Adepoju and I. S. Shehu, "Usability Evaluation of Academic Websites Using Automated Tools," in *3rd International Conference on User Science and Engineering (i-USEr)*, 2014, pp. 186–191.

[8] N. M. Rusli, S. Hassan, and N. E. Liau, "Usability Analysis of Students Information System in a Public University," *J. Emerg. Trends Eng. Appl. Sci.*, vol. 4, no. 6, pp. 806–810, 2013.

[9] S. A. Adepoju, I. S. Shehu, and P. Bake, "Accessibility Evaluation and Performance Analysis of e-Government Websites in Nigeria," *J. Adv. Inf. Technol.*, vol. 7, no. 1, pp. 49–53, 2016.

[10] P. D. D. Dominic, H. Jati, and G. Kannabiran,

"Performance evaluation on quality of Asian e-government websites – an AHP approach," *Int. J. Bus. Inf. Syst.*, vol. 6, no. 2, pp. 219–239, 2010.

[11] N. E. Youngblood and J. Mackiewicz, "A usability analysis of municipal government website home pages in Alabama," *Gov. Inf. Q.*, vol. 29, no. 4, pp. 582–588, 2012.

[12] L. Hasan, A. Morris, and S. Probets, "A comparison of usability evaluation methods for evaluating e-commerce websites," *Behav. Inf. Technol.*, vol. 31, no. 7, pp. 37–41, 2011.

[13] B. Sohrabi, P. Mahmoudian, and I. Raeesi, "A Framework for Improving e-Commerce Websites Usability Using A Hybrid Genetic Algorithm and Neural Network System," *Neural Comput. Appl.*, vol. 21, no. 5, pp. 1017–1029, 2012.

[14] D. Jain, R. Garg, and A. Bansal, "A Parameterized Selection and Evaluation Of E- Learning Websites Using Topsis Method," *Int. J. Res. Dev. Technol. Manag. Sci. –Kailash*, vol. 22, no. 3, pp. 12–26, 2015.

[15] S. Alwahaishi, V. Snášel, and A. Nehari-talet, "Website Evaluation an Empirical study of Arabian Gulf Airlines," *Int. J. Inf. Stud.*, vol. 1, no. 3, pp. 212–222, 2009.

[16] S. Hahn, B. Sparks, H. Wilkins, and X. Jin, "E-service Quality Management of a Hotel Website: A Scale and Implications for Management," *J. Hosp. Mark. Manag. ISSN*, vol. 26, no. 7, pp. 694–716, 2017.

[17] J. Nielsen, "Turn User Goals into Task Scenarios for Usability Testing," *Nielsen Normal Group*, 2014. .

[18] ISO, "Ergonomics of human-system interaction -- Part 11: Usability: Definitions and concepts," ISO, 2018.

[19] S. Joo, S. Lin, and K. Lu, "A Usability Evaluation Model for Academic Library Websites: Efficiency, Effectiveness and Learnability," *J. Libr. Inf. Stud.*, vol. 9, no. 2, pp. 11–26, 2011.

[20] WAI, "Introduction to Web Accessibility," 2006. .

[21] P. Booth, *An introduction to human-computer interaction*. Hove/East Sussex: Lawrence Erlbaum Associates, 1989.

[22] B. Shackel, *Usability-context, framework, definition, design and evaluation.* New York: Cambridge University Press, 1991.

[23] N. Jackob and K. L. Thomas, "A Mathematical Model of the Finding of Usability Problems," in *Human factors in computing systems*, 1993, pp. 206–213.

[24] T. Brinck, G. Darren, and S. Wood, *Usability for the Web: Designing Websites that work*. Elsevier, 2001.

[25] J. Nielsen, "Introduction to Usability," 2012. [Online]. Available: http://didattica.uniroma2.it/assets/uploads/corsi/143228/Nielsen_5_articles.doc.

[26] V. Peppa, S. Lysikatos, and G. Metaxas, "Human-computer interaction and usability testing : application adoption on B2C Websites," *Glob. J. Eng. Educ.*, vol. 14, no. 1, pp. 112–118, 2012.

[27] S. Ssemugabi and R. De Villiers, "Effectiveness of heuristic evaluation in usability evaluation of e-learning applications in higher education," *South African Comput. J.*, no. 24, pp. 26–39, 2010.

[28] M. G. García and A. C. Diaz, "Usable and accessible websites in SMEs. Challenges for the future," *Rev. Lat. Commun. Soc.*, vol. 13, no. 65, pp. 392–409, 2010.

[29] H. Petrie and O. Kheir, "The Relationship between Accessibility and Usability of Websites," in *SIGCHI conference on Human factors in computing systems*, 2007, pp. 397–406.

[30] A. Mentes and A. H. Turan, "Assessing The Usability of University Website: An Empirical Study on Namik Kemal University," *Turkish Online J. Educ. Technol.*, vol. 11, no. 3, pp. 61–69, 2012.

[31] M. Rahman, S. Komal, A. Khan, and T. Qamar, "Usability and Accessibility Evaluation of Pakistan's E-commerce Sites," *ARPN J. Sci. Technol.*, vol. 5, no. 9, pp. 457–475, 2015.

[32] W. A. Rahim, W. M. Isa, M. R. Suhami, N. I. Safie, and S. S. Semsudin, "Assessing the Usability and Accessibility of Malaysia E-Government Website," *Am. J. Econ. Bus. Adm. 3*, vol. 3, no. 1, pp. 40–46, 2011.

[33] R. Dagfinn and D. Svanæs, "Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users," *Univers. Access Inf. Soc.*, vol. 11, no. 4, pp. 375–385, 2011.

LWT-CLAHE BASED COLOR IMAGE ENHANCEMENT TECHNIQUE: AN IMPROVED DESIGN

O. M. Gazal, A. E. Adedokun, I. J. Umoh, M.O Momoh
Department of Computer Engineering
Ahmadu Bello University, Nigeria
bintghazall@gmail.com, wale@abu.edu.ng, ime.umoh@gmail.com, momuyideen@abu.edu.ng

Abstract-color Image enhancement is one of important process and actually a vital precursory stage to other stages in the field of digital image processing. This is due to the fact that the effectiveness of processes in this stage on the output determines the success of other stages for a quality overall performance. This paper presents a color image enhancement technique using lifting wavelet transform (LWT) and contrast limited adaptive histogram equalization (CLAHE) to overcome the issue of noise amplification, over and under-enhancement in exiting enhancement techniques. Test images from Computer Vision Database were used for the proposed technique and the performance was evaluated using PSNR and SSIM. Result obtained shows an average improvement of 56.4% and 20.98% in terms of PSNR and SSIM respectively.

Keywords – color Image, Enhancement, LWT, CLAHE, HSV, PSNR, SSIM.

I INTRODUCTION

Image enhancement can be defined as the process of manipulating an input digital image such that the resulting output is more suitable for a specific task[1]. The task may be for more appealing visual appearance or to prepare it to be suitable for further analysis. Image enhancement techniques are commonly used in areas such as satellite imaging, Medical Imaging (Computerized Tomography (CT) Scans and Magnetic Resonance Imaging (MRI)), Remote observation and sensing, Subaquatic Imaging, Forensic, Digital Camera Applications, among others[2].

Enhancement of image can be done both in the spatial and frequency domain. Enhancement techniques in the spatial domain involves direct manipulation of pixel values based on their topological arrangement and neighborhood this includes global histogram equalization and its different variance and contrast stretching. Spatial domain techniques are usually characterized by inherent noise amplification, over- or under-enhancement, color distortion [3] and sometimes unnatural appearance of the resulting output[4]. However, frequency domain techniques are able to avoid some of these limitations due to their characteristics to distinguish image pixel into smooth and detail region of the image [5] and this can then be enhanced accordingly. There are different techniques used in the frequency domain, this includes discrete cosine transform (DCT), Dual Tree-Complex wavelet (DT-CW), Discrete wavelet Transform (DWT), Fast Fourier Transform (FFT), Lifting wavelet Transform (LWT) nonetheless these methods may cause blurriness in the resulting output image and noise emphasis towards the edges[6].

II RELATED WORKS

There has been several techniques and combination of techniques used for image enhancement but most of it suffers from either over or under-enhancement and most especially noise amplification which prevent the enhanced image to sufficiently perform the intended task. In [7] the authors introduced an image enhancement method for satellite low contrast color image using Discrete Wavelet Transform (DWT) to decompose the input image into four different sub bands namely Low-Low (LL), Low-High (LH), High-Low(HL) and High-High(HH). This was done on each of the three-color spectrum of the input low contrast image (ie

RED-, Blue- and GREEN-spectrum of the RGB) separately. The LL sub band of each of the color specs is then decomposed into series of binary levels, each of which was processed separately. The method was able to remove blurriness in the image however, the resulting image details was not effectively highlighted and noise amplification was observed towards the edges.

Authors in [3] combined spatial- and frequency-domain technique to eliminate the contrast over-stretching and noise amplification problem of spatial domain technique by enhancing only the low frequency component of the image by CLAHE previously split by DWT and introduced a weighted average of the original and enhanced image using a predefined weighting factor to adaptively control the enhancement level of regions with different luminance to get the final output. The technique was tested only on grayscale images although it succeeded in bring out details in a low contrast image, however the details were not efficiently enhanced as some details in the image are lost to blurriness and suffers from noise amplification.

Satellite image enhancement using single value decomposition was proposed in [8], the image was first equalized using GHE and simultaneously decomposed into four sub-bands using DWT, the equalized image was later decomposed also. The singular value matrix was estimated for the two LL sub-bands (the equalized and non-equalized and the SVD values are used to calculate the new LL sub-band, this new LL is added to the equalized high frequency components and the inverse is taken to reconstruct the enhanced image. However, the resulting output suffers from noise amplification.

A color image enhancement technique was proposed in [9] using Laplacian filter and CLAHE. They converted the input image to HSV and the S and V components are filtered using Laplacian, output from this was used to calculate local correlation, variance and luminance to get an enhanced V component which is further enhanced using CLAHE, the Laplacian Filtered S component was further enhanced by stretching using a fixed stretching factor value of 0.77, the final image is then obtain by converting the combination of the H, enhanced S and V components back to RGB.

[10] Proposed an enhancement technique which decomposes the input image into low and high frequency using LWT and enhance the low frequency using CLAHE whilst keeping the high frequency unchanged. The enhanced image obtained is then added to the original image using a weighting factor matrix and brightness compensation factor to control over-enhancement and compensate for the brightness loss respectively. The technique was tested on some test images and several performance metrics like the AMBE, Average information content, contrast improvement index, degree of entropy unpreserved, SSIM and universal quality index was used to evaluate the result obtained and compared with some exiting technique like GHE, BBHE, DSIHE, MMBEBHE, DQHEPL, BHEPLD, NPMHE and CLAHE. The result obtained was observed to outperform all of these techniques on most of the metrics except the CLAHE with which it has comparable results. Although this technique is robust considering it high performance in all the metrics used however, it only focuses on the image contrast and no consideration was given to image details which needs special focus to efficiently enhance an image. Also, the weighting factor process will further add noise to the enhanced image.

To efficiently enhance an image the improved design LWT-CLAHE based technique is proposed where different image components are distinct and enhanced independently to avoid over-enhanced and effectively eliminate inherent noise.

III PROPOSED TECHNIQUE

The proposed technique of color image enhancement using LWT and CLAHE is made up three main stages namely; the pre-processing stage where image where separation of image into color, color intensity and luminance using HSV model and noise removal from the S and V occurs, secondly the detail enhancement stage where the V component is converted to transform domain using LWT and the low and high-frequency is enhanced using CLAHE and developed best-out filter and the last stage involves the contrast enhancement where a stretching

factor is calculated and applied on the S component. Figure 1 shows the flow chart of the proposed technique

A Pre-processing Stage

The input image used in the experiment are standard test images obtained from the computer Vision Database [11]. The input image was converted from its original color format of RGB to HSV so as to effectively separate the color, color intensity and luminance content of the image before enhancement. The S and V component are filtered using median filter to reduce the inherent noise the image must have acquired during acquisition. This de-noise components then serves as input to the succeeding stages. The formula for converting RGB to HSV is given in equations (1 to 5).

Value (V) =Value = max(R, G, B) (1)

Det = Value – min(R, G, B) (2)

Saturation = Det /Value (3)

If R = Value, Hue(H) = 1/6((G – B)/Det) (4)

G=Value, Hue(H)=1/6 (2 +(B – R)/Det) (5)

B=Value, Hue(H)=1/6 (4 + (R – G)/Det) (6)

Where R, G, B is the Red-, Green- and Blue-channel respectively.

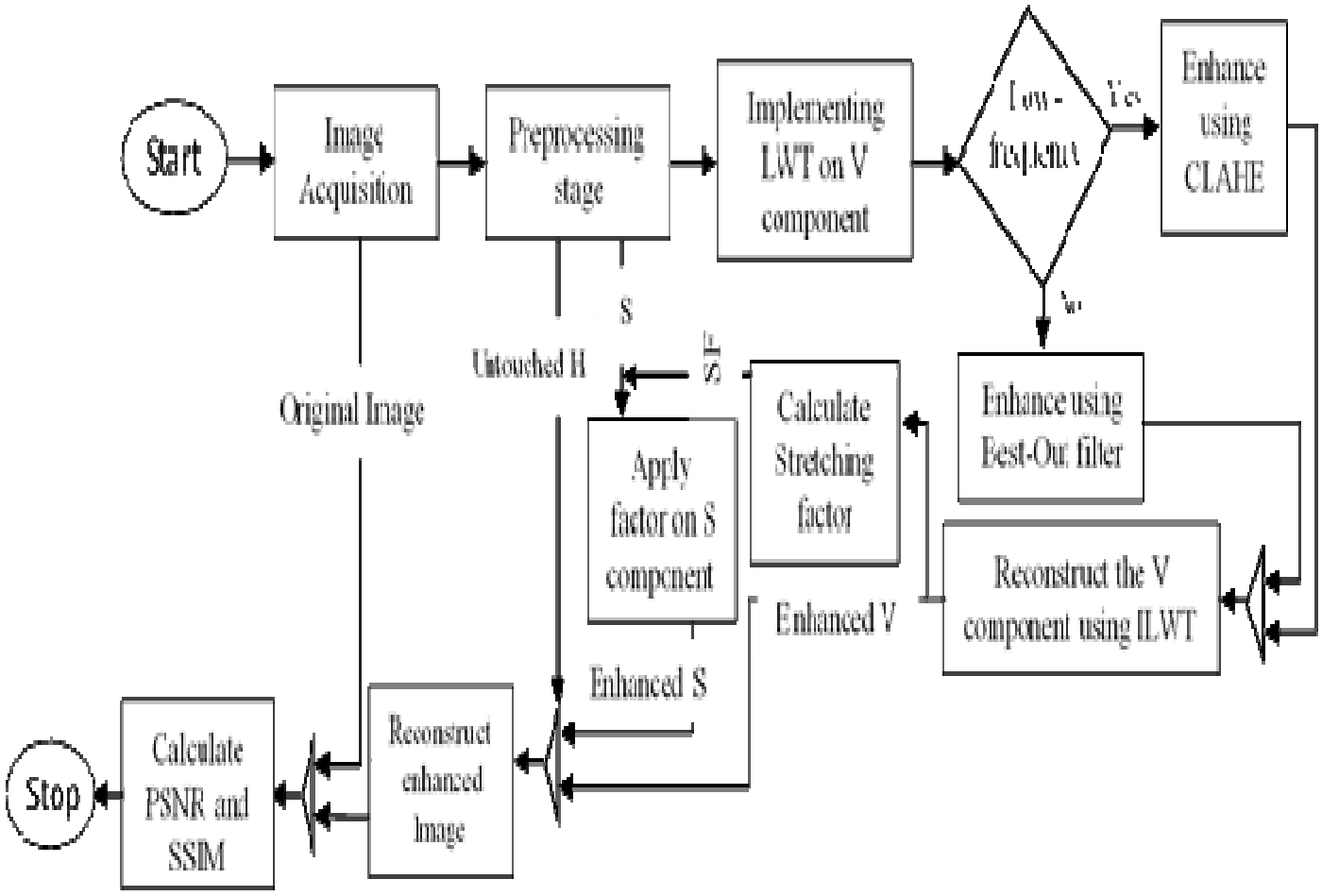


Figure 1: Flow Chart of the proposed Technique

B *Image Details Enhancement*

Structures that make up the outline of details in an image are mostly presented in the luminance (V) component of the HSV model. In order to make the details in the image more pronounced and highlighted the V component has to be effectively enhanced. The LWT, CLAHE and Best-out filter is employed to enhance the image details.

1. *LWT-CLAHE Implementation*

The de-noised V component is transformed into transform domain using LWT. The LWT computational complexity is more effective and speed is faster[12] unlike other transform or frequency domain techniques in enhancement which suffer from computational complexity and expenses(memory)[13], due to the issue of optimal parameters selection [14]. LWT transform parameters (coefficients) are integers which eliminate quantization error that arise from other wavelet transforms.

The Haar wavelet was adopted for the lifting scheme because of its simplicity and faster computation[15]. The LWT-Haar is used to split the transformed V component into low- and high frequency. The low frequency is enhanced using CLAHE, this is adopted because of its efficiency in enhancing local details in an image. The high-frequency is subjected to the best-out filter. The image details enhancement procedure is given as follows;

Step 1: Decomposed the de-noised V component by 2-level LWT using haar wavelet into low- and high-frequency components

Step 2: Enhance the low-frequency component using CLAHE

Step 3: Enhance the high-frequency component using the developed Best-out filter.

Step 4: Inverse transform the enhanced low- and high-frequency using ILWT to obtain the enhanced V component (V_{enh}).

2. *Contrast Limited Adaptive Histogram Equalization*

Since its introduction in 2013, CLAHE has proven to be an effective image local details enhancement technique as stated earlier. There are two important parameters that needs to be defined when using CLAHE. The parameters are the block size to choose the contextual region to avoid over or under-enhancement and clip limit to reduce noise amplification. The whole image is divided into sub-image of non-overlapping contextual region based on the block size while over enhancement is controlled using the clip limit. In this paper a block size of 3X3 to ensure close region equalization and clip limit of 0.01 (to reduce amplification of noise to the barest minimum) is predefined. The procedural step in implementing CLAHE on the low-frequency component of V is given in Figure 2;

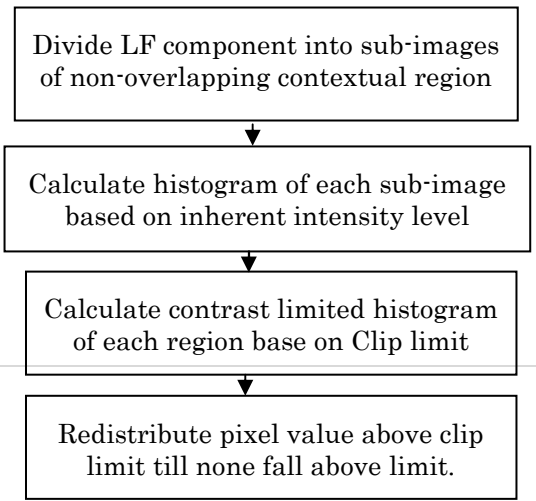


Figure 2: Block Diagram of CLAHE procedure

3 *Best-out filter*

Due to the sensitivity of noise in the image which in most cases is more pronounced in the detail part of the image (high frequency), a filter is developed to manipulate the pixel values but only keep the changes only if the new value is better than the original.

This filter circumvents amplification of noise in the noisy parts and enable enhancement in the less noisy or noise-free part. The images considered here are assumed to be 8-bits and the highest intensity level considered is 255. The procedure of the best-out filter is highlighted as follows;

Step 1: Import the high-frequency (HF) component of the transformed V component

Step 2: Find the size of HF component in rows and column, store as r and c

Step 3: Calculate probability density function (pdf) of each value of the HF component and store in an array.

Step 4: Calculate cumulative density function (cdf) of each of the PDF value from step 3

Step 5: Calculate modified CDF by multiplying each CDF value from step 4 by the highest intensity level in the image (256), add 0.5 and round up to the nearest integer

Step 6: Perform conventional Histogram equalization on the values from step 5

Step 7: Compare each of the original image pixel values with the results from step 6 and keep the best of the two.

C *Contrast enhancement*

The enhanced V component is used to calculate the stretching factor to be applied on the contrast of the image. This is calculated adaptively (to eliminate the problem of over- or under-stretching) by histogram analysis of V and introducing a weighted function. The stretching factor (sf) formula is given by below;

$$stretching\ factor = 1 - cdf_w \quad (7)$$

Where cdf_w is the weighted cdf from the histogram analysis which is dependent on the pdf_w by the equation i and ii;

$$cdf_w(i) = \sum_{i=0}^{i_{max}} \left(\frac{pdf_w(i)}{\sum pdf_w}\right) \quad (8)$$

$$pdf_w(i) = pdf_{max}\left(\frac{pdf(i) - pdf_{min}}{pdf_{max} - pdf_{min}}\right)^{\alpha} \quad (9)$$

Where $pdf(i)$ is the weighted probability distribution function of intensity level i, pdf_{max} and pdf_{min} is the maximum and minimum pdf of intensity level of the image pixel array respectively, $pdf(i)$ is pdf of the pixel intensity of interest i and α is weighted distribution function adjustment parameter which is assumed to be 0.75.

The enhanced contrast is calculated using the stretching factor as given in equation below;

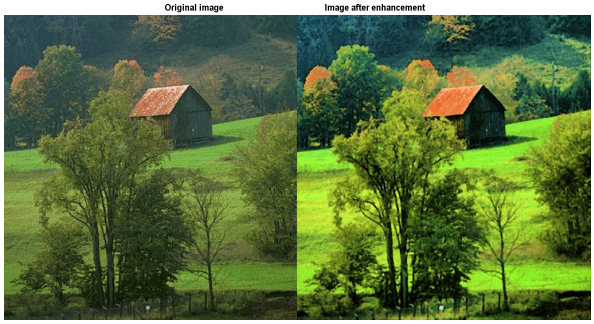
$$S_{enh} = S_d^{stretching\ factor} \quad (10)$$

Where S_{enh} is the enhanced S and S_d is denoised S.

The enhanced V (V_{enh}) and S (S_{enh}) components are added to the unaltered H and converted back to RGB to get the final enhanced image.

IV EXPERIMENTAL RESULTS AND DISCUSSION

The proposed algorithm was implemented in MATLAB 2018a and executed on a computer with Intel® Core(TM) i5-7500 CPU @3.40GHz processor with a 8Gb RAM 64-bits Operating system. Standard color test mages used for the developed algorithm were obtained from Computer Vision Database and the results obtained are evaluated using Peak Signal-to-Noise Ratio (PSNR) and structural similarity index (SSIM) given by the following formula.



$$PSNR = 10\log_{10}\left(\frac{\max Intensity^2}{MSE}\right) \quad (11)$$

Where *maxIntensity* is the maximum intensity level in the image which is 255 for a 8-bits image and MSE is the mean square error given as;

$$MSE = \frac{\sum_{r,c}(I_o(r,c) - I_i(r,c))^2}{rc} \quad (12)$$

Where r, c is the number row and column in the image respectively, $I_o(r,c)$ is the output image and $I_i(r,c)$ is the input image.

Figure 3 and 4 shows two of the test images (Barnfall and Bodie (512 X 512 pixels)) obtained from the Computer Vision Database and the resulting outputs after enhancement using the proposed algorithm.

Figure 3: Original image versus Enhanced image (Barnfall)

Figure 4: Original image versus Enhanced image (Bodie)

In Figure 3 the effective image details enhancement can be observed from the variation of color intensity of green grasses in different area, the roofing sheet rusty color and clearer presentation of details of wire fence at the bottom of the image this can also be observed in Figure 4 on clearer appearance of details of the whole plank house frame, the dimension of the planks use for support of the window by the side and blue color intensity of the sky.

Generally, it can be observed that both the image details as well as the intensity of colors in the image were effectively enhanced using the proposed algorithm and natural visual appearance is retained. It can also be observed that the inherent noise perceived in the original image is less conspicuous in the enhanced image.

The result obtained for PSNR and SSIM using different test images from Computer Vision Database is compared with the work of [9] where the authors used Laplacian filter to enhance V component and fixed value 0.77 of gamma for contrast stretching unlike the proposed algorithm where the contrast of the image is adaptively stretched based on the image luminance composition. Table 1 and 2 presents the PSNR and SSIM results respectively. From table 1 and 2, it can be observed that the images are enhanced better than the Laplacian-CLAHE algorithm and also image structure are better preserved.

Sn	Image Name	PSNR of proposed algorithm (dB)	PSNR of Laplacian +CLAHE (dB)	Improvement (%)
1	Barnfall	76.2062	29.3227	61.52
2	Bodie	66.1031	30.7823	53.43
3	Butterfly	66.4998	28.3227	57.41
4	Clinmill	72.1240	29.5946	58.97
5	Flower	84.8102	29.6242	65.07
6	Tulips	57.6032	27.5110	52.24
7	Voit	57.4301	28.3812	50.58
8	Peppers	62.4275	29.9763	51.98

Table 1: Comparison of PSNR values of test images

Table 2: Comparison of SSIM values of test images

sn	Image Name	SSIM of proposed algorithm (dB)	SSIM of Laplacian +CLAHE (dB)	improvement (%)
1	Barnfall	0.6628	0.5219	21.26
2	Bodie	0.8164	0.5677	30.46
3	Butterfly	0.7840	0.6636	15.36
4	Clinmill	0.6690	0.6311	5.67
5	Flower	0.8408	0.5466	34.99
6	Tulips	0.7105	0.4976	29.96
7	Voit	0.8519	0.7245	14.95
8	Peppers	0.8558	0.7261	15.15

V. CONCLUSION

This paper presented an improved design of LWT-CLAHE image enhancement technique. The technique enhances image details and contrast while reducing noise amplification and retain its natural appearance to overcome issue of under and over enhancement associated with conventional techniques. The original color image is converted to HSV, the luminance (V) component was decomposed to transform domain to separate the low- and high frequency using LWT and enhanced independently. The low-frequency is then enhanced using CLAHE and high-frequency using Best-out filter. The contrast or color intensity is adaptively enhanced by a stretching factor calculated using the enhanced luminance. The enhanced luminance and contrast are then combined with the unaltered color content (Hue) and converted from HSV to RGB to obtain the final output. The result of proposed algorithm is evaluated using PSNR and SSIM.

REFERENCES

[1] R. Dorothy, J. R M, J. Rathish, S. Prabha, S. Rajendran, and S. Joseph, "Image enhancement by Histogram equalizationNo Title," *Int. J. Nano Corros. Sci. Eng.*, pp. 21–30, 2015.

[2] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd Editio., no. September. Pearson Prentice Hall Pearson Education, Inc. Upper Saddle River, New Jersey 07458, 2008.

[3] Z. Wei, W. Jun, H. Lidong, and S. Zebin, "Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement," *IET Image Process.*, vol. 9, no. 10, pp. 908–915, 2015.

[4] R. Mishra and U. Sharma, "Review of Image Enhancement Technique," *Int. J. Eng. Technol.*, vol. 2, no. 8, 2013.

[5] J. Cherian, Mereena, and A. Thomas, "A Survey on DWT and LWT based Digital Image Watermarking," *Int. J. Res. Comput. Appl. Inf. Technol.*, vol. 4, no. 3, pp. 27–32, 2016.

[6] R. Dhanda and k k Paliwal, "Hybrid method for iamge watermarking using 2 level LWT-Walsh Transform-SVD in YCbCr Color Space," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 5, no. 11, 2017.

[7] R. Thriveni and Ramashri, "Satellite Image Enhancement Using Discrete Wavelet Transform and Threshold Decomposition Driven Morphological Filter," *IEEE*, pp. 4–7, 2013.

[8] K. Sugamya, S. Pabboju, and A. Vinayababu, "Image enhancement using singular value decomposition," *Int. Conf. Res. Adv. Integr. Navig. Syst. RAINS 2016*, pp. 0–4, 2016.

[9] S. Bhairannawar, A. Patil, A. Janmane, and M. Huilgol, "Color image enhancement using Laplacian filter and contrast limited adaptive histogram equalization," *2017 Innov. Power Adv. Comput. Technol. i-PACT 2017*, vol. 2017–Janua, pp. 1–5, 2018.

[10] M. Goyal, B. Bhushan, S. Gupta, and R. Chawla, "Contrast Enhancement Technique Based on Lifting Wavelet Transform," *3D Res. Center, Kwangwoon Univ. Springer-Verlag GmbH Ger. part Springer Nat. 2018*, vol. 9, no. 50, 2018.

[11] "http://decsai.ugr.es/cvg/dbimags/." .

[12] A. Tun and Y. Thein, "Digital Image Watermarking Scheme Based on LWT and DCT," *Int. J. Eng. Technol.*, vol. 5, no. 2, 2013.

[13] A. V. Kumari, "A Review of Image Enhancement Techniques," *Int. J. Trend Res. Dev.*, 2015.

[14] J. Xia, K. Panetta, and S. Agaian, "Color Image Enhancement Algorithms Based on the DCT domain," *J. Electron. Imaging*, vol. 21, no. 1117, 2012.

[15] D. Srikanth and M. Mittal, "Implementation of Haar Wavelet through Lifting For Denoising," *Int. J. Eng. Res. Appl.*, vol. Vol. 3, no. Issue 5, p. pp.1311-1314.

Impact of Household Construction Materials on Wi-Fi (2.4 GHz) Signal

M. Ahmed, R. A. Buhari and A. Musa
Centre for Telecommunication Research and Development, Kwara State University, Malete, Nigeria
twhid2001@yahoo.com

***Abstract*—Different construction materials cause significant losses when subjected to wireless signals along propagation path. A change in the signal strength, losses and user experience occur as a result of the complex composition of the materials. This paper investigated Wi-Fi (2.4GHz) signal when subjected to an enclosure dimension (0.24mx0.20mx0.16m) made of three timber species namely: *Chlorophoral excels (Iroko), Afzelia Africans (Apa) and Gosswei Lerodendrin balsamic ferum (Agba)*; a reflective glass and solid reinforced concrete. Received Signal Strength (RSS), Upload and Download speed (Mbps) and Jitter (ms) have been considered. The steps taken involve finding the penetration loss from the RSS with their respective data representation and drawing fact from the represented data. The results show that material C has 35.4% increase in indoor signal strength, material E has 8.8% decrease in indoor signal strength, material A, B and D has 63.7%, 46.3% and 71.2% decrease respectively in indoor received signal strength. The result from QoS parameter (Jitter) are compared for both outdoor and indoor. Relatively, it shows that Material A and Material E will provide improved experience to the user.**

Keywords—RSS; Jitter; Penetration Loss; Upload speed; Download speed; Packet Loss.

I. INTRODUCTION

The need to measure wireless signal parameters is important in every enterprise to determine the type of activities to be carried out and the quality of service obtained from the network. It is very rampant in buildings whereby the mobile users tend to change position from one point to another in other to have a reliable wireless communication.

However, the density, coverage, signal power and other related metrics are to be known in different environmental conditions.

The quality of wireless signal is reduced when they are obstructed by different materials along the path of propagation. Researchers have worked on radio waves propagation in building and statistical results were obtained from their measurements [1]- [[2].

Wireless signals generated by Radio Frequency (RF) circuit or device are affected by various properties, which could be categorized into intrinsic or extrinsic properties. The intrinsic properties refer to the factors that affect the wireless signals internally within the sources where they are generated. This scenario is commonly affected by RF devices, such as Router and RF Antenna operating within 900/1800MHz, 2100MHz, 2400MHz. Examples include short and thermal noise etc. The extrinsic properties refer to the factors that tend to distort wireless signal externally relative to their environment immediately after they are directed away from the sources.

There is little awareness in the construction industry regarding the impact of wireless coverage when it is subjected to different construction materials. A change in building materials to meet the requirement for building regulation needs to be known, such that it can improve the efficiency of the signals. Some materials can help to improve the thermal conductivity properties of houses but can affect the transmission of wireless signals into and within houses. In the future, construction practice and materials may change particularly in response to improve signal transmission within them [3].

The loss experienced by radio waves when subjected to different materials depends on the frequency and angle of incidence of the waves as well as the materials properties. An approved recommendation by ITU-R (P.2014) defines the basic quantities related to the electrical properties of a building materials and penetration loss which describe the effect of the material structure, electrical properties and building loss measurement on radio propagation [3]. It is therefore necessary to differentiate materials such as bricks, wood and glass when they are subjected to wireless signals.

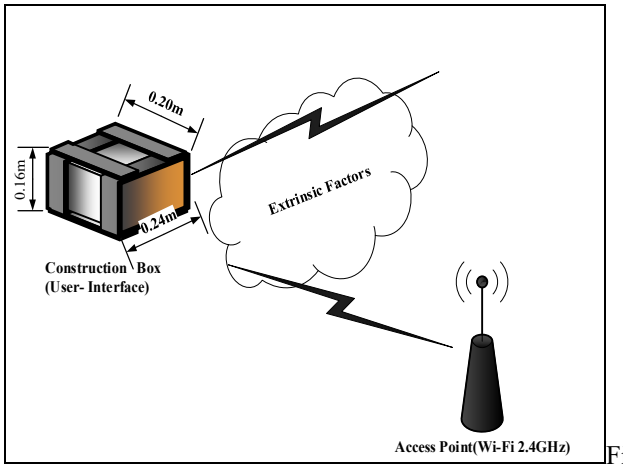


Fig. 1. Illustration of Research Study.

Fig. 1 gives the physical model of the study, showing the access point Wi-Fi for wireless communication. The enclosure is subjected to the Wi-Fi signal and predictive measurement is required for both indoor and outdoor scenarios.

The coverage of RF signal is further decreased by the impact caused by various household materials. The reduction in signal strength is indicated in Table I for various materials.

Table I. Radio Coverage Reduction Loss [4].

S/N	Materials	Range of Reduction of loss
1	Wood, plaster, glass uncoated, without metal	0 - 10 %
2	Brick, Press Board	5 - 35 %
3	Ferro concrete	10 - 90 %
4	Metal, aluminum lining	15-95%

Customers or end-users demand good quality of service (QOS) and coverage. The penetration loss caused by different household construction materials affect signal strength received by the end users. The penetration loss is significant while considering the overall losses in wireless communication [5].

The survey metrics in this project introduce the various parameters to be measured using optimum software and hardware device to detect and measure the metrics that are used as judging criteria to evaluate the Quality of Service (QOS) experience by the end users for various household construction materials as wireless signal propagate through them.

User experience on a given network service differs based on their environment and the medium through which the wireless signal has travelled. The absorption rate of different household construction materials through which wireless signals travel differs from one another. The building materials considered in this project are non-magnetic and dielectric but behave like a lossy dielectric material since they allow signal to penetrate through them.

The unrealistic propagation of the wireless signal when it passes through various medium which is caused by multipath has made it penetrate through various construction materials having different characteristics which account for significant penetration loss. The metrics of focus in this work therefore include Received Signal Strength, Penetration loss, download speed, upload speed, Packet Loss and Jitter.

This work presents the impact of construction materials on Wi-Fi signal by evaluating the QoS in different household materials. The related works have been presented in Section II. Section III compares 2.4GHz and 5GHz, Section IV discusses the methodology, while the results are discussed in Section V. Conclusion is drawn in Section VI.

II. RELATED WORKS

The work in [6], presented the method used to determine the Global System for Mobile Communication (GSM) signal provided by four service in Port Harcourt. Focus was on the penetration loss in five selected buildings constructed with different materials. The measured data was obtained using a Techno Tablet installed with Radio Frequency Signal Tracker (RFST) software. The surveyed parameters were signal strength, and distance, from which the penetration losses were obtained. The data were evaluated using Least Square Line Analysis to obtain the line that best fits the curve. The result obtained showed that the building with alucoboard wall cladding has the highest signal penetration loss while the sandcrete building/unrusted corrugated iron sheet roof has the lowest signal penetration loss.

The data obtained from the services provider were reliable and correlative, since they were taken twice a day in one month. However, the Signal to Noise Ratio (SNR) was not considered. This could be a bottle neck for wireless signal in the environment where external sources contribute majorly to signal distortion which could be referred to as extrinsic factor.

The work in [7] examined the penetration loss of doors and windows inside residence using Integrated Service Digital Network Broadcast-Terrestrial (ISDB-T) Television signal operating at 677MHz frequency. It demonstrated the signal parameters such as received power (Indoor and Outdoor), signal to noise ratio (SNR), field strength with the aid of spectrum analyzer which was connected to Ultrahigh frequency (UHF) antenna known as Rabbit Antenna, directly placed at the doors and windows made of different materials. The indoor signal is measured and compared with the outdoor signal. The discrepancy was used as the penetration loss, as a result of the material characteristics which causes signal degradation. The result obtained from the average penetration loss for each house was computed and compared using commutative distribution function (CDF) in a probability distribution curve which shows that average penetration loss (APL) decreases from heavy to light materials.

More concerned was made on interference by using a desk set spectrum analyzer which aid visualization of operating signal in frequency and time domain. However, as a result of low power level in the device, it might not display accurate result of the surveyed data.

[8] analyzed and evaluated the performance of a radio signal in four different large buildings to give more understanding of what to expect from radio propagation environment in a worst-case scenario. The method used was to provide large set of data, which described attenuation variability of radio signals in various building types in the public safety frequency band. The operating frequency of the signal was 750MHz, in which the measurement was done with three different types of signal measuring devices (Spectrum Analyzer Radio Mapping, Narrow band communication Receiver Radio Mapping and Broadband Synthetic-Pulse measurement) to obtain the RSS, Centre frequency and Root Mean Square (RMS) delay spread of signal power.

The result obtained showed that the present of short pulses in the time-domain waveform was as a result of multipath in a given environment. The RMS delay spread value was used to determine the time it will take the multipath reflection to decay below the threshold level. The median and standard deviation

value were critically observed in most of the measurement in both spectrum analyzer and the receiving system. The result further illustrated that; the RMS delay spread value for the measurement made in the large open floor building were typically two to five times greater than the measurement obtained from the building with narrows corridors.

The use of the three set of measuring devices makes the result correct with little or no discrepancies. However, the RMS spread value was affected greatly by the direction of the antenna and the penetration loss. Therefore, radio signal analysis should not only be focus on the signal but also on the environment and material characteristics which could be classified as either lossy or lossless materials.

The authors in [9] presented a model metrics in relation to evaluate the Quality of Experience (QOE) and signal strength for future QOE optimization using a wireless Local Area Network (LAN) recognized as 802.11b/g in the IEEE standard. From the empirical study performed by the authors, using a mobile web browsing application which was tested on a Personal Digital Assistant (PDA), a result was obtained and evaluated using both the linear and exponential regression. The measurement was taken at four different locations using wireless LAN in the test environment. The result demonstrated that Location 2 with signal strength of -61dBm has a higher MOS rating compared to Location 4 with signal strength of -83dBm. It can be deduced from the result that the better the QOE, the better the QOS. Although, the author was able to give a correlative result between the QOS and QOE based on user perception, more than one user experience should be perceived to obtain more reliable result.

III. THE 2.4GHZ AND 5GHZ FREQUENCY BANDS

Wi-Fi standards such as 802.11b, 802.11g and 802.11n-2.4 utilize 2.400 - 2.500GHz spectrum band, while the 802.11a and the 802.11n use the more regulated 4.915 - 5.825GHz spectrum. The 2.4GHz band has only fourteen unlicensed channels with a 5MHz spacing, four of which are non-overlapping (1, 6, 11 and 14), while the 5GHz band though more regulated has more than 23 non-overlapping 20MHz channels with 20MHz spacing. However, it is worthy of note that the channels used in different countries are subjective. The 2.4GHz band have greater coverage as compared to the 5GHz band, however, due to relative pervasiveness of the 2.4GHz band, interference has become an issue for research and design discussion contrary to the 5GHz band, which has more bandwidth room and less usage by commonly found devices. The 5GHz band offers larger spectrum band, hence more bandwidth compared to the 2.4GHz band [9] and [10].

IV. METHODOLOGY

This work investigates only the 2.4GHz band of the Wi-Fi standard and not the 5GHz band. This is because of its significantly higher coverage range, pervasiveness and easier penetration of obstruction (walls). The popularity of 2.4 GHz devices and its relatively higher permittivity through walls informs our decision make use of the frequency band. The ease of permittivity to walls is a vital requirement to the method adopted in signal parameter measurement. The fact that most of the channels in the 5GHz spectrum lies outside the ISM unlicensed band, hence there is more restriction working in the band for experimental purpose.

A. Description of Materials

The focus of this research was based on five selected construction materials which are: 3 timber species, Concrete Slab, and a Glass-Reflective. The materials were categorized as Material A (*Apa- Afzelia Africans)*, Material B (*Agba-Gosswei lerodendrin balsamic ferum*), Material C (*Iroko- Chlorophora excels*), Material D (concrete) and Material E (reflective glass).

B. Software and Hardware Setup

To achieve the aim of this work, an android smart phone, version 8.0.0 (ABCDEF-180604V15) of model Infinix X-608 was used to measure the network parameters using RF signal tracker and Open Netttest software. The RF signal tracker used by RF Engineers is capable of measuring the RSS with respect to its Geo Location for selected interval of 10s. While the Open Nettest which is co-financed by the European Union can measure the QOS parameters which include Jitter (ms), Upload and Download Speed (Mpbs), and Packet loss (%). The Open Nettest create a server which support IPV4 and IPV6 standard, which allow the network performance to be evaluated depending on the supported scheme.

C. Description of Parameters Obtained

The parameters used in this research are briefly defined below:

Receive Signal Strength (RSS): The Received Signal Strength (RSS) is a metric used by manufacturers to give users an indication of the signal strength their wireless device is receiving for a wireless network, measured in dBm [11].

Upload speed: This involves the interaction from the user terminal to the remote system (BTS) which is set-up for uploading data. It is measured in (Mbps).

Download Speed: This involves the interaction from the user terminal to the remote system (BTS) which is set-up for downloading data. The opposite of the Upload Speed is done with the Download Speed. Its unit is measured in (Mbps).

Jitter: This occurs as a result of packet loss. It is a measure of the delay in packet sent. It is pronounced in video conferencing and VOIP (voice over internet protocol). It is measured in (ms)

Packet Loss: This occurs when one or more transmitted packets fail to arrive at its destination node. It causes observable effect in every digital communication system. It is measured in (%)

$$\textbf{Packet Loss}(\%) = \left(\frac{loss\ packet(s)}{Total packets} \times 100\right) \qquad (1)$$

D. Determination of Penetration Loss (dBm)

The penetration loss is a measure of the absorption of wireless signal when subjected to a material having specific absorption rate. It can be defined as a loss in signal strength

inside the material as compared to outside. This loss occurs in both lossy and lossless material as a result of material composition.

The penetration loss is obtained using the penetration loss formula, which is defined as

$$P_{Loss} = \sqrt{(RSS_{indoor} - RSS_{Outdoor})^2} \quad (2)$$

Alternatively;

$$P_{Loss} = ABS(RSS_{indoor} - RSS_{outdoor}) \quad (3)$$

where P_{Loss} is Penetration Loss in (dBm), ABS is a measure of the absolute value, RSS_{indoor} is the Indoor Received Signal Strength in (dBm) and $RSS_{Outdoor}$ is the Outdoor Received Signal Strength in (dBm).

E. Computation of the RSS Mean Value

The mean or average value of the collected data over 10 samples using the RF signal Tracker to measure the RSS in dBm can be expressed for the RSS_{indoor} and $RSS_{Outdoor}$ for 10 sampled signal using the following expressions.

$$Average\ RSS_{indoor} = \frac{\sum_{k=1}^{n} RSS_{indoor_k}}{n} \quad (4)$$

$$Average\ RSS_{outdoor} = \frac{\sum_{k=1}^{n} RSS_{outdoor_k}}{n} \quad (5)$$

Alternatively using Matlab script;

$$Average\ RSS_{outdoor} = \frac{Sum(RSS_{outdoor})}{Length(RSS_{outdoor})} \quad (6)$$

$$Average\ RSS_{indoor} = \frac{Sum(RSS_{indoor})}{Length(RSS_{indoor})} \quad (7)$$

where n is the total number of the sampled event.

F. Theoretical Expression for Unit Conversion

The standard referenced value of power (dBm) can be converted to Watt using the theoretical expression given below.

$$RSS(dBm) = 10\log_{10}\left(\frac{P_r}{1mW}\right) \quad (8)$$

where P_r is the received power by the smart phone antenna in *W*.

$$P_r(w) = 10^{\wedge}\left(\frac{1}{10} \times RSS(dBm)\right) \quad (9)$$

G. Data Collection Procedure

Procedure I: Measurement for the RSS

Using an android smart phone described earlier, having RF signal Tracker being installed on it accessing only Wi-Fi. (2.4GHz), the GPS receiver of the phone was turned on to give the geo location of the mobile phone. The software user interface capable of giving the signal strength measurement was launched. The graphical and accurate numerical values of the parameters were recorded for both indoor and outdoor. Measurements were sampled for 100s at time interval of 10s and the average values were recorded. The above procedures were repeated for material A, B, C, D and E.

Fig. 2. The different construction materials.

Procedure II: Measurement for the QOS (Quality of Service Parameter)

Using the same android smart phone having Open Net Test software installed on it accessing only Wi-Fi. (2.4GHz), the GPS receiver of the phone was launched to give the geographical parameters of the mobile phone. Other procedures like the ones described in Procedure I were carried out.

Fig. 2. shows the pictures of the construction materials being investigated.

V. RESULTS AND DISCUSSION

For the materials under investigation, the result obtained for both indoor and outdoor scenarios when subjected to Wi-Fi (2.4GHz) are presented in Fig. 4 to Fig. 6. Similarly, the computational mean values for analysis are also presented in Table III to Table XIV.

A. Material A

Material A, termed *Afzelia Africans (Apa)*, is one of the wood species commonly used in Nigeria as household construction materials. The indoor and outdoor measurements are shown in Table III to Table IV.

Table III. Mean value of RSS for Material A

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)
A	-77.3	-72.9

Table IV. QOS Parameter for Material A

Material A Indoor	
QOS Parameter	Measured Value
Jitter	26.8ms
Download Speed	0.12Mbps

Upload Speed	0.017Mbps
Packet Loss	0.00%
Material A Outdoor	
QOS Parameter	**Measured Value**
Jitter	7.02ms
Download Speed	0.051Mbps
Upload Speed	0.34Mbps
Packet Loss	0.00%

B. Material B

Material B, termed as *Gosswei Lerodendrin balsamic ferum (Agba)*, is another commonly used wood species in Nigeria for household construction. The indoor and outdoor measurements are presented in Table V and Table VI.

Table V. Mean value of RSS for Material B

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)
B	-45	-37.7

Table VI. QOS Parameter for Material B

Material B Outdoor	
QOS Parameter	**Measured Value**
Jitter	7.02ms
Download Speed	0.051Mbps
Upload Speed	0.34Mbps
Packet Loss	0.00%
Material B Indoor	
QOS Parameter	**Measured Value**
Jitter	40.2ms
Download Speed	0.080Mbps
Upload Speed	0.020Mbps
Packet Loss	0.00%

C. Material C

Material C is termed as *Chlorophoral excels (Iroko)*. It is also one of the wood species commonly used in Nigeria for household construction. The indoor and outdoor measurements are illustrated in Table VII and Table VIII.

Table VII. Mean value of RSS for Material C

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)
C	-82.4	-84.3

Table VIII. QoS Parameter for Material C

Material C (Indoor)	
QoS Parameter	**Measured Value**
Jitter	36.3ms
Download Speed	0.20Mbps
Upload Speed	0.013Mbps
Packet Loss	0.00%
Material C (Outdoor)	
QoS Parameter	**Measured Value**
Jitter	7.02ms
Download Speed	0.051Mbps
Upload Speed	0.34Mbps
Packet Loss	0.00%

D. Material D

Material D is termed as concrete and is being used in building and construction work and in household construction materials. The indoor and outdoor measurements are shown in Table IX and Table X.

Table IX. Mean value of RSS for Material D

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)
D	-87.1	-81.7

Table X. QOS Parameter for Material D

QOS Parameter	Measured Value Outdoor
Jitter	7.02ms
Download Speed	0.051Mbps
Upload Speed	0.34Mbps
Packet Loss	0.00%
QOS Parameter	**Measured Value Indoor**
Jitter	36.4ms
Download Speed	0.17Mbps
Upload Speed	0.023Mbps
Packet Loss	2.00%

E. Considering Material E

Material E is termed Reflective Glass. This type of glass can allow low or no penetration as a result of the coatings. The indoor and outdoor measurements are illustrated in Table XI and Table XII

Table XI. Mean Value of RSS for Material E

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)
E	-14.6	-14.2

Table XII. QOS Parameter for Material E

QOS Parameter	Measured Value Outdoor
Jitter	7.02ms
Download Speed	0.051Mbps
Upload Speed	0.34Mbps
Packet Loss	0.00%
QOS Parameter	**Measured Value Indoor**
Jitter	33.9ms
Download Speed	0.22Mbps
Upload Speed	0.033Mbps
Packet Loss	0.00%

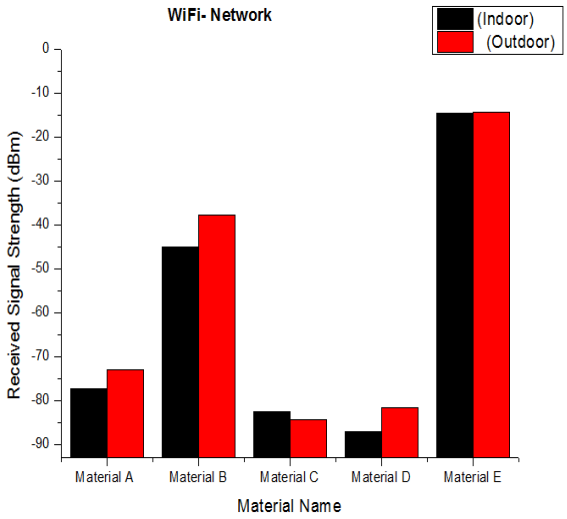


Fig. 4. Mean Value of the Materials indoor and outdoor Signal Strength

F. Average Value of the Penetration Loss

The mean value of the penetration loss also referred to as absorption loss as a result of the material characteristics is presented in Table XIII. The average penetration loss can be computed using (2) and (3) for the five selected materials.

The average value for the RSS which was measured within and outside the materials when subjected to Wi-Fi signals is illustrated in Fig. 6.

Table XIII. Average Penetration Loss for the Materials

Material Under Test	Wi-Fi Signal Penetration Loss (dBm)
A	5.4
B	7.3
C	5.3
D	5.4
E	2.6

Table XIV. Change in Signal Strength Level

Material	Wi-Fi Indoor (dBm)	Outdoor (dBm)	RSS Indoor (fW)	Outdoor (fW)	Percentage Change (%)	Remark
A	-77.3	-72.9	18620.9	51286.1	63.7	Decrease
B	-45	-37.7	31622776.6	16982436.2	46.3	Decrease
C	-82.4	-84.3	5754.4	3715.4	35.4	Increase
D	-87.1	-81.7	1949.8	6760.8	71.2	Decrease
E	-14.6	-14.2	34673685050	38018939630	8.8	Decrease

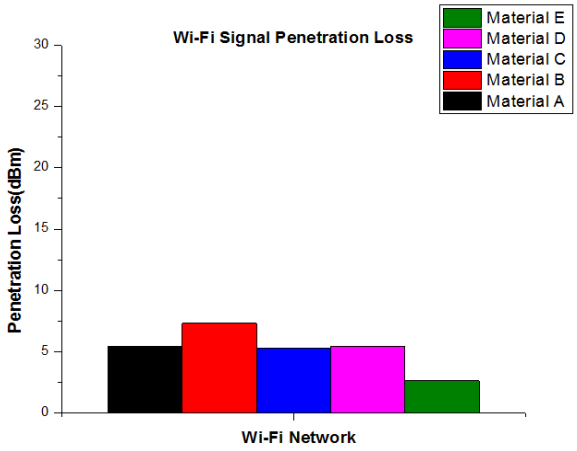


Fig. 5. Materials Penetration Loss

The QoS parameters were obtained from open Net Test in the application Interface as presented in Table III to Table XII describe the quality of experience perceived by the users when subjected to different materials for the indoor and outdoor scenario.

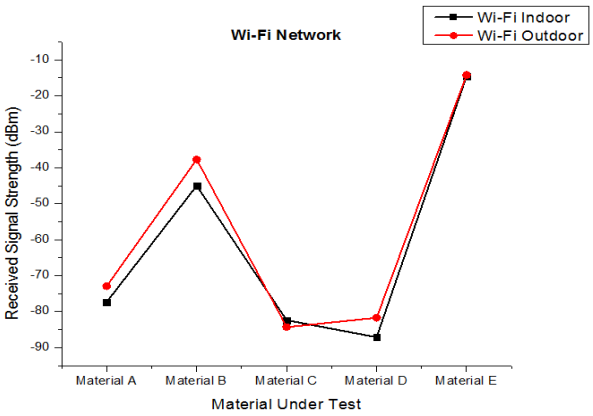


Fig. 6. Indoor and Outdoor RSS Variation

As a result, Jitter is being considered since it gives substantial variation for both the indoor and outdoor scenario. The jitter variation in the selected materials for the Wi-Fi signal is demonstrated in Fig. 7.

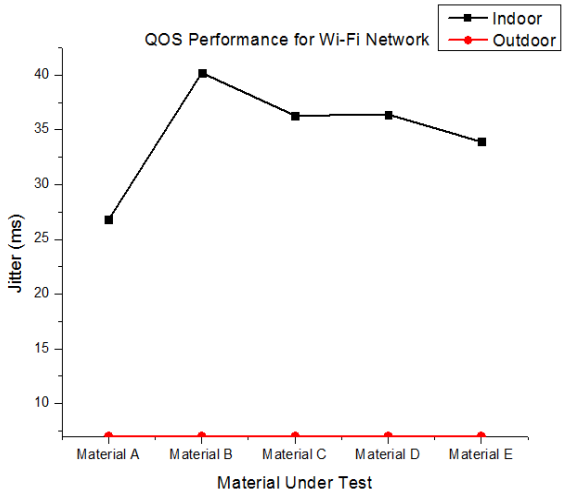


Fig. 7. Jitter Variation in Materials.

VI. CONCLUSION

In conclusion, the results obtained from the surveyed data with the aid of RF signal Tracker and Open NetTest software using Infinix X608 mobile phone showed that the construction materials have significant impact on Wi-Fi signal in terms of RSS and QoS. The result of the penetration loss, indoor and outdoor signal strength are compared. It shows that all the five selected materials have significant effects on signals.

The result from the penetration loss confirms that Material B has the highest penetration loss of 7.3dBm while Material E has the lowest penetration loss of 2.6dBm. This attests for to the fact from the reviewed works that the type of construction materials affects Wi-Fi signal. Material C has 35.4% increase in indoor signal strength, Material E has 8.8% decrease in indoor signal strength, Materials A, B and D have 63.7%, 46.3% and 71.2% decrease respectively in indoor received signal strength.

The result from QOS parameter (Jitter) are compared for both outdoor and indoor. It shows that Material A and Material E will provide improved experience to the user, as a result of their spread. This illustrated that calculating the material penetration loss accounts for attenuation even as the penetration loss is a function of the material composition.

REFERENCES

[1] J. B. Andersen, T. S. Rappaport, and S. Yoshida, “Propagation Measurement and Model for Wireless Communication Channel”, 1995.

[2] C. R. Anderson, T. S. Rappaport, K. Bae, A. Verstak, N. Ramakrishnan, W. H. Tranter, C. A. Shaffer, and L. T. Watson, "In-buiding wideband Multipath characteristics at 2.5GHz and 60GHz," in *IEEE Conference*, vol. 5, pp. 97-101, 2010.

[3] R. Richard, C. Ken, G. Martin, and H. Richard, "Building Materials and Propagation," Department of Signal Science, OFCOM, U.K, Document 2014. [Online]. http://www.zerocarbonhub.org/current-roject/performance-gap

[4] ELDES, UAB, “Radio System Installation and Signal Penetration”, 2016.

[5] T. S. Rappaport, “*Wireless Communications: Principle and Practice*”, 2nd ed. Singapore: Pearson Education PTE Limited, 2003.

[6] E. Promise and O. Otasowle, "Determination of GSM Signal Penetration Loss in Some Selected Buildings in River State, Nigeria", *Nigerian Journal of Technology (NIJOTECH)*, vol. 34, no. 3, pp. 609-615, 2015. [Online]. www.nijotech.com

[7] S. C. Felicito, and C. D. Jeniffer, "Penetration Loss of Doors and Windows Inside Residence Using ISDB-T Digital Terrestrial Television Signal at 667MHz", in *Proceedings of the World Congress on Engineering and Computer Science (WCECS)*, vol. II, San Francisco, 2011.

[8] F. Y. William, A. R. Kate, L. John, L. H. Christopher, G. Chriss, K. Galen, C. Dennis, F. Sander, N. Wouter, and G. Andrea, "Measurement to Support Public Safety Communication: Attenuation and Variability of 750MHz Radio Wave Signals in Four Large Building Structures", U.S Department of Commerce, National Institute of Standard and Technology, Washington, NIST Technical Note 20402-0001, 2009. [Online]. www.gpo.gov

[9] D. Katerien, J. Wout, K. Istvan, T. Emmeric, D. Tom, M. Luc, D. Lieven, "Linking Users Subjective QOE (Quality of Experience) Evaluation to Signal Strength in an IEEE

802.11b/g wireless LAN Environment," *EURASIP Journal*, vol. 2010, pp. 1-12, February 2010.

[10] S. Ahsan, A. Zeeshan, A. Iftikhar, "Analysis and Measurement of Wi-Fi Signals in Indoor Environment", *International Journal of Advances in Engineering & Technology (IJAET)*, vol. VI, no. 2, pp. 678-687, May 2013.

[11] A. Scott, "Guide to surveying building for Wireless Network," *JISC* , July 2011, Published on Jisc community (https://community.jisc.ac.uk). [Online]. http://www.community.jisc.ac.uk/library/advisory-services/guide-surveying-buildings-wireless-networking

CULTURAL EVOLUTION FROM NOMADIC PASTORALISM TO NOMADIC PASTORALIST OPTIMIZATION ALGORITHM (NPOA): THE MATHEMATICAL MODELS

Ibrahim M. Abdullahi
Department of Computer Engineering
Federal University of Technology, Minna
Minna, Niger State, Nigeria
amibrahim@futminna.edu.ng

Muhammed B. Mu'azu
Department of Computer Engineering
Ahmadu Bello University, Zaria
Zaria, Kaduna State, Nigeria
muazumb1@yahoo.com

Olayemi M. Olaniyi
Department of Computer Engineering
Federal University of Technology, Minna
Minna, Niger State, Nigeria
mikail.olaniyi@futminna.edu.ng

James Agajo
Department of Computer Engineering
Federal University of Technology, Minna
Minna, Niger State, Nigeria
James.agajo@futminna.edu.ng

***Abstract*—In this paper, the mathematical models for a proposed novel modified Pastoralist Optimization Algorithm (POA) called the Nomadic Pastoralist Optimization Algorithm (NPOA) inspired by the nomadic pastoralists herding strategies and cultural evolution strategy is presented. The nomadic pastoralist herding strategies which are scouting, camping, herding, splitting and merging were mathematically modeled. The mathematical models will be used to develop the proposed algorithm. The algorithm when developed will be tested on several benchmark functions to ascertain the algorithms exploration and exploitative ability. The performance will also be validated by comparing with POA and other popular and similar metaheuristic algorithms such as GOA, PSO, ABC, BBO and ICA**

Keywords— Optimization; Nature Inspiration; Metaheuristics; Cultural Algorithm, Pastoralist Optimization Algorithm (POA); Nomadic Pastoralist Herding Strategy (NPHS);

I. INTRODUCTION

The rapid growth in technology has brought about faster and more accurate solutions to emerging real world problems. At the heart of this technological advancement and optimal solution seeking is optimization. Optimization is basically a search for optimal solution using the right procedures and mathematical representations [1]. [2] defined optimization (or mathematical programming) as a systematic selection of variable values within some allowed limits whose aim is to minimize or maximize an objective function of a decision problem. Optimization is viewed as optimal seeking in nature in which problem dependent objectives (performance index) must be evaluated or achieved and constraints must be satisfied [3]. Optimization problems (OP) are problems that contains several solutions, variables, constraints and a function or performance measure to measure the optimality of a chosen solution. The general approaches for solving OP can be analytical, experimental, graphical or numerical. [4].

Real world OPs are complex and difficult to solve because of their large number variables and constraints, non-linear and multi-modal objective function and are computational expensive, hence, the need for innovative optimization techniques in solving them [5, 6]. This innovative Nature Inspired (NI) optimization techniques which are mostly population based and metaheuristic Optimization Algorithms (OA) have proven to be very efficient in solving most real world problems. Novel nature-inspired metaheuristic OA are being developed because according to the no free lunch theorem, no OA can optimally solve all Op's, even though they are capable of solving most OP.s [7].

NI-OA are inspired by natural phenomenon, and they are classified as either swarm-based, human-based, evolutionary-based, chemistry-based, physics-based and mathematics-based [8]. Example of some of these algorithms include Particle Swarm Optimization (PSO) [9], Ant Colony Optimization (ACO) [10], Artificial Bee Colony (ABC) [11], Biogeography-based Optimization (BBO) [12], Ant Lion Optimization (ALO) [13], Whale Optimization Algorithm (WOA) [14], Lion Optimization Algorithm (LOA) [15] and Grasshopper Optimization Algorithm (GOA) [16], Pastoralist Optimization Algorithm (POA) [17].

Most of the listed algorithms deploy mostly the biological evolution strategy through mutation or crossover or both and agents share information with a narrow temporal and spatial scale. Cultural evolution strategies on the other hand allows agents to evolve share information through a well-structured belief space. Cultural Algorithm (CA) allows agents to learn from a global knowledge domain rather than local as in the case of most OA. This allows culture to evolve faster than biological and other social evolution strategies. CA have been applied for evolution and modification of some algorithms [18].

In this paper, the cultural evolution strategy was adopted for the evolution of the Nomadic Pastoralist Herding Strategy (NPHS). The strategy has been used to develop a novel POA using the biological (genetic) evolution strategy [17]. Although the algorithm show promising results, there is still need for improvement especially in convergence speed and accuracy.

The remainder of this paper is structured as follows: In Section 2.1, cultural evolution framework is presented, section 2.2, Pastoralist herding strategy is presented, by the mathematical models of NPOA using cultural algorithm evolution strategy and lastly, conclusion and recommendation in Section 4.

II. CULTURAL EVOLUTION FRAMEWORK

A. Cultural Algorithm (CA)

Cultural Algorithm (CA) is a group of computational models that are characterized by three major components; the population space, the belief space and the procedure that describes the knowledge sharing approach between the belief and population space [19]. These models are derivative of the cultural evolution process in nature as shown in the CA framework in Fig. (1).

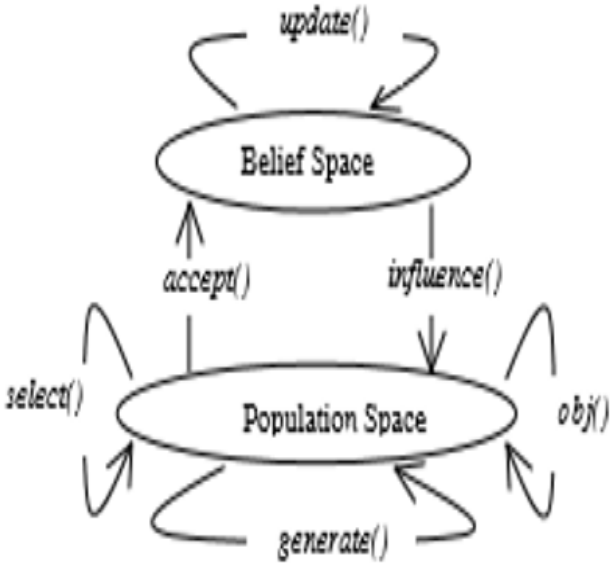


Fig. 1. Cultural algorithm framework [18]

CA is a dual inheritance system that describes evolution in human culture at macro-evolutionary level and micro-evolutionary level which occur in the belief and population space respectively [18]. That is, both the belief and population spaces are updated after each time step based on each other's feedback. As shown in Figure 1, the fitness of each individual is first evaluated using the *obj()* objective function in the population space, after which accepted individuals from the population space are used to update the belief space using the *accept()* function for selecting the individuals. Using the *influence()* function, the individuals to form the population of the next generation are selected using the knowledge from the belief space [20]. For each generation, this processes are repeated until a pre-specified termination condition is met. Fig. (2) shows the cultural algorithm pseudocode.

```
Start
    Initialize
        t=0; B^t; P^M
    Repeat;
        Evaluate Fitness of P^M: ⟹ obj(P^t)
        Select Accepted Individuals: ⟹ Accept(P^t)
        Update Belief Space B^t: ⟹ Update(B^t)
        Generate New Population P^(t+1): ⟹ Influence (B^t)
        Update Time: ⟹ t=t+1
        Until Termination Condition Is Met
end
```

Fig. 2. Cultural algorithm pseudocode [20]

B. Knowledge Sources of CA

CA is built There are five knowledge sources; Situational. Normative, Domain, Historical and Topological.

i. Situational Knowledge (SK): this knowledge source stores the best found exemplars or solution throughout the evolutionary process and is used to lead or guide other individuals on the direction of search, that is towards the exemplars [21, 20, 22].

ii. Normative Knowledge (NK): this knowledge source stores the minimum and maximum values of numeric attributes or a lists of all possible nominal attributes [22]. Normally used during mutation, it guides the adjustment behavior of individuals by determining the step size of search [20].

iii. Domain Knowledge (DK): This knowledge source keeps information about the problem domain used in guiding a search [20]. It also keeps the accepted rules of each generation that are used to guide search for subsequent generations [21].

iv. Topographical Knowledge (TK): This knowledge source is used to diversify the set of rules generated by individual agents in order to prevent local optima entrapment [22]. It was proposed originally to explain region-based functional landscape patterns.

v. Historical Knowledge (HK): is used to store significant events during the search process such as moves, fitness and landscape change in order to guide future moves [20].

These five knowledge sources play different roles in a search process with diverse problem solving capability. The Mathematical representation of the SK and NK knowledge sources which are represented in the belief space, the acceptance function and belief space adjustment, the influence function and all its updating strategies can be found in [23].

III. NOMADIC PASTORALISM

Pastoralism is a livestock production system characterized by extensive movement of animals in search of water and quality pasture [24]. The traditional knowledge of pastoralism allows the pastoralists to manage all entities efficiently using some highly flexible strategies. These strategies help the nomadic pastoralists to survive the unpredictable and potentially hazardous pastoral life [25, 26].

A. Nomadic Pastoralist Herding Strategy (NPHS)

The strategies adopted by the nomadic pastoralist include: Scouting for exploration and search for suitable camp site [27] , camp selection and camping for temporary settlements and daily exploitation [28], herding, which include splitting or herd dispersal for daily herding, risk minimization and trap avoidance [29], finally, merging for camp fitness evaluation and the search for a new camp depending on the quality assessment [30]. The pastoralist herding strategy is shown in Fig. (3).

These strategies have been modelled mathematically using the biological evolution strategy and used to develop a novel POA. The algorithm has been tested on standard benchmark unimodal and multimodal functions and its result were very competitive in terms of its exploration and exploitation capability [17]. However, it suffers from slow convergence which inspires the evolution of NPHS using cultural evolution strategy.

B. Mathematical Models of the Cultural Evolution of NPHS

a) Initialization

The first step in developing the NPOA is to generate the population of pastoralist (nP) randomly because NPOA is a population-based metaheuristic algorithm. In NPOA, a solution is called a pastoralist which is represented in the search space as:

$$P = [P_1, P_2, P_3, \ldots, P_D] \quad (1)$$

where, P is the pastoralist and D is the dimension or number of variables of the optimization problem. The second step is to select (25%) of the pastoralist as scout pastoralist (S) represented as;

$$S = [S_1, S_2, S_3, \ldots, S_D] \quad (2)$$

where, S is the scout pastoralist.

Next, the belief space at time t ($B(t)$) which comprises the situational, normative and domain knowledge component is initialized as shown in Equation (3).

$$B(t) = (\{\varsigma(t)\}, \{N(t)\}, \{\delta(t)\}) \quad (3)$$

where,

$$\varsigma(t) = \{\hat{y}_i(t): , i = 1: ns\} \quad (4)$$

$$N(t) = \{X_j(t): X_j(t) = \left(I_j(t), L_j(t), U_j(t)\right), \ j = 1: D\} \quad (5)$$

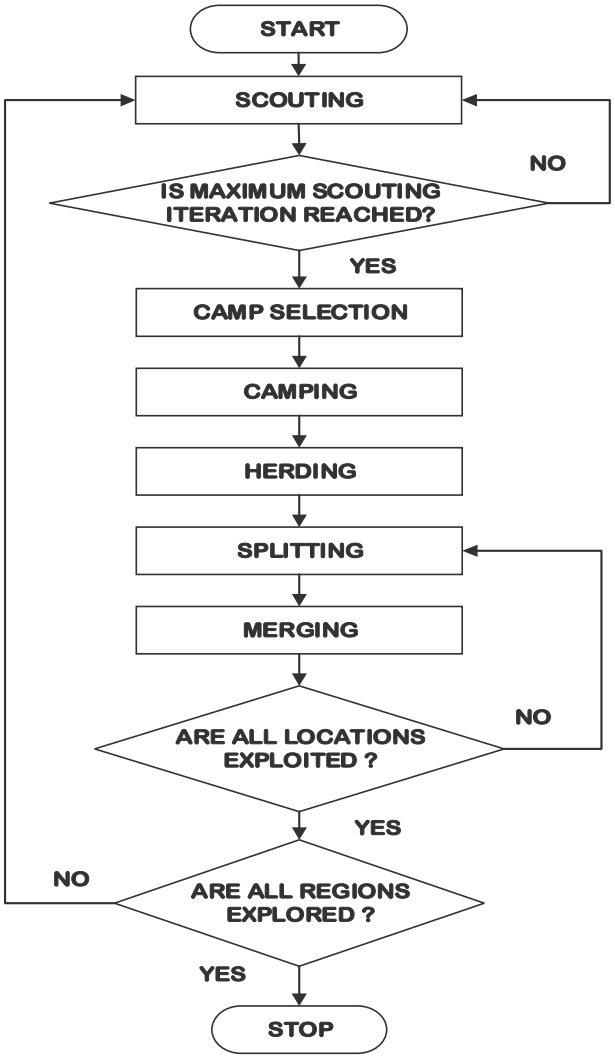


Fig. 3. Nomadic pastoralist herding strategy

S(t) which is the SK component is divided into the Scouters situational knowledge component $S_S(t)$ and the Herders Situational knowledge component $S_H(t)$, that is, $\varsigma(t) \epsilon \{\varsigma_S(t), \varsigma_H(t)\}$. Similarly, the normative knowledge component is also divided into the Scouters NK and Herders NK represented as $N_S(t)$ *and* $N_H(t)$ respectively, where $N(t) \epsilon \{N_S(t), N_H(t)\}$.

The closed interval $I_j(t) = [x_{max,j}(t) - x_{min,j}(t)], U_j(t)$ and $L_j(t)$ are the upper and lower bound score respectively while ns is the number of scouts and D is the dimension of the search problem. $\delta(t)$ represents the domain knowledge component at time *t* which stores the rules in the splitting stage.

b) Scouting

After selecting the number of scout pastoralist, their locations are initialized randomly within the search space using Equation (6) and followed by evaluation of fitness of each scout. The fitness of scout i is evaluated using Equation (4).

$$S_{i,j} = rand([L_b, U_b]^D) \quad (6)$$

$$F(S_{i,j}) = FF(S_j) \quad (7)$$

where $rand([L_b, U_b]^D)$ is a D-dimensional random vector between the lower bound and upper bound of the search space and FF is the fitness function which is problem dependent. Next, the finesses of all scouts were sorted based on their fitness values in ascending order. The best 20% of scout population whose behaviour are acceptable are selected for the belief space adjustment using the pseudocode in Equation (8).

$$S_{accept} = S_{sorted}(1:nAccept) : (nAccept = round(0.2 * nS)) \quad (8)$$

- Situational Knowledge Update: The situational knowledge component is updated using Equation (9).

$$\zeta_s(t+1) = \begin{cases} F(S_i(t)), & if\ F(S_i(t)) < \hat{y}(t) \\ \hat{y}(t), & \text{otherwise} \end{cases} \quad (9)$$

where $min\{F(S_i(t))\}$ is the minimum fitness of scout i, and $\hat{y}(t)$ is the initial g best fitness of all scouts at time t.

- Normative Knowledge Update: The normative component determines the step size of the search, hence, controls the algorithms exploration and exploitation. The rules of updating the normative knowledge components are given as follows;

$$S_{min,j}(t+1) = \begin{cases} S_{i,j}(t), & if\ S_{i,j}(t) \leq S_{min,j} \text{ or } F(S_i(t)) < L_j(t) \\ S_{min,j}(t), & \text{otherwise} \end{cases} \quad (10)$$

$$S_{max,j}(t+1) = \begin{cases} S_{i,j}(t), & if\ S_{i,j}(t) \geq S_{max,j} \text{ or } F(S_i(t)) < U_j(t) \\ S_{max,j}(t), & \text{otherwise} \end{cases} \quad (11)$$

$$L_j(t+1) = \begin{cases} F(S_i(t)), & if\ S_{i,j}(t) \leq S_{min,j} \text{ or } F(S_i(t)) < L_j(t) \\ L_j(t), & \text{otherwise} \end{cases} \quad (12)$$

$$U_j(t+1) = \begin{cases} F(S_i(t)), & if\ S_{i,j}(t) \geq S_{max,j} \text{ or } F(S_i(t)) < U_j(t) \\ L_j(t), & \text{otherwise} \end{cases} \quad (13)$$

$$I_j(t+1) = [S_{max,j}(t+1) - S_{min,j}(t+1)] \quad (14)$$

$I_j(t+1)$ is the size of the normative component at time t + 1.

Due to the diversity of scouters during searching, the step size (that is the normative component size of scout j (I_j)) is high which guarantees effective exploration of scout pastoralists.

- Scout Population Influence: The updated SK and NK components were used to influence the scout population if the maximum scouting rate (α) is not exceeded. The scouts move into a new location guided by normative and situational knowledge component as shown in Equation (15).

$$S'_{i,j} = \begin{cases} S_{i,j} + |\alpha * I_j(t+1) * N_{ij}(0,1)| & if\ S_{i,j} < \hat{S}_j \\ S_{i,j} - |\alpha * I_j(t+1) * N_{ij}(0,1)| & if\ S_{i,j} \geq \hat{S}_j \end{cases} \quad (15)$$

where, $S'_{i,j}$ and $S_{i,j}$ is the next and current position of scout i for variable j, $\hat{S}_j$ is the best scout position so far. The fitnesses of the new scouters are re-evaluated using Equation (7).

c) Camp Selection and Camping

Selection of the best location for camping is obtained by selecting the best scout in terms of their fitness after completing the maximum scouting iteration. The roles of the scout pastoralist are reversed to herders after scouting and they are joined with other pastoralists. The k[th] pastoralist P_K is initialized at a camp using Equation (16).

$$P_{k,j} = \begin{cases} \hat{S}_j & if\ k = 1 \\ \hat{S}_j + rand([-r, r]^D) & if\ k > 1 \end{cases} \quad (16)$$

where $\hat{S}_j$ is the best scout position, r is the camp radius and D is the variable size (j ϵ [1:D]).

d) Herding

The fitness of the k[th] pastoralist is evaluated using Equation (17) during herding. This is followed by sorting and selection of the best 20% of pastoralist or herders' population whose behaviour are acceptable are selected for the belief space using Equation (18).

$$F(P_K) = FF(P_K) \quad (17)$$

$$P_{accept} = P_{sorted}(1:nAccept) : (nAccept = round(0.2 * nP)) \quad (18)$$

- Situational Knowledge Update: The situational knowledge component of herders is updated using Equation (9).

$$\zeta_H(t+1) = \begin{cases} F(P_i(t)), & \text{if } F(P_i(t)) < \zeta_s(t+1) \\ \zeta_s(t+1), & \text{otherwise} \end{cases} \quad (19)$$

where $min\{F(P_k(t))\}$ is the minimum fitness of pastoralist k, and $\zeta_H(t+1)$ is the situational best fitness of the herders at time t+1, while $\zeta_S(t+1)$ is the situational best fitness of all scouters.

- Normative Knowledge Update: The normative components of herders were updated as follows;

$$P_{min,j}(t+1) = \begin{cases} P_{k,j}(t), & \text{if } P_{k,j}(t) \leq P_{min,j} \text{ or } F(P_k(t)) < L_j(t) \\ P_{min,j}(t), & \text{otherwise} \end{cases} \quad (20)$$

$$P_{max,j}(t+1) = \begin{cases} P_{k,j}(t), & \text{if } P_{k,j}(t) \geq P_{max,j} \text{ or } F(P_k(t)) < U_j(t) \\ P_{max,j}(t), & \text{otherwise} \end{cases} \quad (21)$$

$$L_j(t+1) = \begin{cases} F(P_k(t)), & \text{if } P_{k,j}(t) \leq P_{min,j} \text{ or } F(P_k(t)) < L_j(t) \\ L_j(t), & \text{otherwise} \end{cases} \quad (22)$$

$$U_j(t+1) = \begin{cases} F(P_k(t)), & \text{if } P_{k,j}(t) \geq P_{max,j} \text{ or } F(P_k(t)) < U_j(t) \\ L_j(t), & \text{otherwise} \end{cases} \quad (23)$$

$$I_j(t+1) = [P_{max,j}(t+1) - P_{min,j}(t+1)] \quad (24)$$

$I_j(t+1)$ is the size of the normative component at time t + 1.

I_j (the normative component size) is very small during herding because of the closeness of pastoralist in the camps. This allows the algorithm to effectively exploit the camping area.

e) Splitting

Each pastoralist (herders) split to different locations within the camp using the normative, situational and domain knowledge components as shown in Equation (25).

$$P'_{k,j} = \begin{cases} \hat{P}_{k,j} + |\alpha * I_j(t+1) * N_{ij}(0,1)| & \text{if } P_{k,j} < \hat{P}_j \\ \hat{P}_{k,j} - |\alpha * I_j(t+1) * N_{ij}(0,1)| & \text{if } P_{k,j} \geq \hat{P}_j \\ round(\hat{P}_{k,j}) & \text{if } t > 1 \;\&\&\; \delta(t) > \phi \end{cases} \quad (25)$$

where, $P'_{k,i}$ is the next position of k[th] pastoralist, $\hat{P}_{k,i}$ is the previous best position of herders and $P_{k,i}$ is the current pastoralist location. The domain knowledge rule $\delta(t)$ is given as:

$$\delta(t) = \left(\frac{\zeta_S(t+1) - \zeta_H(t+1)}{\zeta_s(t+1)}\right) * 100 \quad (26)$$

where $\zeta_S(t+1)$ and $\zeta_H(t+1)$ scouters and herders situational best fitness, ϕ is a constant representing the branching threshold set as a percentage.

The fitness of the new herders positions $P'_{k,i}$ were evaluated using Equation (17) where P_K is replaced with $P'_{k,i}$, followed by acceptance population selection (Equation (18)). Next, the situational knowledge component is updated using Equation (27) followed by update of the normative knowledge components using Equations (20 to 24).

$$\zeta_H(t+2) = \begin{cases} F(P'_k), & \text{if } F(P'_k) < \zeta_H(t+1) \\ \zeta_H(t+1), & \text{otherwise} \end{cases} \quad (27)$$

where $F(P'_k)$ is the minimum fitness of k[th] pastoralist, and $\zeta_H(t+2)$ is the new situational best fitness of the herders at time t+2, while $\zeta_H(t+1)$ is the situational global best fitness of the herders at time t+1.

f) Merging

During merging, the fitness of the best location within the camp (C_{best}) is updated by comparing the situational best at t+2 and at t+1 as shown in Equation (28).

$$C_{best} = \begin{cases} \zeta_H(t+2), \text{ if } \zeta_H(t+2) < \zeta_H(t+1) \\ \zeta_H(t+1), \text{ otherwise} \end{cases} \quad (28)$$

where, C_{best_i} is the camp best location (that is the best pastoralist within the camp). If all locations in the camp have not been exploited (β not exceeded), the pastoralist splits again to new locations by repeating the steps in section (v).

If all regions have not been explored (maximum iteration not reached), the new regions to be explored by scouters were obtained using the situational knowledge components only as shown in Equation (29).

$$S'_{i,j} = \begin{cases} S_{i,j} + |\alpha * (Vmax_j - Vmin_j) * N_{ij}(0,1)| & \text{if } S_{i,j} < \hat{S}_j \\ S_{i,j} - |\alpha * (Vmax_j - Vmin_j) * N_{ij}(0,1)| & \text{if } S_{i,j} \geq \hat{S}_j \end{cases} \quad (29)$$

where, $S'_{i,j}$ and $S_{i,j}$ is the next and current position of scout i, $\hat{S}_j$ is the previous best position of scouters, $Vmax_j$ and $Vmin_j$ are the upper and lower bound of variable j. The processes in sections (i to vi) are repeated until the stopping criteria (maximum generation is reached). The Global best solution G_{best} is obtained as the last updated situational fitness and position.

Using these mathematical models, a modified POA called the Nomadic Pastoralist Optimization Algorithm (NPOA) will be developed. The algorithm when developed will be applied to solve various combinatorial and numerical optimization problems with the view of obtaining a competitive or better results will be obtained.

IV. CONCLUSION

This paper presents the mathematical models of the evolution from nomadic pastoralist herding strategy to Nomadic Pastoralist Optimization Algorithm (NPOA). The background of cultural algorithm framework and pastoralist herding strategies were first presented followed by the mathematical models of each herding strategy. This models will be used to develop the NPOA and the algorithm will be applied on some standard benchmark test functions. The performance of the algorithm will be compared with existing pastoralist optimization algorithm and some similar and popular metaheuristic optimization algorithms like PSO, GOA, BBO, ABC and ICA.

REFERENCES

[1] C. Arango, P. Corte´s, L. Onieva and A. Escudero, "Simulation–optimisation models for the dynamic berth allocation problem.," *Comput Aided Civil Infrastruct Eng.,* vol. 28, no. 10, pp. 769-779, 2013.

[2] G. Zhang, J. Lu and Y. Gao, Multi-Level Decision Making Models, Methods and Applixations, Bwelin Heidelberg: Springer, 2015.

[3] N. Siddique and H. Adeli, "Nature Inspired Computing: An Overview and Some Future Directions," *Cogn Compu,* vol. 7, p. 706–714, 2015.

[4] A. Antoniou and W. S. Lu, Practical Optimization: Algorithms and Engineering Aoolications, New York: Springer Science + Business Media, LLC, 2007.

[5] S. Sumathi, A. L. Kumar and P. Surekha, Computational Intelligence Paradigms for Optimization Problems Using MATLAB®/SIMULINK®, Boca Raton: CRC Press Taylor & Francis Group, 2016.

[6] F. Glover and A. G. Kocehenberger, Handbook of metaheuristics, Dordrecht: kluwer academic publishers, 2003.

[7] D. H. Wolpert and W. G. Macready, " No free lunch theorems for optimization,," *IEEE Transactions on Evolutionary Computation 1,* pp. 67-82, 1997.

[8] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software,* pp. 51-67, 2016.

[9] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization," in *Proceeding of the IEEE International Conference on Neural Networks,* , Perth, Australia, 1995.

[10] M. Dorigo, V. Maniezzo and A. œ. Colorni, "Ant System:Optimization by a Colony of Cooperating Agents.â ," in *Systems, Man. and Cybernetics-Part B*, 1996.

[11] D. Karaboga and B. Basturk, " A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of Global Optimization,,* vol. 39, no. 3, pp. 459-471, 2007.

[12] D. Simon, "Biogeography-based optimization.," *IEEE Transactions on Evolutionary Computation,* vol. 12, no. 6, pp. 702-713, 2008.

[13] S. Mirjalili, "Ant Lion Optimization," *Advances in Engineering Software,* vol. 83, pp. 80-98, 2015.

[14] Mirjalili, S; Lewis, A, "The Whale Optimization Algorithm," *Advances in Engineering Software,* pp. 51-67, 2016.

[15] M. Yazdani and F. Jolai, "Lion Optimization Algorithm (LOA): A Nature-Inspired Metaheuristic Algorithm," *Journal of Computational design and Engineering,* vol. 3, pp. 24-36, 2016.

[16] S. Saremi, S. Mirjalili and A. Lewis, "Grasshopper optimisation algorithm: theory and application," *Adv Eng Softw,* vol. 105, p. 30–47, 2017.

[17] I. M. Abdullahi, M. B. Mu'azu, O. M. Olaniyi and J. Agajo, "Pastoralist Optimization Algorithm (POA): A Novel Nature-Inspired Metaheuristic Optimization Algorithm," in *International Conference on Global and Emerging Trends (2018)*, Abuja, 2018.

[18] R. G. Reynolds, B. Peng and R. Whallon, "Emergent Social Structures in Cultural Algorithms," 2006.

[19] G. R. Reynolds, "An Introduction to

Cultural Algorithms," in *In Proceedings of the 3rd Annual Conference on Evolutionary Programming,*, 1994.

[20] G. R. Reynolds and B. Peng, "Cultural Algorithms: Computational Modeling Of How," 2005.

[21] G. Gill and P. Gupta, "Enhanced Cultural Algorithm of Data Mining for Intrusion Detection System," *International Jourla of Computer Science and Technology,* vol. 7, no. 1, pp. 148-152, 2016.

[22] S. Srinivasan and S. Ramakrishnan, "Cultural Algorithm Toolkit for Multi-objective," *International Journal on Computational Sciences & Applications (IJCSA),* vol. 2, no. 4, 2012.

[23] M. R. Shahriari, "Acultural algorithm for data clustering," *International Journal of Industrial Mathematics,* vol. 8, no. 2, pp. 99-106, 2016.

[24] D. G. Msuya, "Pastoralism Beyond Ranching: A Farming System in Severe Stress in Semi-Arid Tropics Especially in Africa," *Journal of Agriculture and Ecology Research International,* vol. 4, no. 3, pp. 128-139, 2015.

[25] M. Q. Sutton and E. N. Anderson, Introduction to Cultural Ecology, Plymouth, UK: AltaMira Press., 2010.

[26] C. Hesse and S. Cavanna, (2010). Modern and Mobile: The Future of Livestock Production in Africa's Drylands, UK: International Instutute for Environment and Develpment (IIED) and SOS Sahel International UK, 2010.

[27] J. Gerald and R. Dorothy, Pastoralism in Practice: Monitoring Livestock Mobility in Contemporary Sudan, Kenya: United Nations Environment Programme, 2013.

[28] C. Liao, P. E. Clark, S. D. DeGloria, A. Mude and C. B. Barret, "Pastoral mobility and policy recommendations for livestock herding in the Borana pastoral system in Southern Ethiopia," *Solutions Journal,* vol. 7, no. 5, pp. 17-20, 2016.

[29] d. G. Gert, "Taking Indegenious Knowledge Serious: The case of Pastoral Strategies Among Turkana Nomads of Northwest Kenya," ILEIA Newsletter, Eldoret, Kenya, 2011.

[30] H. K. Adriansen, "Understanding pastoral mobility: the case of Senegalese Fulani," *The Geographical Journal ,* vol. 174, no. 3, pp. 207-222, 2008.

[31] A. Rota and S. Sperandini, " Livestock and Pastoralist,," International Fund for Agricultural Development, 2008.

Sinkhole Attack Detection in A Wireless Sensor Networks using Enhanced Ant Colony Optimization to Improve Detection Rate

Kenneth E. Nwankwo
Department of Cyber Security Science,
Federal University of Technology
Minna, Nigeria
soundnwankwo@gmail.com

Shafi'i Muhammad Abdulhamid
Department of Cyber Security Science,
Federal University of Technology
Minna, Nigeria
shafii.abdulhamid@futminna.edu.ng

***Abstract* - Wireless Sensor Networks (WSN) comprise of tiny sensor nodes that are able sense and process data. Sinkhole attack occurs when an attacker node in a wireless sensor network disguises itself as the legitimate node closest to the base station, in order to have all data pass through hence having the opportunity to modify, drop or delay data going to the base station. we propose a sinkhole detection scheme in this conceptual framework, enhancing ant colony optimization to improve sinkhole detection via packet drop, packet delivery rate, energy exchange and throughput in a wireless sensor network. Initial results and expected outcome are shown and further research is also be discussed.**

***Keywords**: Swarm Intelligence, Ant Colony Optimization, Wireless Sensor Network, Sinkhole Detection,*

I. INTRODUCTION

Wireless Sensor Network (WSN) is an interconnection of sensing nodes that can come in different sizes connecting to a base station that makes meaning out of the data received. Being one of the highly utilized network types and applied in many areas such as health care monitoring, area monitoring, earth and environment sensing including industrial monitoring. These sensors monitor the environment, collecting data and sending to the base station. WSN can be used in an environment that is physically without protection or attended to [1]. The peculiar nature of wireless sensor networks makes them vulnerable to security threats of different types and purposes. With the simplicity of their routing techniques, security is the greatest challenge hence making them more susceptible to many network attacks, some of which are Sinkhole attack, Selective Forwarding attack, Wormhole, Hello Flood, sybil attacks, attack node replication, and Blackhole attack[2]. Sinkhole attack is one of the most formidable as it can lead to everyone of the other attacks mentioned above. Figure 1 depicts a typical WSN networks and all it comprises of.

WSN Uses

Sensor networks are generally sent in an assortment of utilization going from military to natural and medicinal research. In numerous applications, for example, target following, war zone reconnaissance and gatecrasher recognition, WSNs regularly work in hostile and unattended situations. In this manner, there is a solid requirement for ensuring the detecting information and detecting readings. In remote situations, an enemy not exclusively can listen in the radio traffic, yet additionally can catch or interrupt the traded messages. In this way, numerous conventions and calculations don't just work in unfriendly conditions without having sufficient safety efforts. Subsequently, security ends up one of the significant concerns while structuring security conventions in asset compelled WSNs. A portion of the uses of WSNs are in war zone surveillance, medicinal services applications, natural observing, shrewd home and vehicular specially appointed systems (VANETs) and some more.

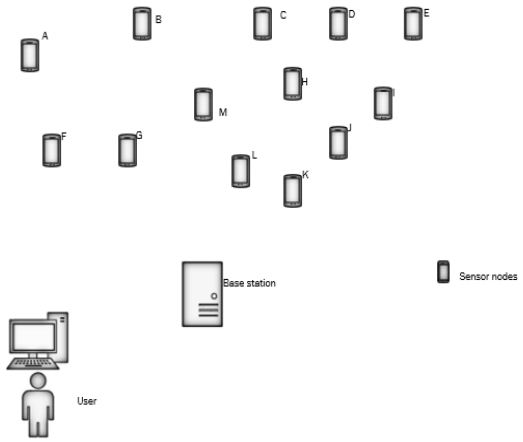


Fig 1. Wireless sensor Network

The major contribution of this conceptual framework paper :

- We describe wireless sensor network and showcase the structure of the WSN architecture

- We analyze and present the sinkhole attack scenario in WSN
- We propose an enhanced ant colony optimization technique (EACO) framework for sinkhole attack detection.

The purpose of our proposed paper is to develop a framework to improve sinkhole attack detection by Enhancing Ant Colony Optimization which a swarm intelligence technique using network simulator NS3 as the simulator. The remaining section of this framework paper is arranged as follows: section 2 shows the over view of sinkhole attack, section 3 gives a summary of recent related works, section 4 describes the problem formulation and the design model, section 5 presents the proposed framework for Enhanced Ant Colony Optimization (EACO) for sinkhole detection, section 6 introduces Ant Colony Optimization and the equations, section 7 shows the initial results describes the expected outcome of our model, while section 8 recommends future research direction and lastly, section 9 concludes the paper.

II. SINKHOLE ATTACK OVERVIEW

In Sinkhole attack a compromised sensor node attempts to get information to it from neighboring node. Thus, the sensor node picks up information, and knows what information is being communicated between neighboring nodes. This attack occurs by disguising the compromised node to be the most attractive to its neighboring node with respect to the algorithm by falsely advertising itself as the node closest to the base station. [3]. Figure 2 clearly illustrates a sinkhole attack in a wireless sensor network. As shown in figure 2, nodes with data seeking to transfer to the base station, first sends a route request (RREQ) to all nearby nodes. Normally the node with the quickest path to the base station will be known from the route reply (RREP) response coming from all the nearby nodes, this is where the sink node (SN) sends (RREP) reply too nodes no matter how far it is but pretends to have the quickest route to base station hence it attracts all data to itself and carries out any number malicious purposes like altering the information, dropping the information or dallying the information from getting to the base station giving birth to other attacks like blackhole, wormhole and grey hole in the wireless sensor network.

An attacker can compromise the integrity, confidentiality and authenticity of the wireless sensor network through a compromised node that becomes the sink node hence the wealth of research on ways to detect and mitigate such an attack on a wireless sensor network.

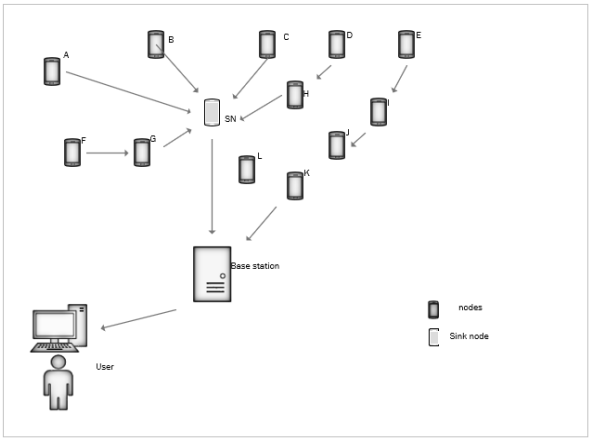


Fig 2. Sinkhole Attack Illustration

III. RELATED WORKS

Sinkhole attack detection and mitigation of in wireless sensor networks have been on for over a decade, the summary of the recent researches include[4]who addressed the problem of packet loss, delivery ratio and energy consumption by enhancing a swarm-based algorithm ABC to find compromised nodes using node ID's rule set comparison. [5] who presented parameter evaluation for ant colony system to obtain the best values for throughput, energy consumption and latency guiding research results to achieve optimal performance for packet routing process. [6] which introduced a detection algorithm using information from data aggregation algorithm to detect a sinkhole attacker in body area network, using omnet++ for simulation achieved a good performance in detection.[7] employed enhanced (ESPO) modifying flocking based clustering algorithm that involved separation, alignment and cohesion of clustered nodes in the WSN employed to mitigate and detect sinkhole attack in a larger instance.[8] proposed an algorithm enhancing ant colony system inspired from a variant of ant colony optimization with the global and local pheromone updates to improve path exploitation and exploration with the effect of reduction in packet loss and increase in energy efficiency of sensor nodes.[9] improved the (KMT) technique using ant colony optimization for path panning to deceive intruders, improve and safe guard data collection from node to base station and vice versa, using COOJA to simulate, taking into consideration ACO key protection allocation and ACO-pheromone vanishing mechanism effectively balancing the blending speed and communication of the nodes.[8] proposed an improved ant colony optimization algorithm to solve packet loss problems for nodes carrying more packets than its capacity. Experiment conducted to compare the performance of the proposed Enhanced Ant Colony System (EACS) using Energy Efficient Ant-Based Routing (EEABR) algorithm and Cost-aware Ant-Routing2 algorithm and the proposed algorithm is promising for implementation in static WSN systems.[10] proposed a flow based mitigation model with of time variant snapshot(FBSD) for sinkhole detection and mitigation. Using the geographical and physical features of the nodes, the base station maintains location details of the nodes enabling it to detect the presence of a sink node in the network. NS-2

simulator was employed for implementation, the proposed method highly reduces over-head generated by flood of control messages. [11] proposed a swarm intelligence algorithm, an Enriched Artificial Bee Colony Optimization (EABC) to observe and detect sink nodes in WSN, monitoring the position of estimated malicious nodes continuously. Implementing this algorithm as an evolutionary algorithm Using MATLAB for performance evaluation shows that with the intimation of the base station, the risk factor of the attacker node it known and hence the sinkhole discovered and its purpose.[12] proposed Artificial Bee Colony (ABC) technique for sink hole detection and compared it with an existing Enhanced Particle Swarm Optimization (EPSO) technique and got better detection rate, false alarm rate, packet delivery ratio and average delay. Using NS-2 as simulator. [13] designed a mechanism for sinkhole detection by first considering three types of sinkhole malicious node in WSN, the sinkhole modification node (SMD), sinkhole message dropping node (SDP), and sinkhole message delay node (SDL), then providing a detection scheme able to detect the different types of sinkhole node in a hierarchical wireless sensor network (HWSN). In this approach the network was divided into clusters with high sensing nodes used as the cluster head (CH) responsible for detecting sinkhole node in the cluster. Using NS-2 for simulation gaining better performance in terms of detection rate.[14] this authors proposed Enhanced Particle Swarm Optimization and the technique tested in a simulated environment. This technique proved to have a better performance than the initial techniques Particle Swarm Optimization (PSO) and Ant Colony Optimization in areas of packet delivery ratio, detection rate and average delay.

IV. PROBLEM STATEMENT

Sinkhole attack has an elusive nature. in a WSN, it can act as either a cluster head (CH) or the legitimate node closest to the base station. Hence there still lies the problem of false alarm rate in WSN, With the wealth of research on-going on the detection and mitigation of sinkhole attack in WSN because its malignant nature, there is room for improvement on solution of false alarm rate with the techniques.

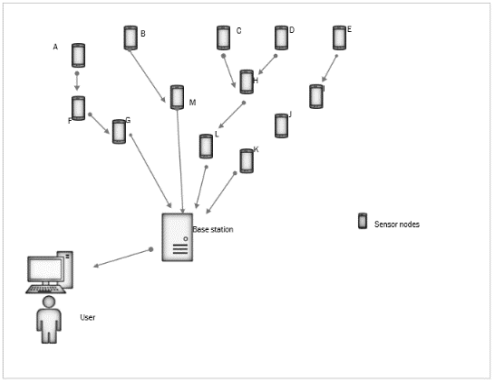


Fig. 3 Scenario I

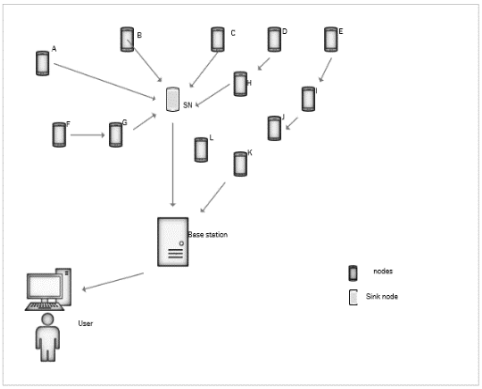


Fig 4. Scenario II

Scenario 1 shows a normal multi-hop many to one WSN node to base station packet delivery nature where the packets are transferred from one node to another based on the node is closest to the base station of the node with stronger link quality. While Scenario 2 shows the presence of an attacker node and it can be seen clearly that it has every packet passing through it giving the attacker the privilege to delay, modify or drop packets. In our proposed enhancement of the Ant Colony Optimization, the sink node is detected and no packet is sent through the compromised node.

V. PROPOSED ENHANCED ANT COLONY OPTIMIZATION FRAMEWORK FOR SINKHOLE DETECTION

This proposed framework comprises of first and second stages. The first stage involves problem formulation and planning, dataset description with design and stage two involves implementation which include Enhanced Ant Colony Optimization (EACO) detection on NS-3.29 simulator, flowchart and pseudocode.

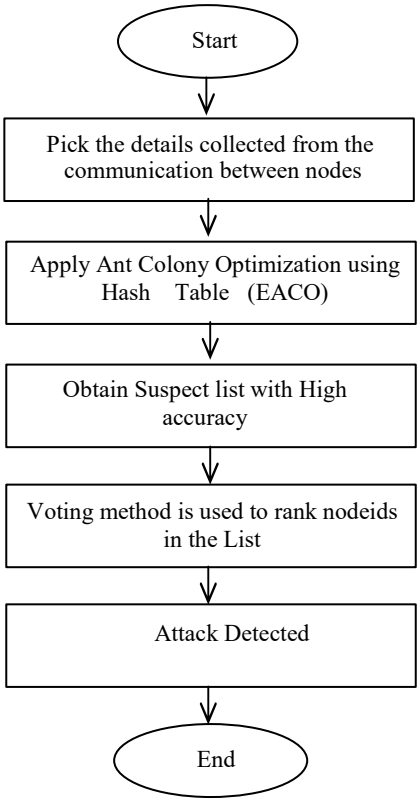


Fig. 5 Flow chart for ACO

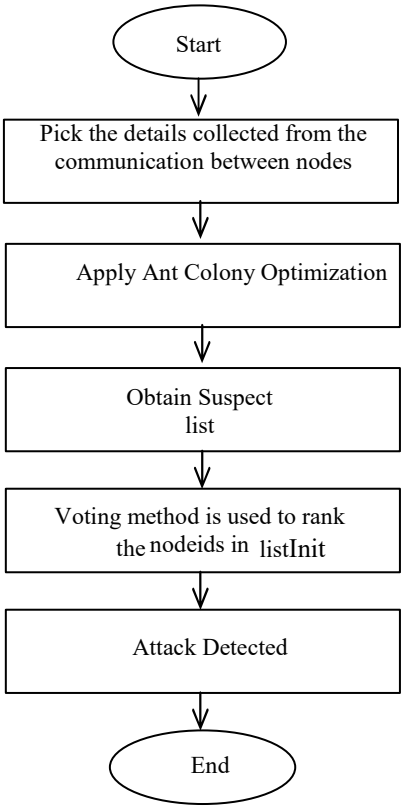


Fig. 6 Flow Chart for Proposed EACO

VI. ANT COLONY OPTIMIZATION

ACO algorithm is derived from the behavior of ants when they search for food, [15-18] derived the initial model of the ACO Algorithm. ACO an intelligent algorithm a type of swarm intelligence technique used to simulate food search process by ants. Ants secrete pheromones as they search for food to help bring them back to their initial location and to help discover the likely fastest distance that will bring success. Since ants have the ability to feel the amount of pheromone, as one ant finds food, others will surely follow the path which it paves to find the same food through the secreted pheromone. Ants always follow paths with high pheromone concentration that hence increases the pheromone secreted on the path which signal success. in Theory, the more the amount of pheromone there are on the path, the more ant connect and follow. The path having the highest amount of pheromone soon becomes the optimal path for the ant colony to the food location.

Formulated model of ACO according to [16] equation below

$$P_{ij}^{k} = \frac{[\tau_{ij}(t)]^{\alpha}.[\delta_{ij}]^{\beta}}{\sum_{k\in[N-tabuu_k]}[\tau_{ij}(t)]^{\alpha}.[\delta_{ij}]^{\beta}} \quad [1]$$

$$\tau_{ij}(t+1) = P * \tau_{ij}(t) + \sum_{k=1}^{N} \Delta\tau_{ij}(t)^{k} \quad [2]$$

$$\Delta\tau_{ij} = \frac{Q}{l_k(t)} \quad [3]$$

$$\delta_{ij} = \frac{1}{d_{ij}} \quad [4]$$

$$(X_{it} - X_{ii+1})^2 + (Y_{it} - Y_{it+1})^2 = d_{ii+1,it^2} \quad [5]$$

Q being constant, P_{ij}^{k} is probability of following node $d_{i_{t+1}}$ is distance between node l_{i+1} while the end node l_t , the $\tau_{ij}(t)$ is amount of pheromone in the curve (i,j). (x_{i_t}, y_{i_t}) and $(x_{i_{t+1}}, y_{i_{t+1}})$ are respectively coordinates l_t and l_{i+1} . δ_{ij} is the heuristic data in curve. α and β are weight factors of $\tau_{ij}(t)$ and δ_{ij}. where N is number of ants, $L_k(t)$ shows the object function.

Once an ant selects the next node location, the roulette wheel technique is used, which continues till the next node destination is obtained by the ants. Pheromones of all the nodes at once are restructured equation according to [16]

$$\tau_{ii,ii+1}(k, N+1) = \rho\tau_{ii,ii+1}(k, N) + \Delta\tau_{ii,ii+1}(k, N) \quad [6]$$

ρ represents the rate evaporation of pheromones

Procedure ACO ( )
{
Input n, α , β , ρ
set the ant colony configuration
set the initial pheromone and heuristic value
get ant colony optimization system based on the calculated

```
cost matrix
i = 1
while (I <= n)
{                    r = 1
While ( r <= i)
{
Reset the ants
Build ant s' solution
initiate local search
Update path best for i
Update pheromones
r = r + 1
}
Choose  path best for i
i = i + 1
}
}
```

Pseudo code for ACO

VII. INITIAL RESULTS AND EXPECTED OUTCOME

To show the workability of our framework, a WSN simulation is set up in NS-3.29 using 22 nodes. It supports simulations of TCP and UDP, MAC layer protocols, multicast and routing protocols in Wireless Sensor Networks. Simulation parameter. In this simulated network, standard routing protocol AODV is used. Number of nodes in the network are 22. The Number of sinkhole attacks varies with the total number of nodes observing there energy levels as they communicate until a sinkhole attack is observed.

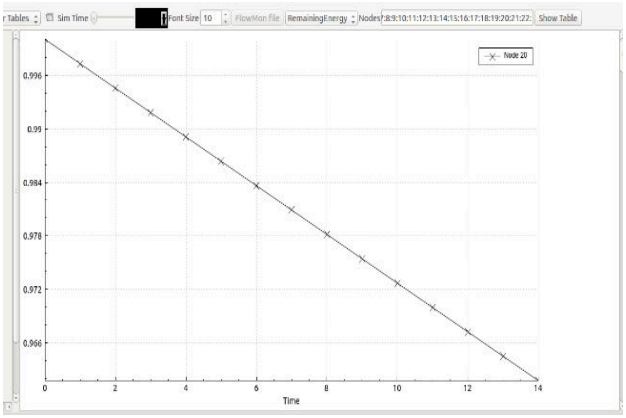


Fig 5. Energy change rate

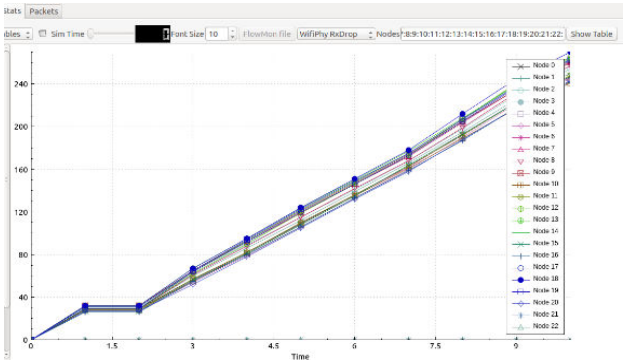


Fig 6. Packet drop ratio

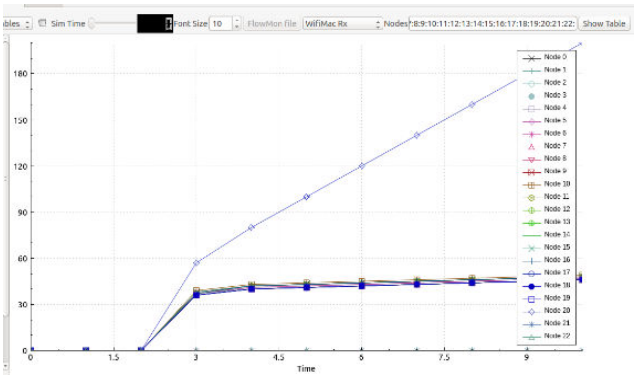


Fig 7. Throughput rate

From the simulation, introduction of the sink node changed throughput, energy level, packet delivery ratio of the nodes.

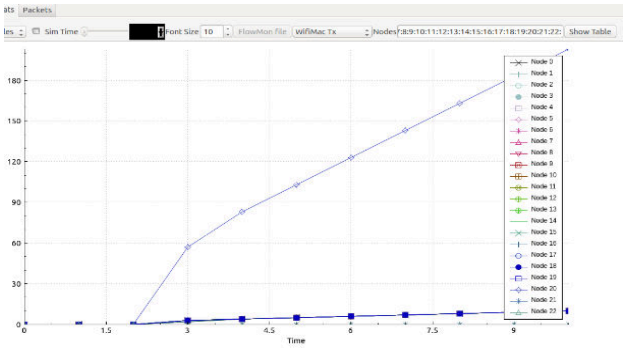


Fig 8. Packet delivery rate

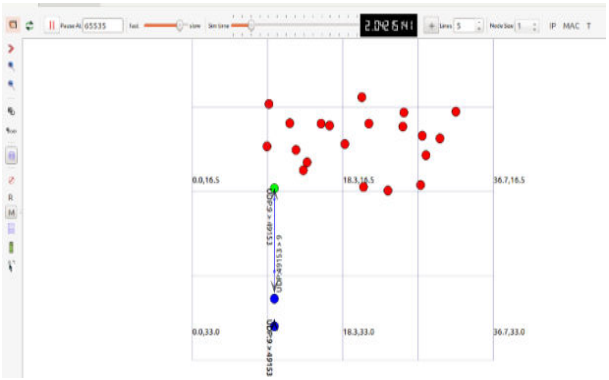


Fig 9. WSN simulation

VIII. FURTHER RESEARCH

With the framework on ground, the future direction is to proceed to apply the EACO algorithm in the wireless sensor network simulation in the NS-3.29, taking note on the effect of the detection rate and false alarm reduction rate and then compare with other known algorithm.

IX. CONCLUSION

We discussed wireless sensor network, highlighting its uses in the world around us and focused on one of the most malignant security challenges sinkhole attack. Presented a framework of our proposed enhancement of the Ant Colony Optimization algorithm to improve sinkhole detection rate and reduce false alarm in wireless sensor network. Presented initial results and future research direction.

X. ACKNOWLEDGEMENT

Our acknowledgement and appreciation are to every staff in the department of computer science and cyber security department, Federal University of Technology Minna.

REFERENCE

[1] G. Sanjeev Kumar and S. Poonam, "Overview of Survey on Security of Wireless Sensor Network," vol. 3, no. 1, pp. 5201–5207, 2013.

[2] S. Mohammadi and H. Jadidoleslamy, "A Comparison of Link Layer Attacks on Wireless Sensor Networks," *Int. J. Appl. Graph Theory Wirel. Ad Hoc Networks Sens. Networks*, vol. 3, no. 1, pp. 35–56, 2011.

[3] C. Karlof, N. Sastry, and D. Wagner, "TinySec: A link layer security architecture for wireless sensor networks," in *Proceedings of the 2nd international conference on Embedded networked sensor systems - SenSys '04*, 2004.

[4] N. Nithiyanandam and P. Latha, "Artificial bee colony based sinkhole detection in wireless sensor networks," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2019.

[5] H. Jamal, A. Nasir, and K. R. Ku-mahamud, "How to cite this article: Nasir, H. J. A., Ku-Mahamud, K. R., & Kamioka, E. (2019). Parameter adaptation for ant colony system in wireless sensor network.," vol. 2, no. 2, pp. 167–182, 2019.

[6] A. Nadeem and T. Alghamdi, "Detection Algorithm for Sinkhole Attack in Body Area Sensor Networks using local information," no. September 2018, 2019.

[7] N. Nithiyanandam, P. L. Parthiban, B. Rajalingam, and R. Scholar, "Effectively Suppress the Attack of Sinkhole in Wireless Sensor Network using Enhanced Particle Swarm Optimization Technique," vol. 118, no. 9, pp. 313–329, 2018.

[8] H. J. A. Nasir, K. R. Ku-Mahamud, and E. Kamioka, "Enhanced Ant Colony System for Reducing Packet Loss in Wireless Sensor Network," *Int. J. Grid Distrib. Comput.*, 2018.

[9] C. Iwendi, Z. Zhang, and X. Du, "ACO based key management routing mechanism for WSN security and data collection," *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 2018-Febru, pp. 1935–1939, 2018.

[10] K. Devibala, S. Balamurali, A. Ayyasamy, and M. Archana, "Flow Based Mitigation Model for Sinkhole Attack in Wireless Sensor Networks using Time-Variant Snapshot," *Int. J. Adv. Comput. Electron. Eng.*, vol. 2, no. 5, pp. 14–21, 2017.

[11] R. S. Raghav, S. Pothula, and D. Ponnurangam, "An enriched artificial bee colony (EABC) algorithm for detection of sinkhole attacks in Wireless Sensor Network," *Int. J. Mech. Eng. Technol.*, vol. 8, no. 8, pp. 193–202, 2017.

[12] Y. N. Dharshini and c N. Chinnaswamy, "SWARM INTELLIGENCE TECHNIQUE FOR SINKHOLE ATTACK DETECTION IN WIRELESS SENSOR NETWORK Performance Comparison of the Algorithms," no. 4, pp. 647–656, 2017.

[13] S. K. and M. K. K. Mohammad Wazid, Ashok Kumar Das, "Design of sinkhole node detection mechanism for hierarchical wireless sensor networks," *Wiley Online Libr.*, vol. 9, no. 19, pp. 4596–4614, 2016.

[14] G. Keerthana, "Detecting Sinkhole Attack in Wireless Sensor Network using Enhanced Particle Swarm Optimization Technique ," *Int. J. Secur. Its Appl.* , 2016.

[15] M. Dorigo and D. C. Gianni, "Ant Colony Optimization: A New Meta-Heuristic." 1992.

[16] J. Zhao, D. Cheng, and C. Hao, "An Improved Ant Colony Algorithm for Solving the Path Planning Problem of the Omnidirectional Mobile Vehicle," *Math. Probl. Eng.*, vol. 2016, pp. 1–10, 2016.

[17] T. M. Ibrahim et al., "Recent Advances in Mobile Touch Screen Security Authentication Methods : A," Comput. Secur., no. April, 2019.

[18] M. Shafie et al., "A Review on Mobile SMS Spam Filtering Techniques," IEEE Access, vol. 5, pp. 2169–3536, 2017.

Application of Filter Bank Multicarrier Offset Quadrature Amplitude Modulation to Large MIMO Systems: A Review

Mahmud Ja'afar, S. M. Sani, A. M. S. Tekanyi
Department of Electronics and Telecommunications Engineering, Ahmadu Bello University Zaria
Email: *emjaysani009@gmail.com, smsani@abu.edu.ng, amtekanyi@abu.edu.ng

***Abstract*-- The present generation of wireless communications systems, to a greater extent, do not optimally utilize the scarce radio spectrum and also not flexible in the allocation of the time-frequency resource. This may also be the case with the next generation systems if a viable solution is not found. Many researchers worked on resolving these two problems but suffered some limitations in their results, tools approaches and techniques used. The main problem was that they failed to collectively utilize the advantages of both the high spectral efficiency of Filter Bank Multicarrier Modulation Offset Quadrature Amplitude Modulation (FBMC-OQAM) and the efficient capacity property of large Multiple Input Multiple Output (MIMO) systems. In this paper, we shall propose to use the advantages of both systems in order to transmit FBMC-OQAM data symbols over a large MIMO system by employing a specific Space Time Transmission Scheme (STTS) of FBMC-OQAM for large MIMO. The transmission, however, will be prior to error correction using Hadamard scheme to double the spectral efficiency by restoring complex orthogonality in the imaginary field. The STTS will eliminate imaginary intrinsic interference, thereby enabling the efficient transmission of the FBMC-OQAM data over the large MIMO system. By employing the Hadamard scheme and STTS, we propose to develop a spectrally efficient large MIMO/FBMC-OQAM system with an improved capacity gain over a conventional MIMO/FBMC-OAQM.**

***Index Terms*-- Filter-bank multicarrier, FBMC-OQAM, Hadamard scheme, Imaginary intrinsic interference, large MIMO, multicarrier modulation, multipath propagation, OFDM, space-time transmission scheme.**

I. INTRODUCTION

The high demand for efficient and effective data transmission in wireless communication calls for better modulation techniques for a transmission[1] over a large Multiple Input Multiple Output (MIMO) system [2]. Orthogonal Frequency Division Multiplexing (OFDM) is presently the dominant transmission method for wireless systems. Filter Bank Multicarrier with Offset Quadrature Amplitude Modulation (FBMC-OQAM) which is based on OFDM, offers higher data and is more spectrally efficient because of its little out of bound emissions [3]. It also increases the capacity gain while maintaining the use of the scarce radio spectrum and it is also robust to multipath fading [4]. These among other reasons, make it applicable in the present and even in the next generation wireless communication systems [5]. FBMC-OQAM is more sensitive to imaginary intrinsic interference and it is also incompatible with MIMO systems because of non-orthogonality of the subcarrier symbols in the imaginary field [6]. The imaginary intrinsic interference complicates the application of MIMO and schemes such as Space-Time Block Coding (STBC) [3, 7]. The effect of the intrinsic imaginary interference can be subdued by spreading symbols in time and frequency domain with a Hadamard Matrix which enables the restoration of complex orthogonality using the STBC to effectively incorporate MIMO with FBMC-OQAM [6]. The application of multiple antennae on a common time-frequency spectrum for huge data transmission at high rate is achieved through MIMO system by employing STBC for symbols transmission [8]. STBC needs huge amount of transmit antennae in order to achieve high data rate and full spatial diversity. Large MIMO, otherwise called massive MIMO requires special transmission scheme [2]. For omnidirectional signaling with spatial diversity in large MIMO, an omnidirectional STBC is designed [9]. A small orthogonal STBC is designed for large MIMO of hundreds of antennae in Rayleigh fading environment [10], while OSTBC for 32 antennae is employed for real symbol transmission [11]. For large MIMO with huge arbitrary number of antennae, STBC offers high data rate only at the expense of symbol decoding complexity, which is impractical for huge number of antennas. Hence, STBC attains full spatial diversity at the expense of data rate [2]. Space-Time transmission Scheme (STTS) for large MIMO is employed to achieve high data rate with low complexity decoding by sacrificing diversity order according to the required quality of service [2]. Multicarrier systems in MIMO improve spectral efficiency link capacity. Hence, the combination of MIMO with FBMC system enhances the overall system performance and achieve better throughput [12]. This review proposes to employs Hadamard technique and STTS in order to develop a

spectrally efficient system of large MIMO with FBMC-OQAM components.

II. Multipath Propagation

In a communication channel, multipath propagation is a process in which a radio signal takes multiple routes to reach the receiving antenna from the transmitter antenna [13]. Multipath propagation is caused by ionospheric refraction and reflection, atmospheric ducting, reflection from surfaces like terrestrial objects and water bodies when there is multiple dispersions of the transmitted signal [14].

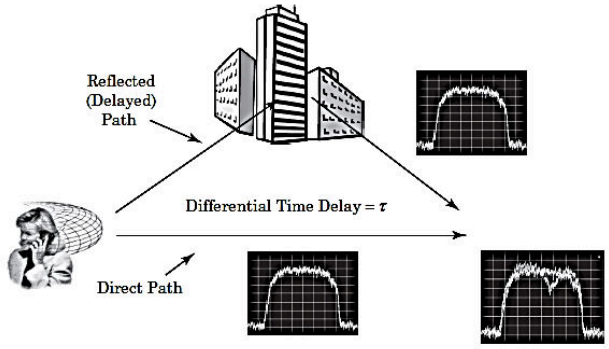


Figure 1: Signal Multipath Propagation Diagram [13]

The signal ray which propagates along a direct path mostly has the highest strength at the receiving antenna. The model equations of the direct ray [$x_D(t)$] and any of the indirect rays $x_{R,N}(t)$ are given [13] as:

$$x_D(t) = \cos(\omega_0 t) \tag{1}$$

$$x_{R,N}(t) = \rho_n \cos(\omega_0 t - \omega_0 \tau) \tag{2}$$

where, ω_0 is the angular frequency, ρ_n is the difference in amplitude between the direct ray and the reflected ray and is a real number, τ is the time difference between the direct ray and the instantaneous reflected ray.

The received signal, y(t) is the summation of all the signal components [13] and is expressed as follows:

$$y(t) = \sum_{n=0}^{N-1} \rho_n e^{j\phi_n} x(t - \tau_n) \tag{3}$$

where, $e^{j\phi_n}$ is a complex exponential component of the amplitude, τ_n is the time difference between the direct ray and the n[th] reflected ray.

III. Multicarrier Modulation

Multicarrier modulation is a technique for fragmenting data into several components for transmission over separate carrier signals of narrow bandwidth [15]. Multicarrier modulation is to some extent immune to multipath fading and Inter-Symbol Interference (ISI) and also less affected by noise, but under marginal condition, synchronization of the carriers is difficult [15]. The general model of a time continuous multicarrier system [16] is given as:

$$x(t) = \sum_{k=1}^{K} \sum_{l=1}^{L} g_{l,k}(t) x_{l,k} \tag{4}$$

where, $g_{l,k}(t)$ is the transmit basis pulse and $x_{l,k}$ is the data symbol at subcarrier position l and time position k. $g_{l,k}(t)$ modulates $x_{l,k}$ and it is basically the time and frequency shifted version of the transmitted prototype filter, $P_{TX}(t)$ and is expressed as:

$$g_{l,k}(t) = P_{TX}(t - kT) e^{j2\pi l(F(t-kT)} e^{j\theta_{l,k}} \tag{5}$$

Received basis pulse, $q_{l,k}$ is the time and frequency shifted version of the received prototype filter $P_{RX}(t)$ and it is similarly expressed as:

$$q_{l,k}(t) = P_{RX}(t - kT) e^{j2\pi l(F(t-kT)} e^{j\theta_{l,k}} \tag{6}$$

where, T is the time spacing, F is the frequency spacing and θ is the polar angle.

Subcarrier spacing in time domain (T) and in frequency domain (F) determine the spectral efficiency of the system. A schematic depiction of flexible resource allocation in time and frequency spacing is shown in Figure 2:

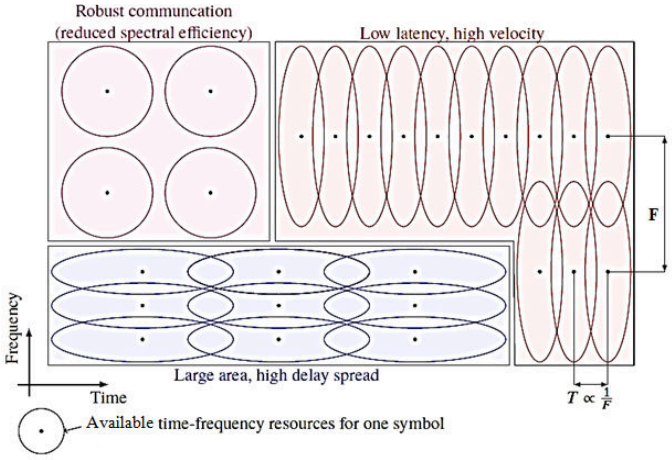


Figure 2: Time-Frequency Resources [17].

The major parameters that describe multicarrier systems are the prototype filters, $P_{TX}(t)$ and $P_{RX}(t)$, time spacing (T), and frequency spacing (F). the relationship of these parameters is given in an ambiguity function $A(\tau, v)$ as follows [18]:

$$A(\tau, v) = \int_{-\infty}^{\infty} P_{TX}(t - \frac{\tau}{2}) P_{RX}^{*}(t - \frac{\tau}{2}) \tag{7}$$

where τ is the time delay and v is the Doppler frequency. The ambiguity function, $A(\tau, v)$ at subcarrier position, l and time position, k is denoted by $A(T(k_1 - k_2), F(l_1 - l_2))$.

The ambiguity function determines the SIR projection and also projects $g_{l,k}(t)$ onto $q_{l,k}(t)$ in order to have an intuitive description of time and frequency offset [18]. Multicarrier system is characterized by some key limitations which are captured in the Balian low theorem. The theorem states that it is mathematically impossible for all the following four features to be satisfied at the same time in a multicarrier system [19]:

i. Maximum symbol density, so that TF equals 1

ii. Time localization, where:

$$\sigma_t = \sqrt{\int_{-\infty}^{\infty} (t-\bar{t})^2 / p(t) /^2 dt} < \infty \quad (8)$$

iii. Frequency localization, where:

$$\sigma_f = \sqrt{\int_{-\infty}^{\infty} (f-\bar{f})^2 / p(t) /^2 dt} < \infty \quad (9)$$

iv. Bi-Orthogonality, where:

$$\langle g_{l_1,k_1}(t), q_{l_2,k_2}(t) \rangle = \delta_{(l_1-l_2),(k_1-k_2)} \quad (10)$$

where δ is a Kronecker delta function describing orthogonality of the transmit and receive basis pulses.

A. Orthogonal Frequency Division Multiplexing (OFDM)

OFDM is a digital signal modulation technique of encoding, in which a single data stream is split over several narrowband channels which have varying orthogonal frequencies [20]. Conventional FDM and OFDM are conceptually alike but the orthogonality in OFDM makes it more spectrally efficient and allows the perfect propagation of multiple data signals across the same channel with a flawless detection [15]. The baseband discrete signal of an OFDM system is modeled as follows [21]:

$$x(i) = \frac{1}{N} \sum_{n=0}^{N} \sum_{m=-\infty}^{\infty} S_{m,n} g(i-mN) e^{j\frac{2\pi}{N}ni} \quad (11)$$

where n is the subcarrier index, $S_{m,n}$ is the quadrature modulated symbol which is the data transmitted on the n^{th} subcarrier of the m^{th} OFDM symbol, N is the total number of subcarriers in the OFDM symbol, $1/N$ is introduced for power normalization ,and g is the rectangular window function that separates the sub channels with their time domain coefficients.

OFDM channel is more spectrally efficient because the signals from various channels are combined into a single multiplexed stream of data to be transmitted by multiple orthogonal subcarrier symbols.

Despite the advantageous features of OFDM, the rectangular pulse shaping of the Fast Fourier Transform (FFT) filters results in a high spectral leakage. At the expense of spectral efficiency, Out of Bound Emissions OOBE is avoided using guard bands in-between adjacent frequency bands. The conventional OFDM is also highly susceptible to channel frequency selectivity [22]. Enhanced versions of OFDM such as: windowed OFDM, filtered OFDM, OFDM with redundant transmitter symbols (cyclic prefix), FBMC-QAM, and FBMC-OQAM are seen to be more efficient and to have better spectral efficiency. In terms of symbol density, time localization, frequency localization, and bi-orthogonality, FBMC-OQAM has the best performance, which is outlined in Table 1.

Table 1: Comparison of Different OFDM Schemes (Schwarz et al., 2017)

Schemes	Maximum Symbol Density	Time Localization	Frequency Localization	Bi-Orthogonality	Independent transmit Symbols
OFDM (without CP)	yes	yes	no	Yes	yes
Windowed/Filtered OFDM	no	yes	yes	Yes	yes
FBMC-QAM	no	yes	yes	Yes	yes
FBMC-OQAM	yes	yes	yes	Real only	yes
Block Spread FBMC-OQAM	yes	yes	Yes,	Yes, after de-spreading	no

B. Filter-Bank Multicarrier (FBMC)

FBMC is a multicarrier modulation technique that uses prototype filter. Prototype filter annuls inter-carrier interference and inter-symbol interference and is suitable for time and frequency localization, thus enabling the scheme to offers a better spectral efficiency [23]. It also offers high data rate transmission by reducing signaling overhead through the collection of non-adjacent bands in asynchronous transmission to acquire huge bandwidth [24]. Multi-user access in uplink transmission is also simplified due to a simplified synchronization requirements [25]

C. Filter-Bank Multicarrier Quadrature Amplitude Modulation (FBMC-QAM)

In the first level of FBMC-QAM systems, the transmit signal, S_{QAM} is obtained by mapping the information symbols to complex QAM symbols, and it is modeled [26] as follows:

$$S_{QAM} = \sum_{m=0}^{2m-1} \sum_{l=0}^{\infty} d_{m,l} g_{m,l}(t) \quad (12)$$

where, $d_{m,l}$ is the complex QAM symbol and $g_{m,l}$ is the information symbol at subcarrier index m and multicarrier index l. Throughput of the system is defined by the lattice/symbol density δ_{TF} which is the amount of

symbols transmitted per unit frequency and time. Figure 3 gives the FBMC-QAM time-frequency lattice and the relationship between δ_{TF}, Δf and Δt is given as follows [26] and it is expressed in $symbol/s/Hz$:

$$\delta_{TF} = \frac{1}{\Delta t \Delta f} \tag{13}$$

where, δ_{TF} is the lattice/symbol density, Δf is the change in frequency and, is the change in time Δt.

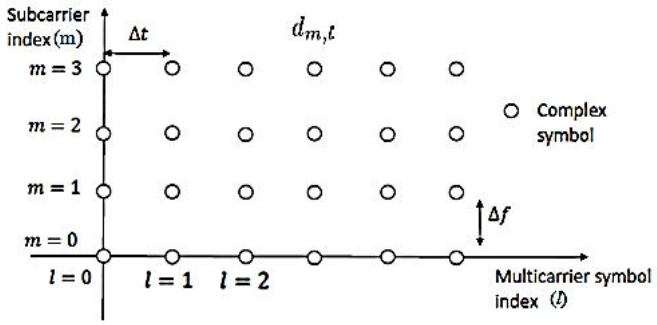


Figure 3: FBMC-QAM Time-Frequency Lattice [26].

As shown in figure 3, orthogonality in the imaginary field is sacrificed because the complex QAM symbols are exclusively real (represented by the symbol ◦). Information carried by QAM is relatively low due to large time and frequency spacing [26].

D. Filter-Bank Multicarrier Offset Quadrature Amplitude Modulation (FBMC-OQAM)

The operation principle of FBMC-OQAM is based on subcarrier modulation at a shift (offset) of half the symbol period between the imaginary and real components of a complex QAM, [27]. FMBC-OQAM is basically the same as FBMC with Cosine-modulated Multi-Time (FBMC-CMT), FBMC with Staggered Multi-Time (FBMC-SMT) and FBMC with Pulse Amplitude Modulation (FBMC-PAM) because of their similar working principle [24]. The block diagram of OFDM-OQAM transceiver is shown in figure 2.6 [12]:

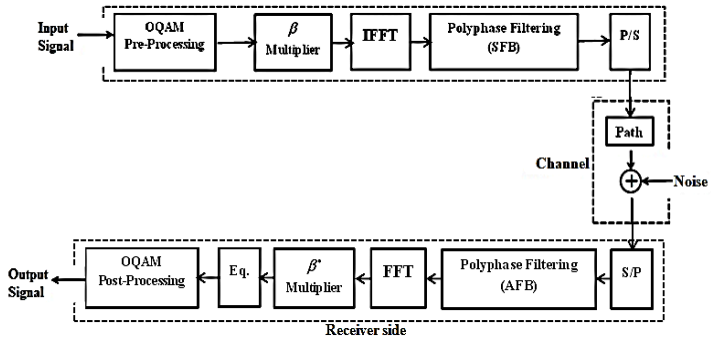


Figure 4: Block Diagram of FBMC-OQAM Transceiver [12].

Figure 4 shows the block diagram of FBMC-OQAM transceiver. The transmitter is composed of OQAM pre-processing, β multiplier, Inverse Fast Fourier Transformer (IFFT), synthesis filter bank (SFB) and Parallel to Serial (P/S) converter which is an up-sampler with delay chain. The receiver is composed of a delay chain with down sampler in form of Serial to Parallel (S/P) converter, Analysis Filter Bank (AFB), Fast Fourier Transformer (FFT), β^* multiplier, channel estimator, equalizer, and finally the OQAM post processor. Figure 2.6 shows the time-frequency symbol density of FBMC-OQAM system.

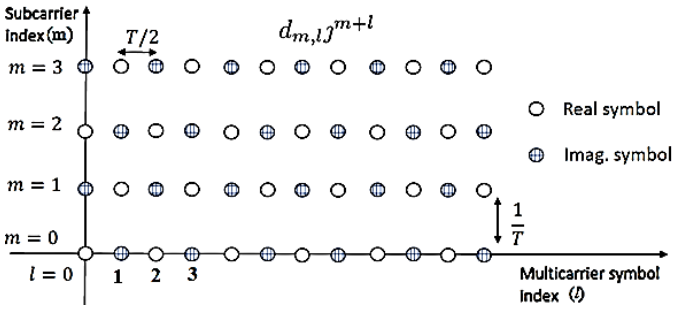


Figure 5: Time-Frequency Symbol Density of FBMC-OQAM System [26].

As shown in figure 2.6, a tight spacing of $T/2$ is observed which makes the symbol density twice as much of FBMC-QAM and due to the phase factor, j^{m+l}, the complex symbols ($d_{m,l}$) are also imaginary unlike in FBMC-QAM that are purely real. The transmit signal of FBMC-OQAM system is expressed [26] as:

$$s(t) = \sum_{m=0}^{2M+1} \sum_{l=0}^{\infty} d_{m,l} p_{m,l}(t) \tag{14}$$

with,

$$p_{m,l}(t) = j^{m+l} p(t - lT/2) e^{j2\pi mf\frac{1}{T}} \tag{15}$$

where, $p_{m,l}(t)$ is the OQAM modulated prototype pulse and $d_{m,l}$ is the complex OQAM symbol at subcarrier index m and multicarrier index l.

Since the complex symbols, $d_{m,l}$ are purely real, orthogonality of the system is one-sided and modeled as follows:

$$\Re\left(\int_{-\infty}^{\infty} p_{m,l}(t) p^*_{m_0,l_0}(t) dt\right) = \delta_{m-m_0,l-l_0} \tag{16}$$

where, $\Re$ is a real operator, $p^*_{m_0,l_0}$ is the conjugate complex of the OQAM modulated prototype pulse at the starting position, and $\delta_{m-m_0,l-l_0}$ is the two dimensional Kronecker delta at subcarrier position shift $m-m_0$ and multicarrier position shift $l-l_0$.

Spreading the symbols in time and frequency domain using suitable coding scheme will restore complex field orthogonality [26]. FBMC-OQAM is most preferred amongst OFDM based modulation schemes because of its higher spectral efficiency, low OOBE, and a maximum symbol density of $TF = 1$ [28]. From Table 1, it can be seen that by applying an appropriate spreading technique such as Hadamard spreading, STTC can then be

effectively utilized in order to enhance compatibility of FBMC-OQAM with Large MIMO.

IV. Error Correction and Interference Cancellation

Intense multipath propagation results in error and weakening of the communication quality. This is due to constructive and destructive interferences, with destructive interference resulting to the fading of the transmission channel [14], which can be mitigated by the following techniques[29]:

i. Error Correction Coding
ii. Channel Equalization
iii. Diversity transmission scheme

While channel equalization is beyond the scope of this review, error correction coding and diversity transmission scheme will discussed in this review with Hadamard coding as the preferred error correction coding scheme and MIMO as the diversity transmission scheme component.

A. Hadamard Coding

Hadamard codes otherwise called Walsh codes or Walsh Hadamard codes are the rows of a Hadamard matrix used as codewords [30]. Hadamard matrix is a square matrix containing only 1's and -1's. Each row of the matrix differs in exactly $N/2$ places from other rows making the entire rows to be composed of $N/2$ 1's and $N/2$ -1's, except for a single row which is entirely -1's [31]. The entire rows and columns are mutually orthogonal. The general representation of a Hadamard matrix H of n order $(size\ n \times n)$ is as follows [32]:

$$HH^* = nI_n \tag{17}$$

where I_n is an identity matrix of size n and H^* is the complex conjugate of H.

The matrix length (n) depicts the code length and every orthogonal row presenting a code. There is no cross correlation between codes which makes the orthogonality perfect. Hence, a 4×4 Hadamard matrix is as follows [31]:

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \tag{18}$$

B. Multiple Input Multiple Output (MIMO) System

MIMO system is a system that uses multiple transmit and receive antennae to exploit spatial diversity in multipath propagation in order to overcome issues such as interference, limited bandwidth, signal fading and low transmit power [33]. MIMO system enables the concurrent transmit and receive of multiple data signal over a common channel amidst multipath propagation to enhance the diversity order or received SNR [34].

With spatial multiplexing, MIMO enhance the data rate and channel capacity simply by increasing the antennae [34]. Consequently, higher throughput is attained with the improved channel capacity [33]. Figure 6 depicts a simple MIMO system [35]:

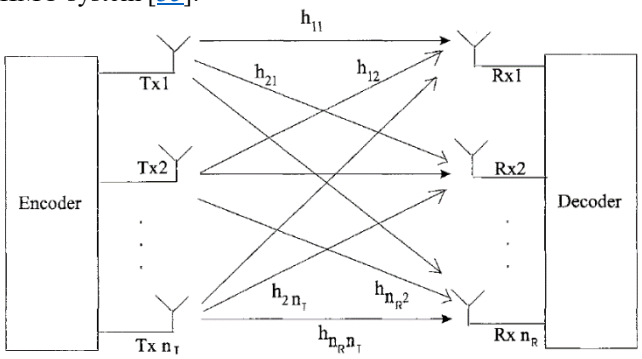


Figure 6: MIMO System [35].

The relationship of the transmit signal matrix $[T]$, the receive signal matrix $[Y]$, and the channel matrix $[H]$ is given by Daniel *et al*., (2017) as follows:

$$[Y] = [H][T] + noise \tag{19}$$

The transmitted signal of a MIMO system with (n_T) transmit antennae and (n_R) receive antennas during every symbol time slot, T_s is given by $n_T \times 1$ column vector, X. The elements of X are x_i for $i = 1, ..., n_T$, which is the transmitted signal at the i^{th} transmit antenna during a symbol time slot. Similarly, the received signal is given by $n_R \times 1$ column vector of y whose elements are y_j for $j = 1, ..., n_R$, at j^{th} receive antenna. The receiver obtains and decodes the signal vector X, noise vector n, and the channel matrix H. The system model for such a MIMO system is as follows [36]:

$$Y = HX + n \tag{20}$$

where y is the receive vector, H is the channel matrix, X is the transmit vector, and n is the noise vector.

The covariance matrix (R_{xx}) of X is given by Daniel *et al.,* (2017):

$$R_{XX} = E\{XX^H\} \tag{21}$$

where $E\{.\}$ is the expectation and $(.)^H$ is the Hermitian transposition operation. The total power, P of transmitted signals during the time slot, T_s is given by Daniel *et al.,* (2017):

$$P = tr(R_{XX}) \tag{22}$$

The channel is represented by the channel matrix which contains elements, h_{ij} called the channel coefficients between the i^{th} transmit antenna and the j^{th} receive antenna while noise at the receiver is given by $n_R \times 1$ column vector, n and having identical variance power, σ^2. If the average total receive power by each receive antenna (p_{rj}) is equal to the average total transmit power from n_T transmit antennae, SNR (ρ) at each receive antenna is modeled as follows [36]:

$$\rho = P/\sigma^2 \tag{23}$$

where, P is the total power of transmitted signals during the time slot T_s, and σ^2 is the variance power.

A large MIMO system which is also called large scale antenna system, massive MIMO, hyper MIMO or Full Dimension MIMO [37] is a multi-user MIMO technique in which a large number of antenna array elements at each base station are used to convey signal to a set of single antenna user terminals along a common time-frequency band [38]. Energy and spectral efficiency in large MIMO are improved with simple linear signal processing to increase power gain for each user terminal [39] and multiplexing gain in large MIMO is shared by all the users in order to cut cost of equipment [40].

Although additional antennas do not increase the training overhead, but as the number of antennae grows, especially in fast fading channel, channel estimation becomes more complicated. Thus, large MIMO is only possible in Time Division Duplex (TDD) systems, where channel reciprocity is exploited [41]. The number of scarce pilot sequence for a particular coherence time have to be reused in adjacent cells at the expense of pilot contamination, which is far greater than multiuser interference and noise [41].

Large MIMO enjoys all the benefits of conventional MIMO such as increased bandwidth, throughput and reliability, improved spectral efficiency, and reduced interference in a much greater scale. Figure 6 illustrates the structure of a massive MIMO system [37]

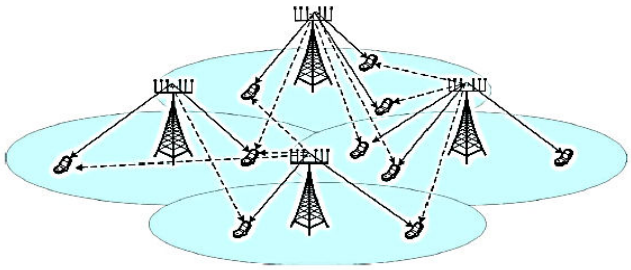


Figure 6: Structure of Massive MIMO system [37]

Huge capacity gain is achieved in MIMO using Space-Time Block Coding (STBC) transmission scheme by taking advantage of spatial diversity. The huge channel coherence time requirement of STBCs makes them unsuitable for large MIMO systems. As a result, Space-Time Transmission Scheme (STTS) which is resourcefully frugal and offers a better data rate is basically necessary for the large MIMO system [2].

V. Space-Time Transmission Scheme (STTC) for Large MIMO

Like STBC, STTC for large MIMO is also an error detection and correction scheme whereby multiple copies of a signal are transmitted for the purpose of reliability [42]. In both STBC and STTS, data are built as matrices. Unlike in STBC where antenna transmission over time is represented by each column and time slot is represented by each row [43], the representation is flipped in STTS [2].

Considering a MIMO system of equation (20), if the system has a large number of antennae N with N_t and N_r representing the transmitter and antennae respectively, $N/2$ time slots is needed to transmit the N complex symbols. The transmit vector, $X \in C^{N \times N/2}$ of the STTS (where C is a set of complex numbers) is in such a way that one symbol of every antenna is transmitted with its original sign in the first time slot and then the sign of every symbol are flipped in the subsequent time slots, except those of the two consecutive previous symbols of the two previous consecutive antennae as given in appendix A. The simplified version is as follows [2]:

$$x = \begin{pmatrix} x_1 & x_1 & -x_1 & \cdots & \pm x_1 \\ x_2 & x_2 & -x_2 & \cdots & \pm x_2 \\ x_3 & -x_3 & -x_3 & \cdots & \pm x_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_N & -x_N & x_N & \cdots & \mp x_N \end{pmatrix} \tag{24}$$

where rows represent transmit antennae and columns represent time interval. The receive vector, Y, channel matrix, H, and noise vector, n of the STTS model are defined as follows [2]:

$$Y = [y_K] \tag{25}$$

where, $y \in C^{1 \times N/2}$, for $k \in (1, \ldots, N/2)$.

$$H = [h_i] \tag{26}$$

where, $h \in C^{1 \times N}$, for $i \in (1, \ldots, N)$.

$$n = [n_j] \tag{27}$$

where, $n \in C^{1 \times N/2}$, for $j \in (1, \ldots, N/2)$.

VI. Imaginary Intrinsic Interference

Imaginary intrinsic interference occurs at the receiver end of the system's transceiver. It is the overlap of the filter response of two adjacent sub-channels in the imaginary field of complex valued symbols. It is caused by Multi-Stream Interference (MSI), ISI, and ICI and has a damaging effect to the system performance [26].

Imaginary intrinsic interference of subcarrier, m and multicarrier, l symbol (u_{m_0,l_0}) is represented as follows [26] :

$$u_{m_0,l_0} = \Im\Big(\sum_{m,l \neq m_0,l_0} d_{m,l} t_{m,m_o,l,l_0} \Big) \quad (28)$$

where, $\Im$ is an imaginary operator, $d_{m,l}$ is the complex symbol, and t_{m,m_o,l,l_0} is symbol duration.

Imaginary intrinsic interference cannot be simply avoided by receiver interference cancellation because of error propagation. As a result, STTS can only be effectively employed after the error has been corrected [7]. The error can be corrected using error correction coding, specifically Hadamard coding, whereby complex orthogonality is restored by spreading the symbols in time and frequency [44].

VII. Conclusion

The scarcity of radio spectrum in wireless communication necessitates the need for efficient schemes that ensures its optimal utilization, especially in the next generation of wireless communications systems. FBMC-OQAM being more spectrally efficient is a promising modulation technique when used over large MIMO transmission is hoped to offer a better data rate by attaining a maximum spatial diversity. However, the spectral efficiency of FBMC-OQAM is at the expense of imaginary intrinsic interference and it cannot be readily and efficiently transmitted over MIMO system. This research proposes to employ necessary techniques in the development of an effective and efficient FBMC-OQAM over large MIMO system that will address the interference limitation of the system.

References

[1] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems," in *Signal Processing Advances in Wireless Communications (SPAWC), 2015 IEEE 16th International Workshop on*, 2015, pp. 66-70.

[2] M. Arti, "A space-time transmission scheme for large MIMO systems," *IEEE Wireless Communications Letters,* vol. 7, pp. 62-65, 2018.

[3] R. Nissel, J. Blumenstein, and M. Rupp, "Block frequency spreading: A method for low-complexity MIMO in FBMC-OQAM," in *Signal Processing Advances in Wireless Communications (SPAWC), 2017 IEEE 18th International Workshop on*, 2017, pp. 1-5.

[4] V. Moles-Cases, A. A. Zaidi, X. Chen, T. J. Oechtering, and R. Baldemair, "A comparison of OFDM, QAM-FBMC, and OQAM-FBMC waveforms subject to phase noise," in *Communications (ICC), 2017 IEEE International Conference on*, 2017, pp. 1-6.

[5] S. Chaudhary and A. Patil, "Performance analysis of mimo-space time block coding with different modulation techniques," *ICTACT Journal on communication technology,* vol. 3, 2012.

[6] R. Nissel and M. Rupp, "Enabling Low-Complexity MIMO in FBMC-OQAM," in *GLOBECOM Workshops*, 2016, pp. 1-6.

[7] R. Zakaria and D. Le Ruyet, "Intrinsic interference reduction in a filter bank-based multicarrier using QAM modulation," *Physical Communication,* vol. 11, pp. 15-24, 2014.

[8] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Transactions on Information Theory,* vol. 47, pp. 1650-1656, 2001.

[9] X. Meng, X.-G. Xia, and X. Gao, "Omnidirectional STBC design in massive MIMO systems," in *Global Communications Conference (GLOBECOM), 2015 IEEE*, 2015, pp. 1-6.

[10] M. Arti, "OSTBC transmission in large MIMO systems," *IEEE Communications Letters,* vol. 20, pp. 2308-2311, 2016.

[11] M. H. Halmi, M. Abdellatif, and N. Nooridin, "Orthogonal space-time block codes for large MIMO systems," in *Communications, Management and Telecommunications (ComManTel), 2015 International Conference on*, 2015, pp. 78-82.

[12] Y. J. Harbi and A. G. Burr, "IIC of the MIMO-FBMC/OQAM system using linear and SIC detection schemes in LTE channel," in *Wireless Communications and Networking Conference (WCNC), 2018 IEEE*, 2018, pp. 1-6.

[13] K. McClaning and T. Vito, "Radio Receiver Design.(The Book End)," *Microwave Journal,* vol. 45, pp. 188-189, 2002.

[14] A. Mitra, "Multipath wave propagation and fading," *Lecture notes on mobile communication,* pp. 75-100, 2009.

[15] N. Malla and H. D. G. Joshi, "Study and Analysis of Carrier Frequency Offset (CFO) in OFDM," 2014.

[16] R. Nissel, E. Zochmann, and M. Rupp, "On the influence of doubly-selectivity in pilot-aided channel estimation for FBMC-OQAM," in *Vehicular Technology Conference (VTC Spring), 2017 IEEE 85th*, 2017, pp. 1-5.

[17] R. Nissel, S. Schwarz, and M. Rupp, "Filter bank multicarrier modulation schemes for future mobile communications," *IEEE Journal on Selected Areas in Communications,* vol. 35, pp. 1768-1782, 2017.

[18] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE signal processing magazine,* vol. 28, pp. 92-112, 2011.

[19] C. Thein, M. Fuhrwerk, and J. Peissig, "About the use of different processing domains for synchronization in non-contiguous FBMC systems," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 791-795.

[20] X. Jianyun, "STUDY OF MULTICARRIER MODULATION," *School of Electrical &*

Electronic Engineering Nanyang Technological University Nanyang Avenue, Singapore 639798, 2018.

[21] R. Zakaria, "Transmitter and receiver design for inherent interference cancellation in MIMO filter-bank based multicarrier systems," Conservatoire national des arts et metiers-CNAM, 2012.

[22] F. Rottenberg, P. Dat, T.-H. Nguyen, A. Kanno, F. Horlin, J. Louveaux*, et al.*, *2x2 MIMO FBMC-OQAM Signal Transmission over a Seamless Fiber-Wireless System in W-band* vol. PP, 2018.

[23] J. Du, P. Xiao, J. Wu, and Q. Chen, "Design of isotropic orthogonal transform algorithm-based multicarrier systems with blind channel estimation," *IET Communications,* vol. 6, pp. 2695-2704, 2012.

[24] A. Zafar, "Filter bank based multicarrier systems for future wireless networks," University of Surrey, 2018.

[25] D. Mattera, M. Tanda, and M. Bellanger, "Analysis of an FBMC/OQAM scheme for asynchronous access in wireless communications," *EURASIP Journal on Advances in Signal Processing,* vol. 2015, p. 23, 2015.

[26] F. Rottenberg, "FBMC-OQAM transceivers for wireless and optical fiber communications," UCL-Université Catholique de Louvain, 2018.

[27] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory," *IEEE transactions on signal processing,* vol. 50, pp. 1170-1183, 2002.

[28] R. Nissel, "Filter Bank Multicarrier Modulation for Future Wireless Systems," 2017.

[29] K. McClaning and T. Vito, *Radio receiver design*: Noble Publishing Corporation, 2000.

[30] M. Cheraghchi, V. Guruswami, and A. Velingker, "Restricted isometry of Fourier matrices and list decodability of random linear codes," *SIAM Journal on Computing,* vol. 42, pp. 1888-1914, 2013.

[31] K. Dostert, *Powerline communications*: Prentice Hall PTR, 2001.

[32] M. Malek, "Hadamard Codes," *California State University,* vol. 112, 2015.

[33] H. Daniel, S. Syed Aziz, S. Syed Ikram, and I. Umer, "MIMO Network and Alamouti STBC (Space Time Block Coding)," vol. 5, p. 42, 2017 2017.

[34] S. R. Chaudhary, A. J. Patil, and A. V. Yadao, "WLAN-IEEE 802.11 ac: Simulation and performance evaluation with MIMO-OFDM," in *Advances in Signal Processing (CASP), Conference on*, 2016, pp. 440-445.

[35] T. A. Wysocki, A. Mertins, and J. Seberry, *Complex Orthogonal Space-Time Processing in Wireless Communications*: Springer Science & Business Media, 2007.

[36] R. Shubhangi, Chaudhary and V. Ashwini, Yadao, "Capacity and Performance Analysis of MIMO-Space Time Block Code System " *IJECT,* vol. 6, p. 3, October, 2015 2015.

[37] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE communications magazine,* vol. 52, pp. 186-195, 2014.

Modified Handover Decision Algorithm in Long Term Evolution Advanced Network

A. S. Sada
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
absmsada86@gmail.com

Z. M. Abdullahi
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
eagbonehime1@gmail.com

A. D. Usman
Department of Electronics and Telecommunications Engineering
Ahmadu Bello University
Zaria, Nigeria
aliyuusman1@gmail.com

***Abstract*— Heterogeneous deployment of access points is considered as a promising technique for improved network performance in 4G networks. While the heterogeneity of the network results in reduced cell sizes thereby improving network coverage, inter-cell handover process becomes more complex and frequent. Therefore, the different velocities and angles of movement of the mobile users leads to excessive scanning process between neighboring cells and poses an inherent delay in the handover decision. In this paper, we present a modified handover decision algorithm whereby pedestrians and motorcyclists access the same small cell thereby reducing the scanning for different target small cells. Signal to interference plus noise ratio was considered to mitigate neighbor cell interference and introducing mobility vector prediction to accurately execute handover process. Probability of unnecessary handover and network throughput were used as metrics.**

Keywords— LTE-Advanced, Handover, HetNet

I. INTRODUCTION

In its quest to increase broadband cellular networks and bring mobile users close to access points, the third generation partnership project (3GPP) developed the Long Term Evolution Networks (LTE). This next generation of wireless cellular technology supports the heterogeneous deployment of access points. Heterogeneity of the networks is the mix deployment of femtocells, picocells and microcells overlaid by a macrocell which covers larger area.

Deploying small cells in a network coverage area comes with a lot of technical, financial and environmental challenges. Technical challenges include frequent cell exchanges as a user crosses a cell boundary, financial include the cost effectiveness associated with the deployment and technical staff for the services and finally, environmental hiccups comprise of location to fix the cells and adverse effects of pollution from power generating sets to power the cell site.

Third Generation Partnership Project (3GPP) in its indefatigable effort to further enhance broadband wireless cellular networks came up with the Long Term Evolution (LTE) network system. Also termed as Fourth Generation (4G) mobile communication standard [1].Classification of base stations are of two major categories. Low-power Home evolved Node B base stations are usually installed indoors while the evolved Node B base stations are installed outdoors to improve capacity and user throughput. LTE enables low-power base stations to connect to the Internet [2]. Handover procedure transfers active call or data session while a user is in motion from one cell to another [3].

Ref. [4] analysed an algorithm that utilizes the predictive values of Received Signal Strength (RSS) to achieve improvement in throughput and reduction in ping pong. The chosen base station was the one with the improved network throughput. Evaluation of the algorithm was based on the following: Outage Probability (OP), Ping Pong Rate (PPR), Throughput, Number of Handover (NHO) and errors associated with the prediction of the RSS.

A handover decision algorithm between femtocells was investigated by [5], using the RSS and user velocity as decision criteria. The algorithm chose a femtocell that is far away from the user equipment compared with other available adjacent femtocells. This idea suppresses unnecessary handovers before getting to the target femtocell. An Inter-femto cell handover scheme for dense femto cellular networks was proposed by [6]. Two algorithms were used to address the movement direction of the user; one for the prediction of the user mobility pattern, while the other for generating mobility rules. This paper proposes a reduced scanning process which translates to a corresponding reduction in the iteration process and reduced power dissipation of the user equipment by considering only two velocity thresholds which are low to medium speed users. Inter-cell interference was also mitigated by selecting users with high signal to interference plus noise ratio. Medium velocity threshold as classified in [7], was set and signal to interference plus noise ratio was incorporated into the mobility vector prediction technique in order to guarantee efficient handover procedure.

Organization of the remaining paper follows suite: Section II explains model. Proposed method was illustrated in Section III. And it concludes with Section IV.

II. SYSTEM MODEL

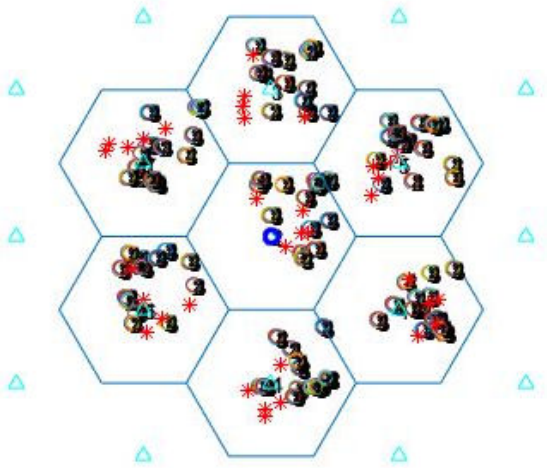


Figure 1: System Model

Consider a seven 7 hexagonal macrocells with an illustrative example depicted in 1. The macrocell area is densely populated with 100 small cells. The blue dot is the centre point, red stars are the user equipment, the black circles are the small cells, and other small cells that are fixed on lamp posts or buildings are indicated with brown colour. Blue triangles are the macrocell areas. The small cells overlaid by the macrocell provide the users with the network coverage in the simulation environment. Low to medium velocity users are covered by same small cell in order to reduce the scanning procedure for three different velocity thresholds. This reduction in the scanning and number of iterations save battery power dissipation of the user equipment and also reduce frequent handing over between small cells as the user is moving. The user equipment adopted a random waypoint model while roaming in the coverage area.

III. PROPOSED METHOD

In the work of [8]; low, medium and high speed users, angles of movement in addition to signal to noise ratio were considered for a successful handover process. Significant improvements are proposed by considering only low to medium users which is aimed at reducing the number of iterations and scanning for a neighbor cell for handover. More so, for improved network throughput; signal to interference plus noise ratio was incorporated to mitigate the effect of interference from neighboring cells. Finally, a mobility vector model was introduced to ensure precise location and accurate handover is executed to the user equipment. Mobility vector prediction technique remembers the previous location of a user equipment and is able to predict the next moving direction. And it supports flexibility in movement patterns of user equipment. Figure 2 illustrates angle of 30^0 at which the user is roaming in the coverage area which is Southwest (SW) movement to East (E).

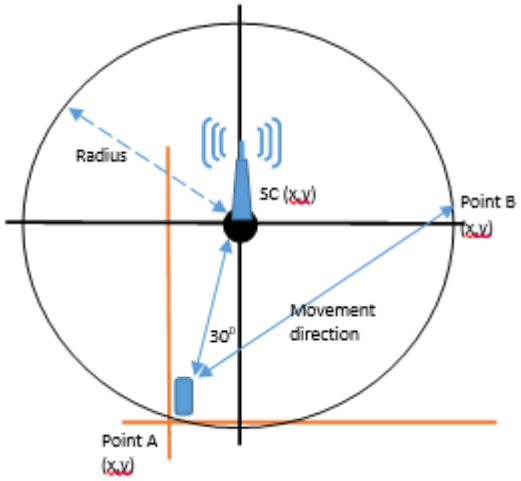


Figure 2: Angle of movement of the UE.

In 3GPP Release 11, three mobility states of users were prescribed and medium to high velocity thresholds used. For the reasons that low velocity moving users are reliably detected by the base station network and it has an ample time to complete the produces for a successful handover. Handover process is best executed before the serving cell's signal quality fades to unbearable degree. Macro-only networks come into play when medium to high velocity users could not detect reliable signal due to speed. Hence, which leads to unnecessary handover especially in massive deployment of low-power access points [9].

Mathematical models for the proposed method are explained as follows:

Set of Candidate small cells for $\boldsymbol{UE_y}$ in one $\boldsymbol{MC}$, is denoted as $\boldsymbol{S_{SC}}$, and can be defined by the following equation:

$$S_{sc} = \left\{ |SC_x \in N_{sc}| \left(d_{act}^{ue_y \to sc_x} \leq d_{th} \right) \wedge \left| \propto_{ue_{yx}} \right| \leq \propto_{xm,th} \right\} \quad (1)$$

The set of SCs are contained in the shortened candidate list from which the small cell with the maximum Signal to Interference plus Noise Ratio (SINR) is selected for handover.

Furthermore, Probability of Successful Handover for the proposed modification is mathematically expressed as:

$$P_{HO} = P_r[V_{ue_y} \leq 15KM/hr \wedge d_{act}^{ue_y \to sc_x} \leq d_{th} \wedge \left| \propto_{ue_{yx}} \right| \leq \propto_{xm,th} \wedge SINR_{sc_x \to ue_y}^{r} > SINR_{m \to ue_y}^{r} + MVP\ ]. \quad (2)$$

where

$V_{ue_y} \leq 15KM/hr$ HO threshold velocity at base station

V_{ue_y} = velocity of the UE

$d_{act}^{ue_y \to sc_x}$ = actual distance between the UE and the SC

d_{th} = distance threshold to form the SC list

$|\propto_{ue_{yx}}|$ = angle between the UEy and SCx

$\propto_{xn,th}$ = angle threshold at which the SCs are included in the candidate list

$SINR^r_{sc_x \to ue_y}$ = Signal to Interference plus Noise Ratio (SINR) measured at user y from SCx

$SINR^r_{m \to ue_y}$ = Signal to Interference plus Noise Ratio (SINR) measured at user y from Macrocell

MVP = Mobility Vector Prediction

The final signal to interference plus ratio for the proposed method is:

$$SINR_{x \to ue_z} = \frac{P^r_{xp \to ue_z}}{\sum_{y=1, x \neq y}^{n_y} \frac{P^t_{y \to ue_z} - g_y\, g_{ue_z}}{l_{o_y} l_{oue_z} \xi_{y \to ue_z} \delta_{y \to ue_z}} + \sigma^2} \quad (3)$$

$P^r_{xp \to ue_z}$ = pilot RSRP received from a target cell x at user z

$P^t_{y \to ue_z}$ = transmitting power of base station y

n_y = total number of interfering base stations

σ = noise power

g_y = antenna gain of base station y

g_{ue_z} = antenna gain of user z

l_{o_y} = base station y equipment loss

l_{oue_z} = user equipment loss z

$\xi_{y \to ue_z}$ = the shadow fading between interfering base station and user z

$\delta_{y \to ue_z}$ = path loss between the interfering base station y and user z

Each cell is assumed to be in a hexagonal shape to coverage holes if it is to be assumed as circular. Therefore, realistically cell borders depend on obstacles, terrains, interference and geographic area [10].

The Mobility Vector (MV) is a mobility scheme that provides a flexible mobility architecture for different motion patterns. The model supports partial changes in the mobility state of a user. It also memory dependent whereby it remembers previous state of user and predicts its next movement direction. Mobility of a user is defined by the sum of two vectors, mathematically expressed as [12]:

$$\vec{B} = (bx_v, by_v) \quad (4)$$

$$\vec{V} = (vx_v, vy_v) \quad (5)$$

where,

$\vec{B}$ is the base vector which defines the major direction and speed of the node and $\vec{V}$ is the Deviation vector that stores the mobility deviation from the base vector.

Therefore, the overall mobility vector from (4) and (5) is expressed as:

$$\vec{M} = \left(\vec{B} + \alpha \vec{V}\right) \quad (6)$$

where α is an acceleration factor.

The base vector is assumed to be either the macrocell or the candidate small cells, while the deviation vector is assumed to represent the user equipment. Thus, the base vector shows the current mobility state of the macro cell or the candidate small cell and the deviation vector defines the direction and speed at which the user equipment moves towards either the candidate small cell or the macro cell.

IV. CONCLUSION

In this paper we proposed a modified handover decision algorithm where mobility vector prediction for handover was incorporated and medium velocity threshold was set for low to medium speed users. We also proposed to mitigate the effect of inter-cell interference by considering signal to interference plus noise ratio. The metrics of probability of unnecessary handover and network throughput were used.

REFERENCES

[1] Jansen, T., Balan, I., Turk, J., Moerman, I., & Kurner, T. (2010). Handover parameter optimization in LTE self-organizing networks. Paper presented at the Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd. 1-5.

[2] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, et. al., "LTE Advanced: an operator perspective,"IEEE Communications Magazine, vol. 50, no. 2, pp. 104–114, February 2012.

[3] Li, Y., Cao, B., & Wang, C. (2016). Handover schemes in heterogeneous LTE networks: challenges and opportunities. IEEE Wireless Communications, 23(2), 112-117.

[4] Kalbkhani, H., Yousefi, S., &Shayesteh, M. G. (2014). Adaptive handover algorithm in heterogeneous femtocellular networks based on received signal strength and signal-to-interference-plus-noise ratio prediction. IET Communications, 8(17), 3061-3071. doi: 10.1049/iet-com.2014.0230.

[5] Rajabizadeh, M., &Abouei, J. (2015). An efficient femtocell-to-femtocell handover decision algorithm in LTE femtocell networks. Paper presented at the 2015 23rd Iranian Conference on Electrical Engineering (ICEE), Tehran. 213-218.

[6] Al-Shahin, F. A. (2015). Femtocell-to-Femtocell Handoff management in Dense Femtocellular networks. International Journal of Computer and Communication Engineering, 4(5), 346-353.

[7] Haddad, Z., Mahmoud, M., Saroit, I. A., & Taha, S. (2016, April). Secure and efficient uniform handover scheme for LTE-A networks. In Wireless Communications and Networking Conference (WCNC), 2016 IEEE (pp. 1-6). IEEE.

[8] F. Bai and A. Helmy, "A survey of mobility models," Wireless Adhoc Networks. University of Southern California, USA, vol. 206, 2004.

[9] Alhabo, Mohanad, L. (2017, February). Unnecessary handover minimization in two-tier heterogeneous networks. In wireless On-demand Network Systems and Services (WONS), 2017 13th Annual Conference on, (pp. 160-164). IEEE

[10] Zhou, Y., Lei, Z., & Wong, S. H. (2014, May). Evaluation of Mobility Performance in 3GPP Heterogeneous Networks. In 2014 IEEE 79th Vehicular Technology Conference (VTC Spring) (pp. 1-5). IEEE.

[11] Alhabo, M., Zhang, L., & Nawaz, N. (2017, May). A trade-off between unnecessary handover and handover failure for heterogeneous networks. In European Wireless 2017; 23th European Wireless Conference (pp. 1-6). VDE.

[12] Chu, X., López-Pérez, D., Yang, Y., & Gunnarsson, F. (Eds.). (2013). Heterogeneous Cellular Networks: Theory, Simulation and Deployment. Cambridge University Press. (BOOK).

Multi-Tenancy Cloud-Enabled Small Cell Security

Babangida Albaba Abubakar
Department of Mathematics and Computer Science
Umaru Musa Yar'adua University,
Katsina, Nigeria
Babangida.Abubakar@umyu.edu.ng

Haralambos Mouratidis
Centre for Secure, Intelligent and Usable Systems
University of Brighton
United Kingdom
h.mouratidis@brighton.ac.uk

Abstract— **The anticipated technological advancement of 5th Generation (5G) network is the ability to apply intelligence directly to network edge, in the form of virtual network appliances through the archetypes of Network Functions Virtualisation (NFV) and Edge Cloud Computing. The adoption and use of innovative technologies, such as Software Defined Networking (SDN) and NFV is the key to making 5G networks more promising. A Cloud-Enabled Small Cell (CESC) provides multi-operator platform to integrates and executes at the virtualised environment. However, implementing these technologies yield to the imaging of new security challenges. Providing services to multiple operators/tenants to access technologies and protocols in unified network architecture requires well-defined security approach in order to deliver secured data communication, privacy and integrity. In this research, CESC security requirements analysis was carried out using Secure Tropos (SecTro) methodology. The paper thoroughly examined the CESC security challenges and provides possible mitigation measures.**

Keywords—5G; CESC; Virtualization; NFV; SecTro; SESAME; Light Data Centre, NFVO, Northbound Interface

I. INTRODUCTION

The high increase in mobile data traffic that is driven by the demand of personalised applications and services is generating high demand in network infrastructure resources to fulfill the required end user experience. The technological advancement has been gearing in fulfilling these increasing demands and improves network coverage, capacity and resource utilization. The long awaited 5th Generation (5G) network framework of achieving high reliability, low latency and managing high applications and demands could be achieved through applications and network resources virtualisation for utmost utilization.

The main anticipated element of 5G technology is the ability to apply intelligence directly to the network edge, in the form of virtual network appliances through the archetypes of Network Functions Virtualisation (NFV) and Edge Cloud Computing. The 5G network infrastructures is envisaged to offer rich virtualisation and multi-tenant capabilities through on-demand processing capabilities, slicing network resources among multiple operators and service deployment close to the end users.

The adoption and use of innovative technologies, such as Software Defined Networking (SDN) and NFV is key to making 5G network more promising. However, implementing these technologies yield to the imaging of new security challenges. Providing services to multiple operators/tenants to access technologies and protocols in unified network architecture requires well-define security approach in order to deliver secured data communication, privacy and integrity.

The complicated security requirement for 5G network manifested not only due to the virtualisation, but also the different security requirements at different network domains. Therefore, an effective security management system is required to address these challenges. The security management system should be able to provide robust authentication and authorisation to control any abuse to the northbound interface controllers. Only authorised applications should have access to the controllers. Also, tenant isolation should be instigated at different network domain to guarantee confidentiality, security and integrity in the multi-tenant environment.

Although the virtualisation of the network infrastructure has been related to its applicability to the Small Cell infrastructures, it has so far received limited attention. Small Cells enhanced the cellular coverage, capacity and applications for homes, businesses, rural and dense metropolitan areas [1][2]. The main role of Small Cells is to offer services in density-populated areas such as stadiums, shopping malls and concert venues. The implementation and maintaining of Small Cell requires high investments of time and money. The investment includes site installation, power supply and establishing dedicated high-capacity backhaul connections. Others include the radio resource management, interference mitigation measures and security of the physical devices, which results to the high cost of maintenance.

A Cloud-Enabled Small Cell (CESC) is a new multi-operator concept that allows Small Cell to integrates and execute at the virtualised environment. The conception of this idea is the

deployment of Virtual Network Functions (VNFs), self-x management support and executing applications and services inside the access network infrastructure.

Small cEllS coordinAtion for Multi-tenancy and Edge services (SESAME), a Horizon 2020 European Commission project in advanced 5G Network Infrastructure proposes the use of network functions virtualisation, mobile-edge computing and cognitive management, to alleviate the high cost of provisioning Small Cells by single operator. One of the focus areas of the SESAME project is the use of cloud technology to virtualised Physical Small Cell (PSC) to provide telecommunication services to Mobile Network Operators (MNOs). The SESAME's main goal is the development and demonstration of an innovative architecture, capable of providing Small Cell coverage to multiple operators 'as a Service'. To achieve the envisaged aim, SESAME proposed virtualisation and partitioning of the localised Small Cell network and shares it to multiple tenants. Moreover, in addition to that, SESAME supports enhanced multi-tenant edge cloud services by enriching Small Cells with micro servers.

This paper is driven by the SESAME network architecture, illustrated in figure 1. The main SESAME innovation is the ability to accommodate multiple operators/tenants in the same infrastructure, through virtualisation of Small Cell and their utilisation. The resources are logically sliced and shared to each operator/tenant separately, depending on their requirements. The paper thoroughly examined the security challenges facing SESAME architecture and provides possible solutions to mitigate those challenges.

II. RELATED WORK

This section provides related work carried out in the area of multiple domain security and virtual environment. A security aware Virtual Machines (VMs) placement scheme for eliminating side channel attacks in the cloud was proposed[3]. The scheme uses an Aggressive Conflict of Interest Relation (ACIR) to describe the constraint relations for users, based on the Chinese wall policy. The scheme also uses forward isolation rules to formulate VMs placement/migration behaviour. The experimental results show that the scheme effectively ensuring isolation of VMs own by conflicting users from residing in the same host which eliminating side channel attack in the cloud. However, the results show that the proposed algorithm affects resource utilisation.

With the high increase in the use of cloud by different applications, security of those applications is a major concern. Virtual network security applications are deployed in the cloud to mitigate the security challenges in cloud environment. In a related work, an optimal algorithm for Placement of Sequentially Ordered Virtual Security Appliances in the Cloud was proposed[4]. The algorithm is to protect traffics from different appliances to avoid data snooping, considering the elastic nature of the cloud. For the traffic of each application to pass through the network security system in a sequential order, it is not only time consuming but requires computational network resource. The Managed Security Service Providers (MSSPs) have been adopted in many cloud infrastructure environment to provide security and privacy solution to the cloud. However, managing different MSSPs in a cloud could be challenging.

An Easy Security services Integration Cloud framework[5] is proposed to simplify the incorporation of multiple managed security services into a cloud environment. It provides client's library with rich APIs and a management console to MSSPs where multiple MSSPs can be integrated into the cloud platform. The proposed framework adapts the use self-analysis by VMs to monitor and analyse execution data. The information is shared among VMs to monitor activities of the guest VMs. However, this process consumes high computational and storage capacity, which is limited in cloud environment.

All the above research works addressed security challenges in single tenant virtual environment, thus the multi-tenancy capabilities for optimum resources utilization remain challenging. In this paper, the security challenges of multi-tenant virtual environment are looked into to and suggestions were made.

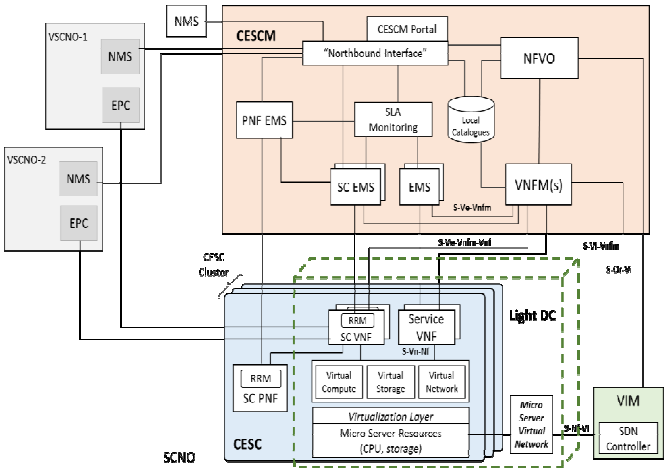


Figure 1: SESAME Architecture

III. SESAME NETWORK ARCHITECTURE

SESAME proposes the new concept of a 'Light Data Centre' (Light DC) to enhance the virtualisation capabilities. Light DC is designed to build a clustered infrastructure for easy manageability and provide friendly ecosystem through optimising power consumption, less cabling, space and cost. The Light DC provides cloud services within the network infrastructure and with combination of the NFV and SDN, it will bring to the reality of achieving a high level of flexibility and scalability. The Light DC is designed to support the execution of different types of VNFs that provides multiple connections from the Small Cells to each network. Figure 2 illustrates the SESAME high-level architecture. The CESC Manager (CESCM) is responsible of providing management and orchestration of the CESC, which include coordinating and supervising the users, the performance and the delivery of

services. It also controls the interactions between the CESC and the network operator, according to the Service Level Agreement (SLA). Also, CESCM monitors and analyse network performance based on the set Key Performance Indicators (KPIs) for effective network management.

The CESCM consists of Virtual Infrastructure Manager (VIM), VNF Manager, Network Functions Virtualisation Orchestrator (NFVO), The Northbound Interface and the Catalog. In brief, the VIM manages the interaction between VNF and resources such as compute, storage and network. While the VNF manager oversees the management of VNF lifecycle. The orchestrator is responsible for orchestration and management of NFV infrastructure as well as software resources and ensuring network services, while the Northbound Interface provides communication channel between the NFVO and the external elements.

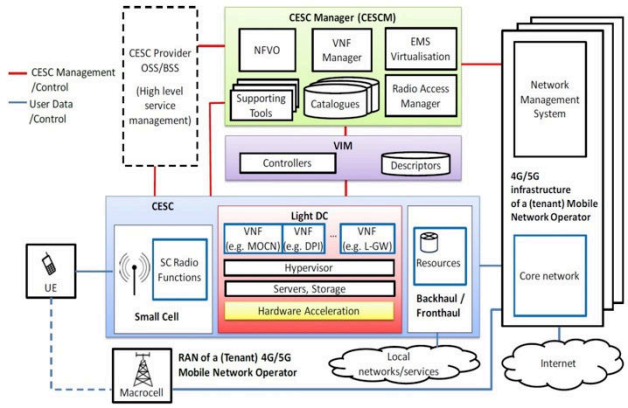


Figure 2: SESAME High Level architecture

IV. SESAME SECURITY REQUIREMENT ANALYSIS USING SECURE TROPOS

The SESAME security requirement analysis was carried out using Secure Tropos (SecTro) methodology. The SecTro is a security analysis modelling tool that allows user to model and analyse security requirements and objectives of a system. The main aim of this methodology is to support the capturing, analysing and identifying security requirement of a system at the early stage of the development process. It allows the system developers to analyse and identify potential security threats to the system at the early developmental stage. The SESAME system architecture was modelled and the security requirements were analysed and discuss in the rest of this section.

The major SESAME stakeholders (Actors) are described as follows:

(i) End User (EU): Is the consumers of communication services through cellular network. The EU could be a mobile device such as smart-phone, tablets or a laptop.

(ii) Infrastructure Owner (IO): Is the cellular infrastructure owners, the infrastructure could be the Small Cell or Micro Cell.

(iii) Venue Owner (VO): Is the owner of the venue such as mall, stadium, enterprise or municipality where the PSC are to be deployed.

(iv) Small Cell Network Operator (SCNO): Is a firm that possesses the radio communication equipment to provide access and services to the end users locally using Small Cells.

(v) Virtual Small Cell Network Operator (VSCNO): This is a firm that leases the equipment to provide radio communication services to end users using virtual technology.

(vi) Mobile Network Operators (MNO): Is the entity that provides the mobile network services to the end users.

(vii) SESAME Platform: This is a platform that provides the CESC, The CESCM and the Light DC services.

ORGANISATIONAL VIEW

The organisational view of SESAME architecture with the inter-relationship between major actors is illustrated in Figure 3. As it has been illustrated in the diagram, the end User (EU) depends on the Mobile Network Operator to provide the services such as voice call, video streaming, etc. To provide these services, the Mobile Network Operator depends on the SESAME platform to provide the Virtual Network Functions, Orchestration, CESC, CESC Manager (CESCM) etc. Also SESAME platform depends on Virtual Network Operators to provide virtual network infrastructures. For the Virtual Network Operators to provide these services, it depends on the infrastructure owners to provide them with the physical infrastructures. The Infrastructure owners depend on the venue owners to provide the site where the physical infrastructures are to be deployed.

From the security point of view, the infrastructure owner depends on the VO to provide the physical security of the equipment. However, this responsibility can be negotiated between them. Also the SESAME platform depends on the VSCNO to provide the security of the entire virtual environment. While, the MNO depends on the SESAME platform to provide security of the CESCM and the Light DC. Also the MNO depends on the EU to ensuring confidentiality of the login credentials.

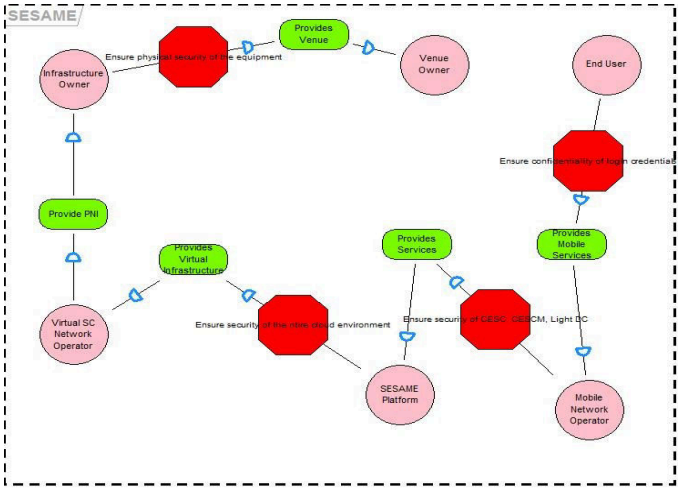


Figure 3: SESAME Organisational View

SECURITY REQUIREMENT VIEW

As described in the previous section, the CESCM is responsible for the management of the entire SESAME system. It consists of the NFV management, Service Function Chaining (SFC) management, Virtual Infrastructure (VI) management and the Orchestration. However there are two main areas of concern in CESCM in relation to the security. The first is the northbound interface, which provides access to the system from the external environment, and the second is the Orchestrator, which provide the coordination and management of the system.

Therefore, in this study emphasis was given to the CESCM and its two main actors. Figure 4 presents the Security Requirement View of the main actors of the CESCM (NFVO and Northbound Interface).

The Northbound Interface

The main goal of northbound interface Identified in SecTro is to provide accessibility to the NFVO. The security constraints identified are:

1) The service request data should be able to pass through the interface to the orchestrator.

2) The interface should only be accessible to the authorised users.

3) It should support the encryption mechanism used for the user credential.

4) The interface should be available to the authorised users at all times.

The Security objectives that satisfy the above mentioned constraints are; availability, accessibility, communicative and Encryption. Numbers of potential threats associated with the interface were identified, such as DoS, Information disclosure, Data tempering and elevation of privilege. Security mechanism is identified for each security objective. For accessibility and availability, user should be authenticated at the minimum with strong and unique username and password. For communicative, data flow control should be set to allow data forwarding to the correct interface. For the encryption, cryptographic algorithm to be used has to be correctly defined in the interface.

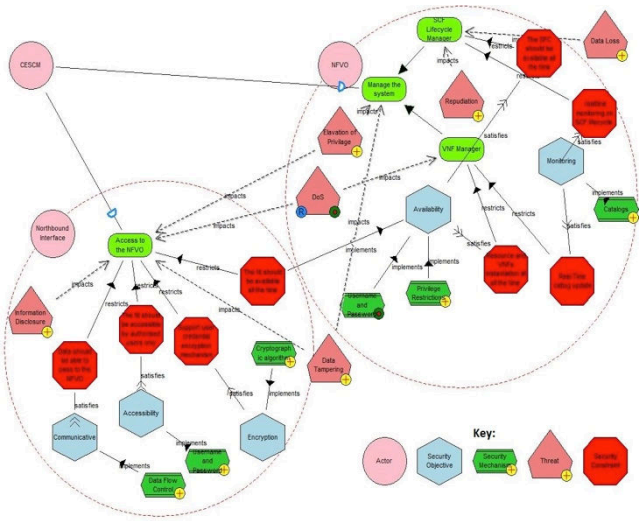


Figure 4: SESAME Security Requirements View

The NFVO

The responsibilities of the NFVO are to managed and coordinate the computational, storage and network resources as well as monitoring of services. Two main goals were identified; 1) the management of the VNF and 2) SCF lifecycle management. Security constraints that are related to VNF manager include constant instantiation of resources and VNF and real-time VNF status update. The security objective identified is that the VNF manager should be available to the VNFO at all times.

V. MULTI-OPERATOR/TENANCY CLOUD-ENABLED SMALL CELL SECURITY

To define holistic security architecture for SESAME, Northbound, Southbound and Westbound interfaces as well as orchestrator must be considered. However, northbound interface is the most important interface that need to be thoroughly secured. It is the interface that provides communication with the external element of the system. While southbound and westbound interfaces provide communications within the internal modules. The orchestrator provides automotive management solution for the system through creating, monitoring, and deployment of resources in virtual environment. Figure 5 illustrates the communication channels. Providing security to the northbound interface and the orchestrator will serve as providing security to the entire system. The secured northbound interface means secured communication with any external system. Also securing orchestrator means securing the management of the system.

Therefore, in this work northbound interface and orchestrator are considered.

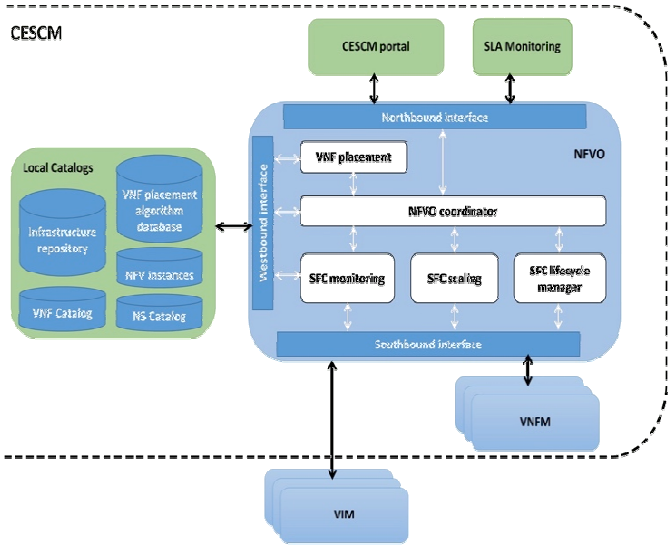


Figure 5: Cloud-Enable Small Cell Management

The northbound interface is an abstraction layer to the internal modules of the Network Function Virtualisation Orchestrator (NFVO) for the requests coming from the CESCM. It is popularly explained in the SDN terminology to provide an abstraction layer that serves an entry point to the internal NFVO components and it is placed in the upper part of the NFVO architecture diagram. The interface connects SDN applications to the controller. An application can request information, such as statistics and incoming connections from the controller. It can also sends commands to the controller, in order to control the network, such as added or removed flow rules. In order to prevent security issues in this interface, every call to the NFVO's interface need to be secured and done by authorised users and modules. Threads can be identified using STRIDE threat modelling[6]. Table 1.1 illustrates STRIDE categories and security property for each category.

STRIDE threat categories	**Threat category Security property**
Spoofing	Authentication
Tampering	Integrity
Repudiation	Non-repudiation
Information Disclosure	Confidentiality
Denial of Service	Availability
Elevation of Privilege	Authorization

Table 1: STRIDE threat categories

There are number of threats vectors related to the SDN northbound interface [7], which are categorised into two; 1. Attacks on and vulnerabilities in controllers (including applications) and 2. Lack of mechanisms to ensure trust between the controller and management applications. The first category is further subdivided into multiple vulnerabilities related to the northbound interface. These vulnerabilities are as follows:

1. Service Chain Interference: When a message is forwarded from application to application, a malicious application may not send the message on to the next application, causing that message to be lost. This results in Control Message Drop or Infinite Loops.
2. Control Message Abuse: A malicious application can arbitrarily insert new flows in the switches' flow table. This results in Flow Rule Modification or Flow Table Clearance
3. Northbound API Abuse: A malicious application may request the controller to disconnect other applications. This results in Event Listener Un-subscription or Application Eviction.
4. Resource Exhaustion: A malicious application can continuously sending requests to the controller, consuming all its resources. This results in Memory Exhaustion or CPU Exhaustion.

VI. STRIDE THREAT MODEL

The first step in STRIDE modelling is to create a data flow diagram and the second step is to identify threats category using STRIDE method explained above.

Spoofing: Illegally accessing and using another user's authentication information, such as username and password. Only authenticated actors should be able to use the northbound interface. Authentication credentials should not be guessed, listened or obtained.

Tampering: Data tampering involves the malicious modification of data. Data sent over the northbound interface should not be tampered with.

Repudiation: Repudiation threats are associated with users who deny performing an action without other parties having any way to prove otherwise. When there is no secure log of commands sent over the northbound interface, it is relatively easy for a malicious actor to perform actions anonymously. A ***Nonrepudiation mechanism should be put in place in*** the northbound interface.

Information disclosure: The northbound interface handles information about the network state and its configuration. While not extremely sensitive, this data can be valuable to some parties. Therefore, this information should not be disclosed to unintended parties. Encrypting network traffic and requiring authentication will aid in preventing information from being disclosed.

Denial of Service (DoS): The northbound interface is very important for the SDN infrastructure. When it is unavailable, applications cannot do their work. DoS poses a serious threat to this interface. A malicious user could either send a large amount of traffic to the interface, or he could send resource-intensive requests to the controller, both resulting in the interface becoming unavailable. Solutions for this include making the interface accessible only from a trusted network, or using other traditional DoS mitigation techniques.

Elevation of Privilege: Applications should only have the least amount of privileges needed for their operation. For

example, a monitoring application should not have the right to write to the controller, only reading will suffice. In addition, there will be very few applications which need the right to alter the device configuration. The controller should enforce some kind of access control on its API functions to prevent an application from accessing too much of the API.

VII. CONCLUSION

In the analyses carried out in SecTro, it show that northbound interface provides access to NFVO from external elements with security constraints of service request, encryption machanism and avaibility. The NFVO security constraints include constant instantiation of resources and VNF and real-time VNF status update. While the southbound and westbound interfaces provides communications within the internal modules. Also the analysis shows that providing security to the northbound interface and the orchestrator using STRIDE threat model secured the entire CESC system. Hence, recomends the adoption of muti-tennency virtual environment using SESAME architecture.

REFERENCES

[1] 4G Americas, J. Ren, L. Liu, H. Zhou, Q. Zhang, S. Yu, X. Gui, F. Tian, P. Yang, A. Shameli-sendi, Y. Jarraya, and M. Fekih-ahmed, "Developing and Integrating a High Performance HET-NET," 2012.

[2] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," IEEE Commun. Mag., vol. 51, no. 3, pp. 136–144, 2013.

[3] S. Yu, X. Gui, F. Tian, P. Yang, and J. Zhao, "A security-awareness virtual machine placement scheme in the cloud," in Proceedings - 2013 IEEE International Conference on High Performance Computing and Communications, HPCC 2013 and 2013 IEEE International Conference on Embedded and Ubiquitous Computing, EUC 2013, 2014, pp. 1078–1083.

[4] A. Shameli-Sendi, Y. Jarraya, M. Fekih-Ahmed, M. Pourzandi, C. Talhi, and M. Cheriet, "Optimal placement of sequentially ordered virtual security appliances in the cloud," in Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015, 2015, pp. 818–821.

[5] J. Ren, L. Liu, H. Zhou, and Q. Zhang, "ESI-Cloud : Extending Virtual Machine Introspection for Integrating Multiple Security Services." 2016

[6] S. Hernan, S. Lambert, and T. Ostwald, "Uncover Security Design Flaws using The STRIDE Approach" msdn. microsoft. com," Design, no. 1, pp. 1–8, 2006.

[7] D. Kreutz, F. M. V. Ramos, and P. Verissimo, "Towards secure and dependable software-defined networks," Proc. Second ACM SIGCOMM Work. Hot Top. Softw. Defin. Netw. - HotSDN '13, p. 55, 2013.

[8] ETSI and NFV, "Network Functions Virtualisation (NFV); Managment and Orchestration," ETSI GS NFV-MAN 001, 2014.

[9] B. Jaeger, "Security Orchestrator," in 2015 IEEE Trustcom/BigDataSE/ISPA, 2015, pp. 1255–1260.

[10] K. Meyler, P. Zerger, M. Oh, A. Bengtsson, and K. Van Hoecke, System Center 2012 Orchestrator Unleashed. Pearson Education, Inc, 2014.

Improving Energy Efficiency of Low Density Parity Code system

[1]Lukman A. Olawoyin, [2]Olaniran K. Olawale, [1]Nasir Faruk, [1]Abdulkarim A Oloyede, [1]Olumuyiwa B. Ayeni, [3]S.K. Adewole and [4]Ismaeel A Sikiru

[1]Department of Telecommunication Science, University of Ilorin, Nigeria
[2]Federal Polytechnic, Ede, Osun State, Nigeria
[3]Depatment of Computer Science, University of Ilorin, Nigeria
[4]Department of Information and Communication Science, University of Ilorin. Nigeria

Email: olawoyin.la@unilorin.edu.ng; niran4uk2005@gmail.com; adewole.ks@unilorin.edu.ng; oloyede.aa@unilorin.edu.ng; faruk.n@unilorin.edu.ng; o_bayles@yahoo.com

Abstract: ***The amount of energy consumed by low density parity check (LDPC) hybrid automatic repeat request (HARQ) is high, hence, this work investigates the efficient use of energy in LDPC HARQ system. A new technique is proposed in HARQ scheme that can be adopted so as to improve the energy efficiency by determining the maximum iteration number of decoding required. The existence of energy efficiency and convergence of word error rate (WER) and throughput are proved to indicate the better maximum iteration number. After retransmission and combining, HARQ also improves the energy efficiency. The simulation results obtained showed that energy efficiency in the new scheme out-performed conventional HARQ scheme.***

Keywords: energy efficiency; HARQ; LDPC; max iteration number, throughput

I. INTRODUCTION

In recent times, the research interest in energy and environment had been in increased. Energy Efficiency (EE) has become a focal point where energy conservation and energy harvesting are two aspects which are of major interest nowadays. In order to minimize energy cost, Zafer and Modiano [1] in their work considered an optimal rate-control transmission approach while the authors in [2] analyzed best modulation strategy, it was shown in their work that some considerable gains are achieved if the uplink receiver uses multiple antennas and, also the authors in [3] proposed power control scheme using coding techniques, the result obtained shown that the use of coded MQAM is not only power efficient but also bandwidth efficient than uncoded MFSK. Additionally, the author in [4] examined the maximum transmission policy for limited used of energy for battery energy harvesting node.

Since data service has been developing quickly, Hybrid Automatic Repeat Request (HARQ) has able to find its application in wireless communication systems such as HSPA, LTE, 802.16e [5-7]. In recent years, EE in HARQ has also become an area of interest for many researchers. The author, Choi et al, discussed Energy-Delay Tradeoff (EDT) of HARQ when channel state information (CSI) can indicate the type of fading signals the channel experienced [8]. In [9], the authors analyzed EDT of HARQ in multi-user systems in downlink channels, the author proved that by allocating equal power, it can provide a near optimal EDT in the high power regime while the relay collaborative systems was discussed in [10] while Tumula and Erik optimized power allocation of CC-HARQ in [11]. However, all the researches work done so far paid more attention on energy allocation or the relationship with delay whereas energy consuming itself should desired to be given more consideration/attention.

In recent times, applications of TURBO codes [12] and Low-Density Parity-Check (LDPC) codes [13, 14] are now chosen for Forward Error Coding (FEC) in HARQ. Both the TURBO and LDPC codes are using iterative decoding techniques. Hence, this work investigates the relationship between iterative decoding and EE of LDPC-HARQ system. In simple FEC without combining with ARQ, although reducing iteration number of decoding can save energy however, there is performance degradation.

The rest of the paper is organized as follows: Section II briefly introduces the LDPC iterative decoding module, and analyzes the determination of maximum iteration number without HARQ. The optimization of maximum iteration number in LDPC-HARQ is proposed in Section III. In section IV, numerical simulation results are presented while section IV concludes the paper.

II. DETERMINATION OF MAXIMUM ITERATION NUMBER OF LDPC DECODING

In order to determine the message passing algorithm for LDPC code, the Belief Propagation (BP) based technique is used which passes message iteratively from variable nodes to check nodes in the Tanner graph. The soft information of a binary bit b is defined as:

$$\eta_x = \ln \frac{\Pr(x=0)}{\Pr(x=1)} \tag{1}$$

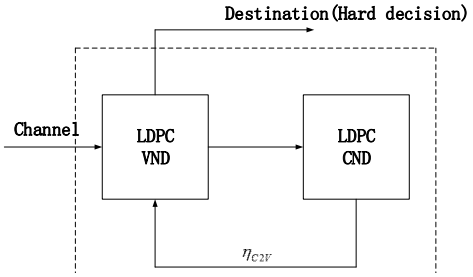


Fig.1 Iterative decoding of LDPC

The process of iteration decoding mainly constrains Check Node Decoder (CND) and Variable Node Decoder (VND). Define $C = \{1,2,...,m\}$ as the set of all check nodes, $V = \{1,2,...,n\}$ as the set of all variable nodes. The soft information propagating in the LDPC decoder can be stated in Fig.1 and further categorized into the following types:

- the full soft information of $b_i : \eta(i), i \in V$
- the soft information input to the decoder: $\eta_{ch}(i), i \in V$;
- the soft information from $j-th$ check node to $i-th$ variable node: $\eta_{C2V}(j,i), i \in V, j \in C$;
- the soft information from $i-th$ variable node to $j-th$ check node: $\eta_{V2C}(j,i), i \in V, j \in C$;

Hence, the following relationships can be established:

$$\eta(i) = \eta_{ch}(i) + \sum_{j:h_{j,i}=1} \eta_{C2V}(j,i) \tag{2}$$

$$\eta_{V2C}(j,i) = \eta(i) - \eta_{C2V}(j,i) \tag{3}$$

$$\eta_{C2V}(j,i) = f(\eta_{V2C}(j,i); i : H_{j,i} = 1) \tag{4}$$

where the multi-dimensional function $f()$ is the check node update function. The standard algorithm for the algorithm is the sum-product algorithm (SPA) [15] and there are also many simplified versions for it, such as min-sum algorithm (MSA)[16], Normalized-MSA[17], Offset- MSA[18],etc. Different iterative decoding algorithm can be developed with different algorithm function for message passing.

The soft information is propagated along the edges and is processed at nodes in an iterative way. At the end of each iteration, the decoder may take the hard decision as represented by

$$\hat{b}_i = 1/2\,(1 - \mathrm{sgn}[\eta(i)]) \tag{5}$$

while hard decision is used to calculate the syndrome $\hat{b} = (\hat{b}_1, \hat{b}_2, ..., \hat{b}_n)$:

$$p = \left(\hat{b}H\right)^T \tag{6}$$

If $p = 0$, the output of decoder $\hat{b}$ terminates. Otherwise, the iteration will continue until the maximum iteration is achieved. Also if the decoder does not terminate within the predefined maximum iteration number $N_{\max}$, a decoding failure is declared.

In the analysis of the characteristic of LDPC iterative decoding without HARQ, the LDPC (576,768) code is applied as FEC, coded bits are modulated by MPSK when $M = 2$ and transmitted over an Additive White Gaussian Noise Channel (AWGN). The results obtained show several curves of WER and throughput come from different max iteration number $N_{\max}$. The Fig.2, shows that WER decreases as SNR increases. Meantime, the more maximum iteration number, the

less WER obtained. However, when the number $N_{\max}$ is larger than number set, the difference of WER is small. In Fig.3, throughput of LDPC shows similar phenomenon that the value could increase little when $N_{\max}$ is set more than 80. From mentioned above, maximum iteration number $N_{\max}$ should be set to 80. It was obtained that there no difference in the result when applied to many other coding techniques. Nevertheless, this design is suitable for the LDPC system without HARQ.

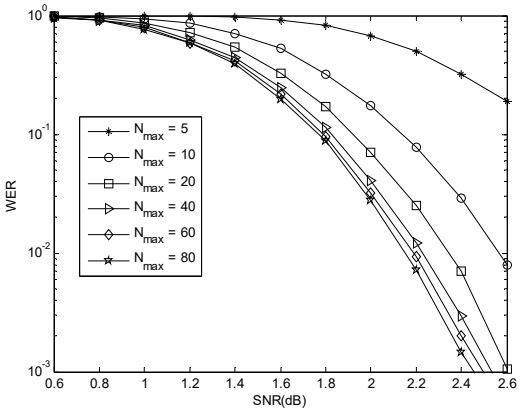


Fig. 2 WER of LDPC without HARQ

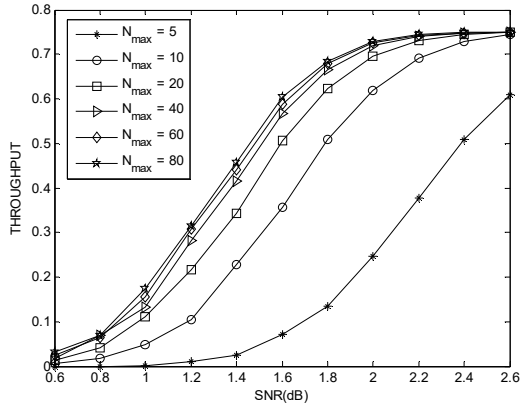


Fig. 3 Throughput of LDPC without HARQ

III. OPTIMIZATION OF MAXIMUM ITERATION NUMBER IN LDPC-HARQ DECODING

In this system, Chase Combining (CC) HARQ is used to demonstrate the analysis of maximum iteration number $N_{\max}$ in HARQ decoding. By using Incremental Redundancy (IR) HARQ, it is very easy to directly extend the principle. When the decoding of receiver is failed, the transmitter received NACK. Then, retransmission could be triggered and repeated codeword would be transmitted. The maximum retransmission number $M_{\max}$ is the significant parameter here, it determines the delay system can tolerant and the throughput system can achieve. Furthermore, this number influents energy efficiency deeply. If decoding is successful, the feedback changes to be ACK which indicates the transmitter is free to send a new data.

The design of $N_{\max}$ in simple FEC may be unsuitable for HARQ systems. On the one hand, each codeword has only one time of decoding in simple FEC. The iteration number should be large enough for more successful decoding and better throughput. On the other hand, if decoding failed, retransmission and combining will guarantee the performance in HARQ systems. Thus the energy consuming of failed iteration decoding is wasteful while retransmission may lead to successful decoding. Reducing the waste is helpful to improve EE of HARQ when performance loss should be considered. From Figures 4 and 5, it can be observed the WER and throughput of different retransmission number M, where Max Iteration Number $N_{\max}$ changes from 0 to 70 and SNR=-2dB. $M=0$ stands for the first transmission while $M=1,2,3$ means the successful rounds of retransmission. WER can converge more quickly when increasing retransmissions happens. In other words, the more the retransmission the less the iterations need to decode successfully. From Fig. 4, it can be seen that WER of the second and third retransmission converge at $N_{\max}=20$ and $N_{\max}=5$ respectively and the same result can be found in Fig.5, where the converged points move left after retransmissions. Therefore, Max Iteration Number $N_{\max}$ can be optimized for improving energy efficiency in HARQ system.

IV. SIMULATION RESULTS

The parameters of LDPC channel coding in 802.16e standard is applied here. Information bits are coded by LDPC (576,768) modulated by BPSK. After transmitting over AWGN channel, signals are received and decoded by MSA. Then feedback of ACK/NACK is sent to trigger CC-HARQ or new data transmission. Delay constrain could be met by the setting $M_{\max}=4$.

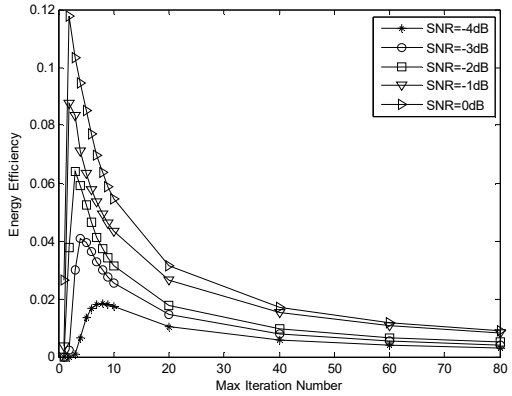


Fig.6 EE of HARQ schemes

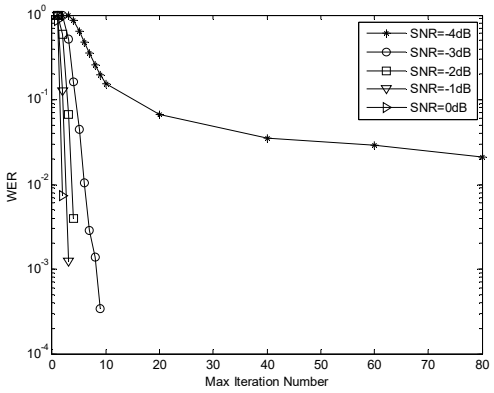


Fig.7 WER of HARQ schemes

The performance of EE after 4 transmissions can be observed in Fig.6. It is clear that the pole $N_{\max}$ exists even SNR is different. And it move left when SNR increases. All the Maximum Iteration Numbers mapping to the poles are much smaller than the traditional value (80). Selecting the $N_{\max}$ mapping to the pole is best for EE. Then we confirm the viewpoint by WER and choose one value suitable for all SNRs. In Fig.7, WER curves of different SNR are depicted. After 4 retransmissions, the difference of WER is extremely small when $N_{\max}$ is large. This means $N_{\max}$ can be selected smaller to improve EE. In the communications system of mentioned above, $N_{\max}$ is modified to 10.

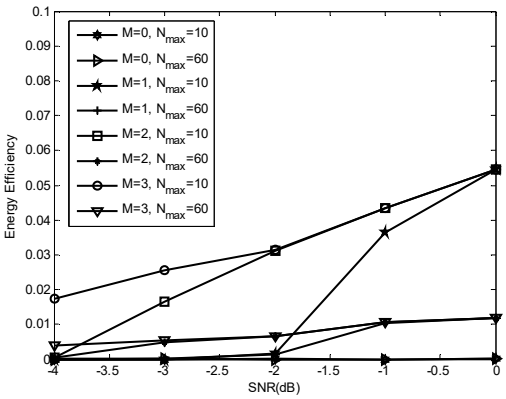


Fig.8 EE of two HARQ schemes

Fig.8 compares energy efficiency of traditional and proposed example when no more than 4 transmission are transmitted. The difference between two schemes is that $N_{\max}$ is changed from 60 to 10. Most values of EE in new scheme are much better than in old one except the first transmission in which both of them are zero. For instance, at -4dB, EE in proposed scheme is about 4 times as much as in traditional one in the third retransmission (M=3). At 0dB, this gain increase to about 5 times. Obviously, the improvement of energy efficiency in the proposed scheme is great after $M_{\max}$ retransmission. Additionally, the gain is raised when retransmission round is increasing. For example, the proposed scheme, where $N_{\max}=10$, has the largest EE when M=3. In other words, the more retransmission works, the larger EE can be obtained. It is easy to find that HARQ can also bring gain of EE remarkably.

V. CONCLUSION

In this work, it have been investigated the iterative decoding of LDPC-HARQ to improve energy efficiency. The design of maximum iteration number is to obtain best performance of WER and throughput in traditional scheme. Since it has been established that the existence of the pole of energy efficiency and the convergence of WER and throughput in HARQ, optimization of maximum iteration number can avoid energy wasting in iterative decoding. The result obtained from simulation shows that proposed scheme performed better than traditional scheme when the loss of throughput is small. Additionally, HARQ can bring significant gain of energy efficiency by retransmission and combining.

References

[1] Meshkati, F and Poor H V and Schwartz S C and Mandayam N B. [Buzzi, 2008 #49]. IEEE Transactions on Communications, 2005, 53(11): 1885-1894.

[2] Zafer M and Modiano E. Minimum energy transmission over a wireless channel with deadline and power constraints. 2009, 54(12): 2841-2852.

[3] 3GPP Technical Specification 25.212, Multiplexing and Channel Coding (FDD), V6.5.0, 2005.

[4] 3GPP Technical Specification 36.212, Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding, V11.2.0, 2013.

[5] IEEE Std 802:16TM-2004. IEEE standard for local and metropolitan area networks Part 16: air interface for fixed broadband wireless access systems. 2005.

[6] Choi Jinho. Energy-Delay Tradeoff Comparison of Transmission Schemes with Limited CSI Feedback. IEEE Transactions on Communications, 2013, 12(4):1762-1773.

[7] Choi Jinho and Ha Jeongseok and Jeon Hyoungsuk. On the Energy Delay Tradeoff of HARQ-IR in Wireless Multiuser Systems. IEEE Transactions on Communications, 2013, 61(8): 3518 - 3529.

[8] Jinho Choi and Weixi Xing and Duc To and Ye Wu and Shugong Xu. On the Energy Efficiency of a Relaying Protocol with HARQ-IR and Distributed Cooperative Beamforming. IEEE Transactions on Communications, 2013, 12(2): 769 - 781.

[9] Tumula V. K. Chaitanya and Erik G. Larsson. Optimal Power Allocation for Hybrid ARQ with Chase Combining in i.i.d. Rayleigh Fading Channels. IEEE Transactions on Communications, 2013, 61(5): 1835-1846.

[10] Near optimum error correcting coding and decoding: turbo-codes. IEEE Transactions on Communications, 1996, 44(10): 1261-1271.

[11] Gallager R G. Low-Density Parity-Check Codes. IEEE Transactions on Information Theory, 1962,8(1): 21-28.

[12] Mackay D J C and Neal R M. Near Shannon limit

performance of low-density parity-check codes. Electronics Letters, 1996, 32(18):1645-1646.
[13] Chen J H, Fossorier M P C. Near optimum universal belief propagation based decoding of low-density parity-check codes. IEEE Transactions on Communications, 2002, 50: 406–414
[14] Chen J H, Fossorier M P C. Density evolution for two improved BP-based decoding algorithms of LDPC codes. IEEE Communications Letters, 2002, 6: 208–210

5)

Tools for Analytical and Simulation Investigation of Named Data Networking

Muktar Hussaini,
InterNetWorks Research Laboratory, School of Computing, UUM Malaysia,
Sintok Kedah, Malaysia.
elhussenkaz@yahoo.com

Habeeb Bello-Salau
Department of Computer Engineering, Ahmadu Bello University Zaria,
Nigeria.
bellosalau@abu.edu.ng

Hiba Jasim Hadi
IASDO Research Laboratory, School of Computing, UUM Malaysia, Sintok
Kedah, Malaysia.
hebaj81@gmail.com

Shahrudin Awang Nor
InterNetWorks Research Laboratory, School of Computing, UUM Malaysia,
Sintok Kedah, Malaysia.
shah@uum.edu.my

Amran Ahmad
InterNetWorks Research Laboratory, School of Computing, UUM Malaysia,
Sintok Kedah, Malaysia.
amran@uum.edu.my

Ibrahim Muhammad Hassan
Department of Computer Science
Hussaini Adamu Federal Polytechnic
Kazaure, Jigawa Nigeria.
ibrahimhass003@gmail.com

***Abstract*—Named Data Networking is a consumer-driven network architecture that supports content consumer mobility due to the nature of in-network catching. The architecture is proposed to supplement and replace current Internet architecture by using named content in place of IP-address. In this paper mobility problem is selected to presents and discussed about the tools necessary for problem investigation in NDN. Usually modeling, simulation and testbed methods are used for investigation of any problem in the field of networking. Analytical investigation and simulation tools are covered, and the investigation can be done using network analysis model with any technique such as hop count method for analytical investigation. While simulators such as ndnSIM, CCNLite, OMNET ++ and ccnSim can be used as tools for simulation.**

***Keywords*— Network analysis model, named data networking, simulation tool, analytical investigation, simulation investigation.**

I. INTRODUCTION

Named Data Networking (NDN) architecture is completely clean-slate architecture that evolved from the IP Internet architecture retain and expand the end-to-end principle with routing and forwarding plane separation [1], [2]. The architecture of NDN maintained the IP hourglass architecture with the replacement of a thin waist with hierarchical named content in place of IP addresses. NDN focus on the name of the data and do away with the location of data [3]. The names were structured as hierarchical form and self-certifying with a digital signature. Moreover, the packet's route by name discourses some problems pose in IP: address management, address space exhaustion, the mobility that requires changing address and NAT traversal [1], [4]. Unlike in IP architecture that secured the connection from the source to destination, NDN secured the content for integrity, authenticity, and confidentiality of data. The architecture comprises three nodes, consumer, producer, and routers. Also, two packets Interest and Data (Fig. 2) were managed by three data structures Content Store (CS), Forwarding Information Base (FIB) and Pending Interest Table (PIT) to enable the named based routing [5], [6].

NDN is an entirely new architecture whose design principles evolved from the successes of the current Internet, that is IP architecture [1], [2]. There are six architectural principles that guided the design of NDN such as (i) the hourglass architecture centres on IP universal network layer, (ii) provision of security built in the architecture, (iii) retain and expand end-to-end principle, (iv) self-regulating flow-balance data delivery, (v) separation of routing and forwarding plane, and (vi) user choice and competition [2].

The thin waist of the current Internet hourglass architecture located at the centres of the universal IP network layer, shown in Figure1. Which provide global connectivity, allow upper-layer and lower-layer technologies to transform and designed independently for network communications. Due to the rapid growth of data demand over the Internet, led the dominant use as a distribution network that makes it complicated for point-to-point communication protocol to solve [2]. However, NDN maintains the hourglass architecture, but replaces the thin waist with name objects or content chunks instead of IP addresses or communication endpoints in every node, as shown in Figure 1. The name can represent a data chunk, endpoint, command, etc. The concept changed by NDN provides named data security, flow balance, in-network storage, multipath forwarding, and solve end-to-end communication, content distribution, and control problems [1].

This paper contributes the following:

- Named Data Networking overview is provided based on the architectural principle and its benefits.
- The NDN mobility problems with regards to content producer mobility are presented.
- The tools for simulation analysis and their capabilities are discussed.
- The analytical investigation for future Internet, especially NDN are presented for researchers' accessibility.
- NDN simulator that is designed purposely for sake on named content simulation is used and evaluate some scenarios.

This paper is organized in to five section, besides the introductory section, the tools for solving mobility problems based on the existing literature are reviewed and discussed. Section three presented the tools for NDN simulation,

Section four is the result discussion and analysis. and finally, a conclusion section.

II. Routing and Forwarding in NDN

NDN is a consumer-driven network that communications initiated from the interest of the consumer. A consumer sends and Interest that carries content prefix or name for the deification of desired data. A router that receives the Interest look up the CS for content availability, if found, the router replied back with data packets to the consumer, otherwise, the router recorded the interface of incoming Interest in PIT, then look up FIB for the information about the content populated by NBR protocol and forward the Interest further. When the Interest reaches any router or content producer that has requested data, a data is sent back with name, data content and producer's signature via the path recorded in the PIT to the consumer. The data packet looks up the PIT, save in CS and forward downstream.

Among the Information-Centric Networking (ICN) approaches, Named Data Networking is the most recent approach; infrastructures less network architecture that provides network scalability due to its nature of hierarchical naming [7]. NDN is hope and promising architecture for the future Internet that has been used and applied in the different fields, such as Space-Terrestrial Integrated Network [8] WSN, Vehicular Network [9], [10] and real-time application [11]. By default, ICN approaches designed to support mobility. However, many challenges arose in some approaches like NDN in support of content mobility [12]–[15]. Hierarchical naming structure of NDN offers some benefits, such as route aggregation, improvement of scalability, multicasting. However, the naming structure causes a significant challenge when a content producer moves to a new location. Zhu et al. [4] ascertain the content consumer mobility support in NDN, and content producer faces similar handoff challenges like an IP architecture.

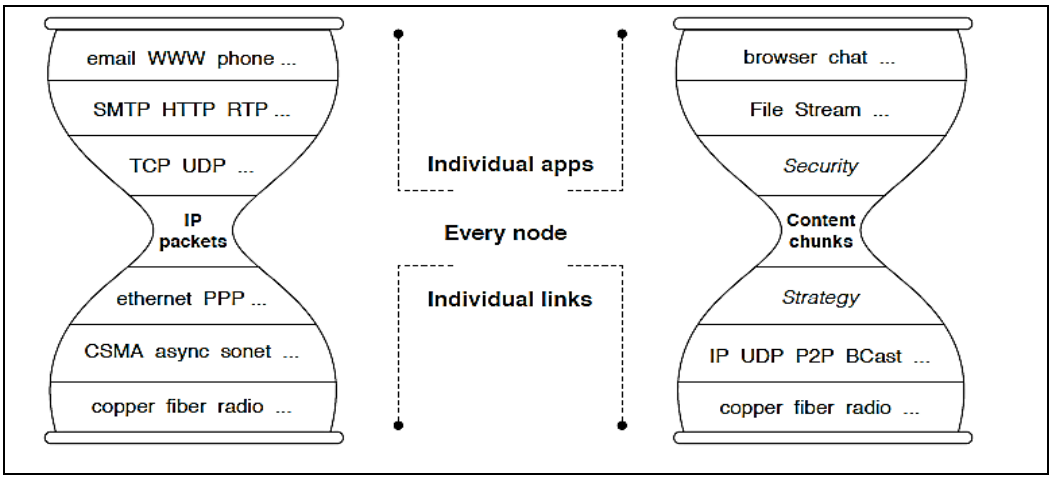


Fig. 1. NDN Data and Interest Format

II. Tools For Solving Mobility Problems

Mobility problem is selected in this paper to analyze the tools necessary for problem investigation in NDN. Typically modeling, simulation and testbed methods are used for investigation of any problem in the field of networking. This exactly happens in the field of future Internet networking such as NDN, CCN, DONA, NetInf, CBN etc. In this paper, Analytical investigation and simulation tools are covered.

A. Analytical Modeling Tools

Analytical models were used to describe the behavioural performance of the predicted system using logical and mathematical principles, that are typically complicated and requires assumptions [16]. Analytical modeling required less time, cost-effective, and is suitable for protocol analysis that required less or no scalability constraint and can provide better predictive behavior of the system compared to simulation and testbed [17].

Mobility handoff problems such as handoff latency, signaling overhead cost and path optimization costs were addressed using analytical investigation. Handoff latency and cost have been mathematically formulated from the existing proposed schemes and approaches to evaluate the concept and address handoff performance. Mathematical models are easier to represents logical behavior between network elements [18], also covers the bound of research question by showing the borderline of the network behavior or characteristics [19]. For the analytical investigation, many researchers formulated an analytical model for the evaluation of their proposed ideas. Deyun Gao [20] proposed network mobility and analysis model for evaluating the producer mobility support solution. Cha et al., [21] proposed a network analysis model for mobility link service scheme to support consumer mobility in NDN. Qin, Zhou, and Xu [22] used network analysis model and formulated a solution to the mobility management scheme for a mobile producer. Gao and Zhang [23] used a network analysis model for handoff and communication latency analysis for total overhead cost.

Also, Do and Kim [24] developed a network analysis model to formulate and evaluate their scheme. Also compared the performance of their scheme and other schemes for benchmarking analysis. Recently, Rui, Yang, and Huang [11] formulated a mobility scheme using network analysis model for analytical investigation of a

producer mobility scheme for real-time multimedia delivery in NDN. However, in their formulation of analytical model movement behavior of mobile producer is neglected, even though a mobile producer must have a mobility component. Consequently, the content producer or consumer mobility behavior that is affecting the performances of handoff latency and signaling cost [25] is ignored. Thus, mobile content producers' or content consumers' movement pattern plays a significant role in the study of NDN mobility support performance analysis.

Furthermore, Rao, Luo, et al., [26] proposed a locator-based mobility scheme to support mobile producer. The scheme improved the forwarding efficiency and routing scalability on novel locator-based mobility support scheme [27]. The author tried to incorporated movement behavior in the formulation of handoff signaling cost and latency by using portable movement model. However, the portable movement model was proposed to update portable location before connection after movement and paging traffic in personal communication services [28], which are more of time and cost analysis for mobile call service. The model does not have the mobility component of the mobile node such as velocity, pause and static. Hence it is not suitable to determine the exact movement pattern of the mobile producer.

Because of the advantage of the random movement pattern of the mobile node, a random model is used frequently in mobile networking simulation. The model mimics the mobility behavior of mobile nodes in a simplified manner [25]. Also, it is widely accepted because of its simplicity and adequately captured mobility characteristics without geographic restriction, spatial and temporal dependency [29]. Therefore, this research proposed an efficient mobility mechanism function that incorporates movement behavior of mobile producer for handoff signaling cost and latency formulation using random movement model.

B. Network Analysis Models

For the analytical investigation, many types of research [20]–[24], [26], [27] been conducted, developed and built an analytical model for the validation of their concept and evaluation of handoff problems.

Cha et al., [21] proposed a network analysis model to evaluate the performance of NDN face mobility link service scheme to support consumer mobility in NDN. In addition, a scalable mobility management (SMM) scheme [23] for a content producer in NDN, that adopts three different separation mechanisms, access/core separation, locator/ID separation, and management/routing separation. Gao and Zhang [23] used a network analysis model for handoff and communication latency analysis of the total overhead cost. Moreover, Deyun Gao [20] used a network mobility and analysis model to evaluate the performance of proposed and existing producer mobility support solution. Also, Ren, Qin, Zhou and Xu [22] in their study mobility management scheme for a mobile producer in CCN that was proposed based on SDC to reduce signaling exchange, overhead cost and routing update processes. A network analysis model was proposed to evaluate the performance of SDC-CCN scheme.

Furthermore, Rao, Luo et al., [26] and Rao Gao et al., [27] proposed a locator-based mobility scheme to support content producer on mobile and improved forwarding efficiency and routing scalability. For the performance analysis of the proposed scheme, producer movement behavior was analyzed, and the network analysis model was built to analyze the handoff latency and overhead handoff cost [26], [27]. Therefore, in this research network analysis model is used to formulate the mathematical models for the verification and validation of proposed producer mobility support scheme and evaluation of handoff performance.

C. Analytical and Simulation Investigation Method

Some researchers used analytical investigation [22]–[24], [26], [27], [30]–[33] or simulation investigation [34], [35], [44], [36]–[43] to validate and evaluate their proposed ideas or concept. Some combined both analytical and simulation investigation for the evaluation of their proposed ideas [20], [21], [45], [46]. The combination of two methods depends on the research questions that research is designed to address. Predominantly, handoff problems can be investigated using analytical method whereas other issues like the determination of packets delivery, packets losses, scalability, and throughput needs to investigated using the simulation for a better result. This is because of simulation provided control and scalable environment for an extensive investigation of the system behavior [16], [19].

To solve the producer mobility problem exist in NDN architecture, Gao et al. [20] proposed proxy-based mobility support to minimizes handoff latency, handoff cost and Interest/data packets losses. An analytical investigation was used for the formulation of handoff latency analysis and handoff cost analysis [20]. Simulation investigation was used to analyzed inter-domain handoff cost, latency and to evaluate the packets losses.

Analytical and simulation investigation was performed for NDN face mobility links service proposed by Cha et al. [21] to manage connections and transaction of data and Interest packets during handoff processes. The scheme was proposed to support consumer mobility services evolved independently. An analytical investigation was used to determine handoff overhead cost while simulation investigation was carried out using ndnSIM to determine the handoff performance that covers handoff delay for the scheme [21].

Lanlan, Yang and Huang [45] proposed a scheme to support consumer mobility in NDN by the formulation of the proactive multi-level cache selection, that focus on handoff situation when consumer disconnects and moves to another PoA before receiving the content requested. The mathematical formulation of the scheme are converted into algorithms, then simulated for the evaluation of total handoff cost and total handoff period or latency [45].

TABLE VII
SUMMARY OF ANALYTICAL AND SIMULATION INVESTIGATION CONDUCTED

Evaluation Technique	Number of Research	Percentage (%)	References
Simulation	17	49	[9], [34], [43], [44], [47]–[51], [35]–[42]

Analytical	11	31	[8], [11], [33], [22]–[24], [26], [27], [30]–[32]
Both	7	20	[20], [21], [45], [46], [52]–[54]

Based on the literature reviewed in this research, simulation investigation covered 49% of the researches carried out and 31% for analytical investigation. Also, 20% covered both analytical and simulation investigation, as highlighted in Table I. Moreover, it is shown in Figure 2. This indicated that research problems always determine the type of design and evaluation method or technique would be used. Therefore, in this research, a combination of analytical and simulation investigation is used to clearly investigate and effectively addressed the research problems presented in Chapter One. The analytical investigation covered objective one and two, while simulation investigation covered objective three.

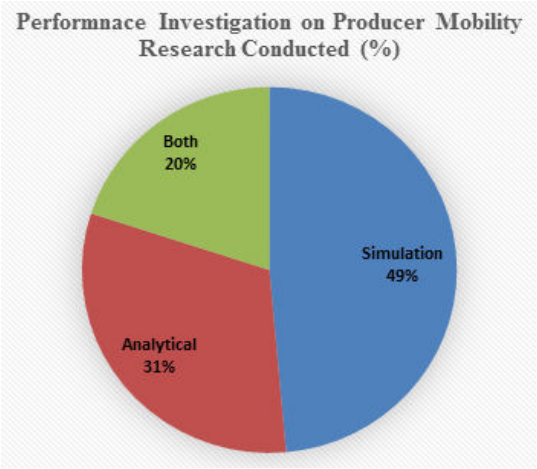


Fig. 2. Percentage Based on Analytical and Simulation Investigation

III. Tools for NDN Simulation Analysis

For the analysis of research conducted in the area of Named Data networking, three methods can be used such as analytical, simulation and testbed method.

A. Simulation Method

Simulation is one of the recognized methods for performance analysis and evaluation of communication networks, algorithms and protocols, which is easily applied for both the existing and non-existing system, or at the design phase of the system. Simulation is cost-effective and requires reasonable time, also provide a controlled and scalable environment that the entire and details behavior of the system can be investigated widely [16], [19]. Moreover, network simulation provides a set of models that support and describes the process and operation of the communication channel, network component, and environmental factors, like network and node mobility support [16]. Moreover, simulation techniques offer researchers a controlled environment that the system is investigated widely [19]. Also, Gyires [18] mentioned that, in highly complex analysis, simulation scenario is more appropriate as a technique to measure the performance.

Research on networking practice of computing devices is growing and becoming more complex that force the needs for a sophisticated, affordable, accurate, and scalable network simulator despite the emergence of testbeds. Simulators are commonly used for network research to design a model, prototyping, development, and testing. Many networks simulators were developed are NS-3, OPNET, OMNET++, CCNx, ndnSIM, CCN-Lite, ccnSim, mini-CCNx etc. However, some of these simulators were designed to support extensive network simulation without full or partial support for the named data networking.

B. Simulation Tools

There are many simulation tools for the analysis and investigation of any future Internet. The standard tools, especially for NDN, are presented below:

NS-3: Network Simulator version 3 (NS-3) is the discrete-event network simulator design to maintained the codebase and extend the functionality of Network Simulator version 2 (NS-2), for simulation of Internet systems that put more emphasis on network layer 2, 3 and 4, targeted for research and educational purposes [19], [55], [56]. NS-3 simulator is an open source software that includes other open networking software and supports the integration of other software for clarity, extensible, and scalable simulation-based experiment [55]. Moreover, NS-3 can be easily installed on Linux, Window OS, Solaris, OS X for both desktop and server platforms. The simulator has models for different network elements such as network, packets, nodes and devices, communication channels and protocols. It provides output trace files including statistics, log, and animation for easy performance evaluation.

OMNeT++: Objective Modular Network Test-bed in C++ (OMNeT++) is an extensible discrete event, component-based and modular C++ simulation library [19], [57]. OMNET++ is public-source simulation tool that is widely used for academic institutions, for non-profit research purposes of simulating communication networks both wired and wireless, also for distributed or parallel systems. It supports the extension for database integration, real-time simulation, system integration, network emulation, and several functions, designed to be general instead of particular simulator [57]. For the design and simulation of NDN architecture, OMNeT++ supports the extension of CCN-Lite and ccnSim for NDN/CCN models design, implementation, and evaluation.

ccnSim: Content Network Simulator (ccnSim) [58] was developed using OMNeT++ framework for the scalability issue of NDN networks caching performance analysis. However, it was designed as chunk-level simulator only suitable for caching analysis and optimized for the research and experimentation for NDN cache replacement policies [6], [59]. Hence, ccnSim cannot be used for the analysis of the main component of NDN architecture that deals with the application and forwarding strategy layer.

CCN-Lite: CCN-Lite [60] is a functionally interoperable lightweight of CCNx and NDNx implementation protocols meant for classroom and experimental extension using OMNeT++ simulation

platform. Explicitly designed for packets scheduling and fragmentation [6], [59]. However, CCN-Lite does not support the full implementation of CCNx and NDN protocol architecture and limited with speed optimization because of reliance on link list data structure. For a complex simulation analysis that represents entire NDN architecture with full exploitation of forwarding daemon, CCN-Lite cannot provide the necessary support.

ndnSIM: Mastorakis et al. [59] proposed ndnSIM (NDN Simulator) to support large scale experimentation of various design in the NDN architecture, to provide a common, open-source and user-friendly simulation platform for NDN research community [6], [59], [61]. The simulator is the extension of NS-3 based platform which completely implements the essential design components and features of NDN network in the modular technique using C++ programming. The protocol featured by ndn-cxx Library with new NDN Forwarding Daemon that can support different real application experiment. However, due to the NFD integration, the researchers have the flexibility to implement different simulation scenarios, with different strategies. Hence, ndnSIM offers to the researchers and developer the integrated environment to simulate and evaluate the real-world applications. Also provides the necessary underlying L3 protocols to be installed in any NDN node and easy modification, replacement and extension of any component.

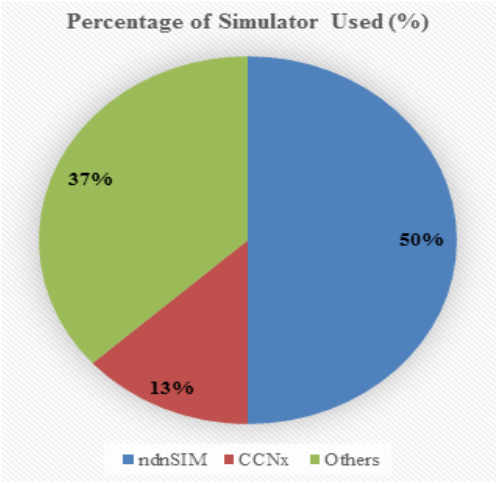


Fig. 3. Justification for Selecting ndnSIM as Simulation Tool

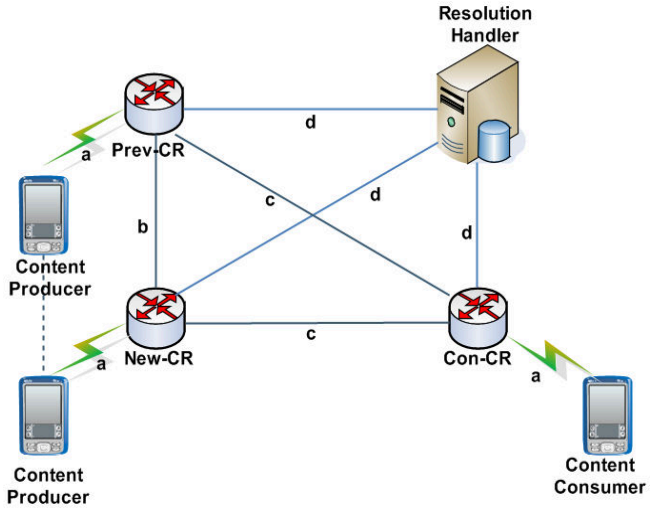


Fig. 4. Network Analysis Model of Producer Mobility

Traditionally, in the field of NDN research, many researchers are using ndnSIM for the simulation of their proposed design. Figure 3 shows the summary based on the percentage of some researches that aimed to provide producer mobility support, and they used ndnSIM more than CCNx simulator and others. Some researchers used other simulators such as packet-level simulator, VoCCN simulator, discrete-event simulator and NS-3 or built their simulators. The extension of NS-3 network simulator called ndnSIM was developed with essential components of NDN network, and flexible tool for researchers is fully implemented [61]. Hence, ndnSIM is suitable and used for this research. The network analysis model was built and presented in Figure 4 as in previous researches as in [23], [24], [26] for the validation and evaluation of proposed producer mobility support schemes ideas, model, concepts, etc.

TABLE II NETWORK ANALYSIS MODEL PARAMETERS

Category	Notation	Parameters	Value
Packets	S_{name}	Size of the signaling packet	+16 byte
	S_{data}	Size of data packets	2000 bytes
	S_{iint}	Size of Interest packet	40 bytes
Latency	L_{par}	Transmission latency between producer to AR	*a*
	L_{car}	Transmission latency between consumer to AR	*a*
	L_{sar}	Transmission latency between Server to new AR	*d*
	$L_{o\text{-}nar}$	Transmission latency between old AR to new AR	*b*
	L_{ars}	Transmission latency between new ARs/Anchors	*c*
	L_{pn}	Time interval btw producer disconnection and reconnection from old AR to new AR	I_{pn}
Signaling Cost	C_{par}	Transmission cost hop/packet producer to AR	*a*
	C_{car}	Transmission cost hop/packet consumer to AR	*a*
	$C_{o\text{-}nar}$	Transmission cost hop/packet old AR to new AR	*b*
	C_{ars}	Transmission cost hop/packet ARs/Anchors	*c*
	$C_{s\text{-}nar}$	Transmission cost hop/packet AR to Server	*d*

B. The formulation for Handoff Latency

In this section, the formulation of handoff latency is presented as an example of how to use analytical modelling for solving the mobility problem. The formulation is deducted from Figure 5 and the parameters presented in Table II are used for the hop count formulation to represents the transmission cost and latency between hops.

Hop count is the number of intermediate nodes through which data and Interest must pass between content producer and consumer. Also, hop count measure the distance between two nodes, that is consumers, producers, routers or servers with their queuing delay, link delay and bandwidth for the wired or wireless connection. Therefore, the total delay that a content producer spends for the disconnection, reconnection and the arrival of the first Interest packet through new PoA is called the handoff latency. The handoff latency of a mobile producer can be acquired when there is a disconnection from the current PoA and reconnection to new PoA as a result of the producer's movement. Equation (1) measured the delay for a wired link between two consecutive hops and Equation (2) for wireless link delay [24].

$$Lw_{name} = \left(\frac{S_{name}}{(Bw+Ldw+Qd)}\right) \quad (1)$$

$$Lwl_{name} = \left(\frac{1+q}{1-q}\right) \times \left(\frac{S_{name}}{Bwl+Ldwl}\right) \quad (2)$$

Where B_w and B_{wl} are bandwidth for the wired and wireless link, Ld_w and Ld_{wl} are linked delay wired and wireless, q is the probability of link failure and the queuing delay Q_d.

The hop count handoff latency formulation of BMA model can be express as Equation (3). When content producer disconnected form old-AR and reconnected to a new-AR and sends a mobility Interest to the new-AR, its completes the handoff over the transmission latency of the time interval *Ipn*. The new-AR add the mobility status and tag the MI packets, indicating that the content name is on mobile. Then broadcast the MI to update the intermediate routers within the domain, with transmission latency *Lpar,* and *Lars*. The content consumer sends normal pending Interest packets directly to the new location of content producer via an optimal path, with transmission latency *Lcar,* and *Lars*. The formulation is presented in Equation (3):

$$\begin{aligned} L_{BMA} &= L_{pn} + L_{par} + 2L_{ars} + L_{car} + L_{sar} \\ &= l_{pn} + a + 2c + a + c \\ &- l_{pn} + a \times Lwl_{nint} + 2c \times Lwl_{nint} + \\ &a\, Lwl_{Int}\, c \times Lw_{Int} \end{aligned} \quad (3)$$

III. RESULT ANALYSIS

A. Analytical Result Analysis

Figure 5 shows the variation of transmission latency between the old Access Router (AR) and new AR, that is a parameter (*b*). The result proved that BMA reduced the handoff latency compared to DMA and LMA In Figure 6, the latency of DMA and LMA schemes were increased significantly as a parameter (*d*) increases. Due to the number of messages sent for updating the server, while that of BMA remains constant, as there is no transmission delay of handoff messages via *d* routes in the absence of server. The server in DMA schemes are placed to provides a producer, and consumer means to update and query the server before the handoff became successful. Further, Figure 7 presents the handoff latency result by varying transmission latency between Ars with parameter (*c*). The result shows that all schemes are affected, but BMA has the lowest and better latency. Therefore, the BMA has the lowest handoff latency compared to other solutions by varying *b*, *c*, and *d*.

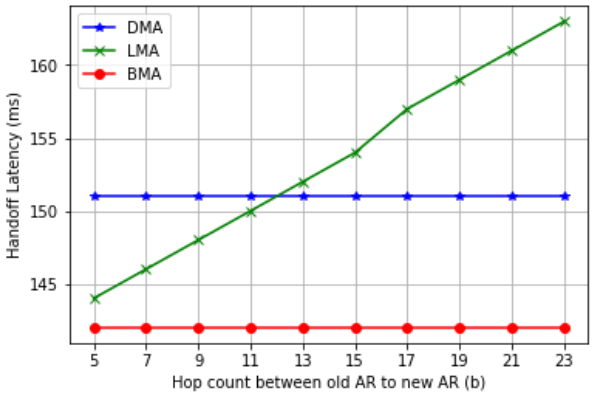


Fig. 5. Handoff Latency Analysis for DMA, LMA and BMA for varying transmission cost between P-AR and N-AR.

Two more results are presented for varying other parameters to show how the analytical result will be presented. The figures are shown in Figure 7 and 8.

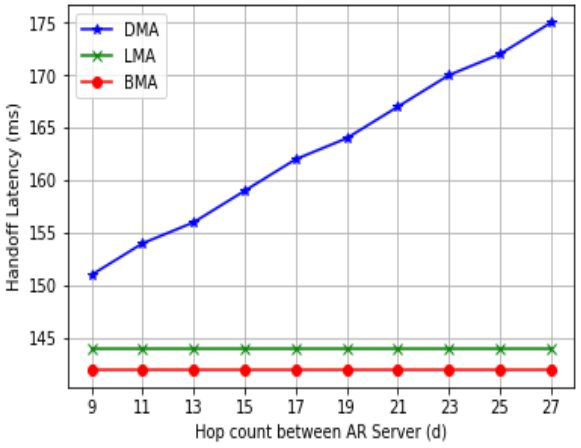


Fig. 6. Handoff Latency Analysis for DMA, LMA and BMA for varying transmission cost between Producer, Consumer and Server.

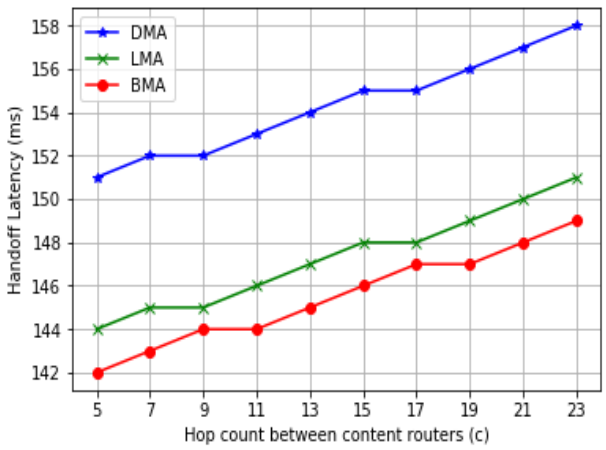


Fig. 7. Handoff Latency Analysis for DMA, LMA and BMA for varying transmission cost between Routers.

B. Simulation Result Analysis

The experimental design started with the definition of network topology that represents the experimental networking environment, which contained consumers, producers, and intermediate routers. Alternatively, a network topology reader can be used to read the topology content from an external file. Then configuration of content store, FIB, selection, and configuration of forwarding strategy and installation of NDN stack on consumers, producers and routers, the processes are presented in details for both simulation and analytical experimentation in these literatures [62]–[65].

The simulation results can be obtained by connection to one or more generated trace file that can keeps trace readings for each node as a result provided by ndnSIM classes. The result of the tracing especially for the rate is stored in bytes and number of packets forwarded by an NDN node as Interest or data packets. The output file is formatted with tab-separated values in rows. The first row is specifying the names of the columns as shown in Table III. It is also possible to use existing trace helpers, which collects, and aggregates requested statistical information in text files.

TABLE III SIMULATION DATA RATE TRACE FILE PARAMETERS

Column	Description
Time	Time for the simulation
Node	ID for each Node which is unique
FaceId	Interface ID (-1 for combined metric)
Type	Type of measurements: • In-Interests measurements • Out-Interests measurements • In-Data measurements • Out-Data measurements • In-Nacks measurements • OutNacks measurements • Satisfied-Interests measurements for all faces • Timed-Out-Interests measurements for all faces • In-Satisfied-Interests measurements per incoming face • In-Timed-Out-Interests measurements per incoming face • Out-Satisfied-Interests measurements per outgoing face • Out-Timed-Out-Interests measurements per outgoing face
Packets	Estimated rate of packets within the last averaging period
Kilobytes	Estimated rate within last averaging period
PacketsRaw	Absolute number of packets within the last averaging period
KilobytesRaw	The absolute number of kilobytes transferred

There are many simulation scenarios that can be used to conduct a simulation. The scenario can be represented based on the kind of topology such as Tree, Grid, Rocketsfuel, Star, Dumbbell etc. These scenarios will be use and trace where the producers, routers and consumers are installed. The ndnSIM have a CS-Tracer that is used to obtain statistics of cache hits and cache misses on every node simulated. The successful codes that are run will create a file named rate-trace.txt in the current directory, or any directory assigned by the user which can be analyzed manually. In addition, it can be used as input to some packages that can statically plots graphs to represents the output of the simulation. Therefore, the result can be presented as in Figure 8.

TABLE IV
SIMULATION CACHE TRACE FILE PARAMETERS

Column	Description
Time	Simulation Time
Node	Node ID, globally unique
Type	Type of counter for the time period. Possible values are: • CacheHits: Specifies the number of satisfied • CacheMisses: Specifies the number of Interests not satisfied
Packets	The number of packets for the period

In Table, IV shows the output file format, which is tab-separated values, with the first row specifying names of the columns. Refer to the following table (Table IV) for the description of the columns. The successful run will create a drop-trace.txt file in the current directory, which can be analyzed manually or used as input to some graph/stats packages.

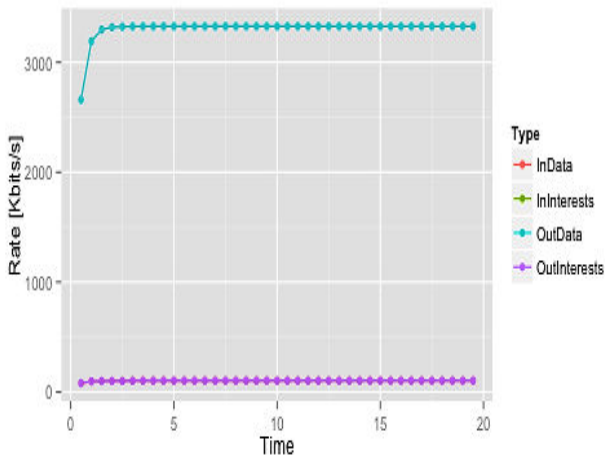


Fig. 8. IN and OUT Interest Rate Analysis from ndnSIM simulator.

IV. CONCLUSION

This paper presented some trend of Information-Centric Networking approach and reviewed mobility support for NDN to presents a tool needed for the evaluation of NDN experiment selected as a research domain. NDN is among the approaches that need further research to serve prominent future Internet architecture. Also, NDN is the most recent approach with infrastructures less network architecture, that provides network scalability due to its nature of hierarchical naming. Further, some conceptual theories and schemes for the existing solutions of mobility support for NDN.

Mobility problem is selected in this paper to analyze the tools necessary for problem investigation in NDN. Normally modeling, simulation and testbed methods are used for investigation of any problem in the field of networking. This exactly happens in the field of future Internet networking such as NDN, CCN, DONA, NetInf, CBN etc. In this paper, Analytical investigation and simulation tools are covered. Predominantly, the handoff performance of mobility support in NDN is investigated using analytical, simulation or both. The handoff signaling cost, latency and data packet delivery cost can be investigated using the analytical method, whereas other issues like the determination of packets delivery ratio, packets losses, scalability and throughput needs to be conducted through simulation investigation for a better result. Henceforth, in this research, the network analysis model is reviewed for the formulation of a mathematical model of proposed producer mobility support concept and the design of the proposed mobility mechanism. Also, simulation investigation presented especially using ndnSIM for the performance evaluation.

REFERENCES

[1] L. Zhang *et al.*, "Named Data Networking," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 66–73, 2014.

[2] L. Zhang *et al.*, "Named Data Networking," *NDN, Tech. Rep. NDN-0001*, no. October, 2010.

[3] A. Dowling, M. Babaeianjelodar, Y. Liu, and K. Chen, "Performance evaluation of NDN applications in low-interference Mobile Ad Hoc Environments," in *2018 IEEE International Conference on Communications*, 2018, pp. 1–6.

[4] Z. Zhu, A. Afanasyev, and L. Zhang, "A new perspective on mobility support," *Named-Data Networking Project, Technical Report NDN-0013*, pp. 1–6, 2013.

[5] S. Mastorakis, A. Afanasyev, and L. Zhang, "On the Evolution of ndnSIM," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 47, no. 3, pp. 19–33, 2017.

[6] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2 : An updated NDN simulator for NS-3," *NDN, Tech. Rep. NDN-0028, 2015*, pp. 1–8, 2016.

[7] M. Hussaini, S. A. Nor, A. Ahmad, I. S. Maijama'a, A. Isah, and A. Aminu, "Mobility support and operation of Information Centric Networking approach," *Int. J. Comput. Netw. Inf. Secur.*, vol. 10, no. 10, pp. 18–25, 2018.

[8] D. Liu, C. Huang, X. Chen, and X. Jia, "Space-terrestrial integrated mobility management via Named Data Networking," *Tsinghua Sci. Technol.*, vol. 23, no. 4, pp. 431–439, 2018.

[9] M. Gohar, N. Khan, A. Ahmad, M. Najam-ul-islam, S. Sarwar, and S. Koh, "Cluster-based device mobility management in Named Data Networking for Vehicular Networks," *Mob. Inf. Syst.*, vol. 2018, pp. 1–7, 2018.

[10] J. M. Duarte, T. Braun, and L. A. Villas, "MobiVNDN: A distributed framework to support mobility in vehicular named-data networking," *Ad Hoc Networks*, vol. 82, pp. 77–90, 2019.

[11] L. Rui, S. Yang, and H. Huang, "A producer mobility support scheme for real-time multimedia delivery in Named Data Networking," *Multimed. Tools Appl.*, vol. 77, no. 4, pp. 4811–4826, 2018.

[12] M. Hussaini, S. A. Nor, and A. Ahmad, "Producer mobility support for Information Centric Networking approaches : A review," *Int. J. Appl. Eng. Res.*, vol. 13, no. 6, pp. 3272–3280, 2018.

[13] D. Kutscher *et al.*, "Information-Centric Networking (ICN) Research Challenges," *Internet Research Task Force (IRTF) Request for Comments: 7927*, pp. 1–38, 2016.

[14] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile Information-Centric Networking: Research issues and challenges," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 2353–2371, 2018.

[15] M. Hussaini, S. A. Nor, and A. Ahmad, "Producer mobility support scheme for Named Data Networking: A survey," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, pp. 31–42, 2018.

[16] E. Weingärtner and K. Wehrle, "Synchronized performance evaluation methodologies for communication systems," Hochschulbibliothek der Rheinisch-Westfälischen Technischen Hochschule Aachen, 2013.

[17] M. M. Kadhum, "Network Performance and NS2 Tutorial," in *NETAPPS2010*, 2010.

[18] T. Gyires, *Algorithms of informatics*, vol. 2. Budapest, Hungary: AnTonCom, 2010.

[19] K. Wehrle, M. Güneş, and J. Gross, Eds., *Modeling and Tools for Network Simulation*. Berlin, Heidelberg: Springer, 2010.

[20] D. Gao, Y. Rao, C. H. Foh, H. Zhang, and A. V. Vasilakos, "PMNDN: Proxy Based Mobility Support Approach in Mobile NDN Environment," *IEEE Trans. Netw. Serv. Manag.*, vol. 14, no. 1, pp. 191–203, 2017.

[21] J. H. Cha, J. H. Choi, J. Y. Kim, S. G. Min, and Y. H. Han, "A mobility link service in NDN face to support consumer mobility service," in *International Conference on Ubiquitous and Future Networks, ICUFN*, 2017, pp. 444–449.

[22] F. Ren, Y. Qin, H. Zhou, and Y. Xu, "Mobility management scheme based on Software Defined Controller for Content-Centric Networking," in *Proceedings - IEEE INFOCOM*, 2016, vol. 2016-Septe, no. 2, pp. 193–198.

[23] S. Gao and H. Zhang, "Scalable mobility management for content sources in Named Data Networking," *2016 13th IEEE Annu. Consum. Commun. Netw. Conf. CCNC 2016*, pp. 79–84, 2016.

[24] T. X. Do and Y. Kim, "Optimal provider mobility in large-scale Named- Data Networking," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 10, pp. 4054–4071, 2015.

[25] C. Bettstetter, "Smooth is better than sharp: A Random Mobility Model for Simulation of Wireless Network," *Proc. 4th ACM Int. Work. Model. Anal. Simul. Wirel. Mob. Syst. - MSWIM '01*, pp. 19–27, 2001.

[26] Y. Rao, H. Luo, D. Gao, H. Zhou, and H. Zhang, "LBMA: A novel Locator Based Mobility support Approach in Named Data Networking," *China Commun.*, vol. 11, no. 4, pp. 111–120, 2014.

[27] Y. Rao, D. Gao, and H. Luo, "NLBA: A novel provider mobility support approach in mobile NDN environment," in *2014 IEEE 11th Consumer Communications and Networking Conference, CCNC 2014*, 2014, no. CCNC, pp. 188–193.

[28] Y. Fang, I. Chlamtac, and Y. B. Lin, "Portable movement modeling for PCS networks," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1356–1363, 2000.

[29] F. Bai and A. Helmy, "A Survey of Mobility Models in Wireless Adhoc Networks," in *Wireless Ad Hoc and Sensor Networks*, University of Southern California,U.S.A, 2004, pp. 1–30.

[30] Z. Yan, S. Zeadally, S. Zhang, R. Guo, and Y.-J. Park, "Distributed mobility management in Named Data Networking," *Wirel. Commun. Mob. Comput.*, vol. 16, no. 13, pp. 1773–1783, Sep. 2016.

[31] J. Tang, H. Zhou, Y. Liu, H. Zhang, and D. Gao, "A source mobility management scheme in Content-Centric Networking," *2014 IEEE 11th Consum. Commun. Netw. Conf. CCNC 2014*, no. Ccnc, pp. 176–181, 2014.

[32] F. Ren, H. Zhou, L. Yi, Y. Qin, and H. Zhang, "A hierarchical mobility management scheme for Content-Centric Networking," in *2014 IEEE 11th Consumer Communications and Networking Conference, CCNC 2014*, 2014, no. CCNC, pp. 170–175.

[33] Z. Yan, G. Geng, S. Zeadally, and Y. J. Park, "Distributed All-IP Mobility Management Architecture Supported by the NDN Overlay," *IEEE Access*, vol. 5, no. c, pp. 243–251, 2017.

[34] Z. Zhou, X. Tan, H. Li, Z. Zhao, and D. Ma, "MobiNDN: A mobility support architecture for NDN," in *Proceedings of the 33rd Chinese Control Conference, CCC 2014*, 2014, pp. 5515–5520.

[35] D. Kim, J. Kim, Y. Kim, H. Yoon, and I. Yeom, "End-to-end

mobility support in Content Centric Networks," *Int. J. Commun. Syst.*, vol. 28, no. 6, pp. 1151–1167, Apr. 2015.
[36] B. Hu, Z. Li, and S. Zou, "Provider Mobility Management Based on Domain Proxies in Content Centric Networks," in *2014 International Symposium on Communications and Information Technologies (ISCIT)*, 2014, pp. 393–397.
[37] Y. Zhang, H. Zhang, and L. Zhang, "Kite," *NDN, Technical Report NDN-0020, 2014.*, no. 1, NDN Project, pp. 1–10, 2014.
[38] Y. Zhang, H. Zhang, and L. Zhang, "Kite: A Mobility Support Scheme for NDN," in *Proceedings of the 1st ACM Conference on Information-Centric Networking- (ICN) ACM-ICN '14*, 2014, no. 1, pp. 179–180.
[39] A. Azgin, R. Ravindran, and G. Wang, "A Scalable Mobility-Centric Architecture for Named Data Networking," *Arxiv Prepr. arXiv1406.7049.*, 2014.
[40] A. Azgin, R. Ravindran, and G. Wang, "Mobility Study for Named Data Networking in Wireless Access Networks," in *Communications (ICC), 2014 IEEE International Conference on*, 2014, pp. 3252–3257.
[41] J. Lee and D. Kim, "Partial path extension scheme for mobile content source in content-centric networking (CCN)," *EURASIP J. Wirel. Commun. Netw.*, vol. 2015, no. 1, p. 212, 2015.
[42] J. Lee, S. Cho, and D. Kim, "Device mobility management in Content-Centric Networking," *IEEE Commun. Mag.*, vol. 50, no. 12, pp. 28–34, 2012.
[43] L. Wang, O. Waltari, and J. Kangasharju, "MobiCCN: Mobility support with greedy routing in Content-Centric Networks," *GLOBECOM - IEEE Glob. Telecommun. Conf.*, pp. 2069–2075, 2013.
[44] D. Han, M. Lee, K. Cho, T. Kwon, and Y. Choi, "Publisher mobility support in content centric networks," *Int. Conf. Inf. Netw. 2014*, no. February 2014, pp. 214–219, 2014.
[45] L. Rui, S. Yang, and H. Huang, "A proactive multi-level cache selection scheme to enhance consumer mobility support in Named Data Networking," *Int. J. Distrib. Sens. Networks*, vol. 13, no. 4, p. 155014771770089, 2017.
[46] Y. S. Chen, C. S. Hsu, and D. Y. Huang, "A pipe-assisted mobility management in Named Data Networking networks," *APNOMS 2014 - 16th Asia-Pacific Netw. Oper. Manag. Symp.*, 2014.
[47] S. Korla and S. Chilukuri, "T-Move: A Light-Weight Protocol for Improved QoS in Content-Centric Networks with Producer Mobility," *Futur. Internet*, vol. 11, no. 2, p. 28, 2019.
[48] J. Augé, G. Carofiglio, G. Grassi, L. Muscariello, G. Pau, and X. Zeng, "MAP-Me: Managing anchor-less producer mobility in Information-Centric Networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 15, no. 2, pp. 596–610, 2018.
[49] H. Farahat and H. S. Hassanein, "Proactive caching for Producer mobility management in Named Data Networks," in *2017 13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017*, 2017, pp. 171–176.
[50] R. A. Rehman and B.-S. Kim, "LOMCF: Forwarding and Caching in Named Data Networking-based MANETs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 1–1, 2017.
[51] J. V. Torres, I. D. Alvarenga, R. Boutaba, and O. C. M. B. Duarte, "An autonomous and efficient controller-based routing scheme for networking Named-Data mobility," *Comput. Commun.*, vol. 103, pp. 94–103, 2017.
[52] G. Zheng, A. Tsiopoulos, and V. Friderikos, "Optimal VNF chains management for proactive caching," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 10, pp. 6735–6748, 2018.
[53] J. Cha, J. Choi, J. Kim, Y. Han, and S. Min, "A Mobility Link Service for NDN Consumer Mobility," *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–8, 2018.
[54] H. Farahat and H. Hassanein, "Optimal caching for producer mobility support in Named Data Networks," in *2016 IEEE International Conference on Communications, ICC 2016*, 2016.
[55] T. R. Henderson, S. Roy, S. Floyd, and G. F. Riley, "ns-3 Project Goals," in *Proceeding from the 2006 workshop on ns-2: the IP network simulator*, 2006, p. 13.
[56] T. R. Henderson, M. Lacage, and G. F. Riley, "Network Simulations with the ns-3 Simulator," *SIGCOM Demonstr.*, vol. 14, no. 14, p. 527, 2006.
[57] A. Varga and R. Hornig, "An Overview of the OMNeT++ Simulation Environment," in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops*, 2008, p. 60.
[58] R. Chiocchetti, D. Rossi, and G. Rossini, "ccn S im : an Highly Scalable CCN Simulator," in *Communications (ICC), 2013 IEEE International Conference on. IEEE*, 2013, pp. 2309–2314.
[59] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2 . 0 : A new version of the NDN Simulator for NS-3," *NDN, Tech. Rep. NDN-0028.*, pp. 1–8, 2015.
[60] T. Christian, C. Scherb, D. Mansour, L. Beck, M. Sifalakis, and U. Schnurrenberger, "CCN-lite Project," 2013. [Online]. Available: http://ccn-lite.net/#doc.
[61] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN Simulator for NS-3," *NDN, Tech. Rep. NDN-0005*, pp. 1–7, 2012.
[62] M. Hussaini, M. A. Naeem, B.-S. Kim, and I. S. Maijama'a, "Efficient producer mobility management model in Information-Centric Networking," *IEEE Access*, vol. 7, pp. 42032–42051, 2019.
[63] M. Hussaini, S. A. Nor, A. Ahmad, R. Mustapha, and A. F. Abdulateef, "A Conceptual Model of Producer Mobility Support for Named Data Networking using Design Research Methodology," *IAENG Int. J. Comput. Sci.*, vol. 46, no. 3, pp. 1–11, 2019.
[64] M. Hussaini, S. A. Nor, and A. Ahmad, "Optimal broadcast strategy-based producer mobility support scheme for Named Data Networking," *Int. J. Interact. Mob. Technol.*, vol. 13, no. 04, pp. 4–19, Apr. 2019.
[65] M. Hussaini, S. A. Nor, and A. Ahmad, "Analytical modelling solution of producer mobility support scheme for named data networking," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 5, pp. 3850–3861, 2019.

Mobility Support Challenges for the Integration of 5G and IoT in Named Data Networking

Muktar Hussaini,
InterNetWorks Research Laboratory, School of Computing, UUM Malaysia,
Sintok Kedah, Malaysia.
elhussenkaz@yahoo.com

Shahrudin Awang Nor
InterNetWorks Research Laboratory, School of Computing, UUM Malaysia,
Sintok Kedah, Malaysia.
shah@uum.edu.my

Habeeb Bello-Salau
Department of Computer Engineering, Ahmadu Bello University Zaria,
Nigeria.
bellosalau@abu.edu.ng

Hiba Jasim Hadi
IASDO Research Laboratory, School of Computing, UUM Malaysia, Sintok
Kedah, Malaysia.
Hebaj81@gmail.com

Aminu Abbas Gumel
Department of Computer Science Hussaini Adamu Federal Polytechnic
Kazaure, Jigawa Nigeria.
aminuabbas360@yahoo.com

Kabiru Abdullahi Jahun
Department of Computer Science Hussaini Adamu Federal Polytechnic
Kazaure, Jigawa Nigeria.
kabjah@yahoo.com

***Abstract*—Named Data Networking (NDN) is a consumer-driven network that supports content consumer mobility due to the nature of in-network catching. Consumer mobility in NDN architecture is ultimately supported unlike content producer mobility that is not supported, also faces many challenges from initial architectural design. Therefore, there is needs to provide substantial producer mobility support to minimize the handoff latency, handoff signaling overhead cost, reduce the unnecessary Interest packets loss to improve data packets delivery once a content producer relocated. In this paper, the Broadcasting Approach is proposed using broadcast strategy as a solution to the problem of the mobile producer in NDN. Consequently, the result may solve the inherited problems of triangular routing in NDN network mobility, especially for the tracing-based mobility called Kite. Also, may have significant implication to support the integration of 5G, Mobile Ad-hoc Networks (MANET), Delay-Tolerant Network, Vehicular Ad-hoc Networks (VANET) and IoT network into NDN for mobile application without a special mechanism.**

***Keywords*—Tracing-based mobility, handoff latency, Named data networking, Producer mobility support, broadcast strategy, mobility Interest**

I. INTRODUCTION

Named Data Networking proposed as clean-slate Internet architecture under Information Centric Networking (ICN), as an evolution of IP architecture, to discourse the weaknesses of the current Internet, by transforming address-centric nature of point-to-point communication to content-centric nature. Some researchers' confirmed that NDN rooted from earlier Content Centric Networking (CCN) project for the enhancement and provision of standard Future Internet [1], [2]. Further, many challenges face NDN such as forwarding strategy, scalable forwarding, mobility support, security and privacy, application development, etc., as attributed to the umbrella of ICN challenges [3], [4].

However, with the rapid development in networking and the vast emergence of mobile devices, many researchers' individuals, academic and organization studied how to manage the mobility of such devices that were connected to the network. Thus, to make the Internet functions better, **Internet Engineering Task Force (IETF), organized by Internet Society,** provided many standard protocols that were developed to support nodes on mobile such as Mobile IPv4 (MIP) [5], Mobile IPv6, Proxy Mobile IP (PMIP) and many extensions [6]. Also, the protocols were based on IP architecture operated on addresses, yielding some problems such as, triangular routing that causes low efficiency and delay, high handoff cost and latency, high level of packets droplet and signaling overhead.

Moreover, NDN that is anticipated as the future Internet to address these problems has emerged. By default, NDN is designed to support mobility, but many challenges aroused with regards to mobility support. Mobility support allows mobile devices to relocate between different Point of Attachments (PoAs) without disrupting the content availability and minimal hand-off delay. PoA is a device that allows the mobile node to connect to the network. Therefore, mobility in ICN is categorized into network, consumer and producer mobility [7]. The support that make content producer to move around and relocate without disrupting the network for content availability within minimal handoff latency [7]. However, producer mobility was unanimously agreed in many studies that are not supported in NDN [8]–[12].

This paper contributes the following:

- The mobility problems are presented from IP based Internet to future Internet, based on ICN.
- Research problem with regards to producer mobility support in NDN are discussed.
- Tracing based mobility support solution is presented, and improved solution are proposed in this paper.
- Proposed broadcasting-based mobility support for NDN producer is investigated and results are provided, analyzed and presented.

The rest of the paper is organized as Section two presented the research problem and motivation for NDN, Section three discussed about the existing tracing-based mobility approach, which is proposed as Kite. Moreover, Section four presented the idea of proposed solution that provides the path optimization after handoff and finally a conclusion section after validation and result discussions.

II. RESEARCH PROBLEM AND MOTIVATION

In NDN technical report for the new perspective on mobility support for NDN, Zhu et al. [8] ascertained the unsupport in NDN architecture of content producer mobility. In addition, expressed that the architecture faces similar problems as in

IP architecture for the mobile content producer. This is due to the increase of routing table size and scaling problem that is not solved in NDN. Also, namespace structure using hierarchy contributes towards the substantial challenges when content producer moves to another location [12]. Similarly, for a moving object, a new route needs to be announced and propagated to replace old routing information causing a similar problem as in IP. To solve the problem by improving namespaces, whereby changing to flat name violates NDN architecture [13]. Content producer in NDN architecture causes a problem that resulted in long handoff latency, signaling overhead cost and unnecessary Interest loses. It happens due to the transmission of pending Interest towards the previous location of content producer on mobile [9], [10]. When the producer moves to a new location, Interest packets were still following the prefix trace in the Forwarding Information Base (FIB) that leads them to be dropped, without reaching the producer. Hence, the situation makes the handoff time to be very long.

Recently, in survey articles for mobility support in NDN by Feng, et al. [7] and Saxena, et al. [14], reported that consumer mobility was inherently supported in NDN while producer mobility was not supported, also faces many challenges from initial architectural design. Furthermore, based on the literature, almost all ICN approaches support consumer mobility, and some like Data-Oriented Network Architecture (DONA) and Network of Information (NetInf) supported both consumer and producer mobility. However, in a case of Publish Subscribe Internet Technology (PURSUIT) and Publish-Subscribe Internet Routing Paradigm (PSIRP), there are uncertainties of producer mobility support and unanimously agreed that producer mobility was not supported in NDN. However, in accordance to the previous researchers [8]–[12], it was proved that producer mobility was really in needs of further research to find a concrete solution, for NDN architecture to confidently replace and solves the shortcomings of the current Internet.

Consider Figure 1 that shows a scenario of consumer mobility support for general approaches of ICN. The figure shows consumer one connected with Content Router (CR3) 3, sends an interest to the location of producer that is connected to CR6. A mobile consumer two is also connected to CR2 sending an Interest to the same producer. Suddenly, the consumer move to CR5. Let assumed another content consumer is in need of the same content, suddenly connected to the network and sent an Interest packets to the access point or router. The router will check for the related content and sent back to the consumer, if not found it will forward to available faces of the network as indicated in step 1. Upon reached the source content, which is a producer, the requested content will be forwarded back to the consumer that requested it through breadcrumb style as labeled in step 2. At the same time the intermediate routers have the capability to cache the content and stored for subsequent uses.

In addition, many consumers can request the same content once connected to the router and sent an Interest packets. The current router will serve them without forwarding the request into the network due to the cached copy that is available in the content store. The consumers are allowed to move while receiving the requested content, the pending Interest packets are resent to the network immediately after reconnection with any available router.

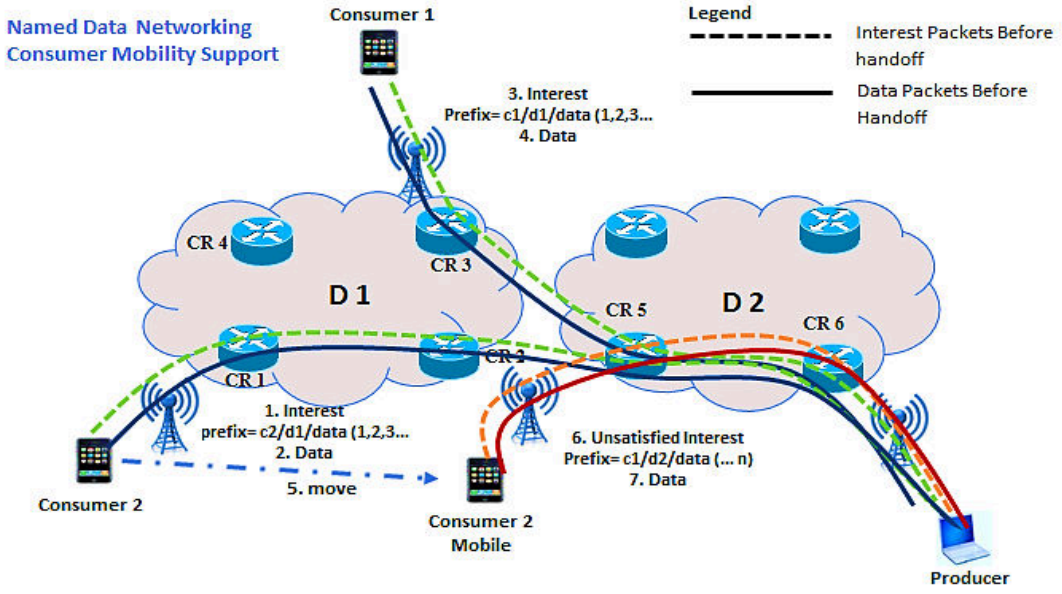


Fig. 1. NDN's Consumer Mobility Support

The architectural perspectives of NDN categorized mobility as mobile content consumer and producer. In the survey report of Feng et al. [7] and Saxena et al. [14], stated that consumer mobility in NDN is fundamentaly supported, when content consumer moves to another location, or new PoA, it will sends the unsatisfied Interest packets with power of NDN consumer-driven nature [15]–[18]. However, the content producer faces some challenges similar to IP architecture when content producer move to new location, such as scaling problem of routing table [8], scalability challenge that significantly affected the entire network [12]. Moreover, handoff delay, latency are getting longer and

unnecessary Interest packet losses towards looking of the new location of content producer [9], [10], the frequent routing update for a large network domain can cost high bandwidth utilization [12], [19] triangular routing and the tunnelling problems. The novel concepts of NDN such as security, in-network caching, hierarchical naming or namespace, named content, name-based routing suit the requirements of Internet of Things (IoT) [20]. Despite the benefits of NDN in IoT network, the mobility perspective needs to be studied and provided a concrete solution to support mobile IoT [20]. Besides, the services that are based on NDN with blockchain technology can be applied in IoT for secured data management [21], [22]. Moreover, decoupling sender and receivers, in-network caching and hop-by-hop transmission characteristics of NDN enable it to support 5G network [23]. The encouraging results of ICN motivated certain technical specification for an organization to promotes ICN as an enabler of 5G [24]. Therefore, the inherent benefits of ICN with support of name-based networking and mobility is to solve the inherited problems of the 5G network [24]. Also, the mobility management in 5G network slicing-based faces challenges caused by high density and high mobility of devices [25]. Consequently, the researchers are applying NDN to the heterogeneous network towards 5G as one of the potential developments [25], and the characteristics features emerged ICN as a promising 5G technology candidate [26].

Conclusively, there is intense anticipation that NDN can successfully become future Internet architecture with the capability to incorporate other networks without additional mechanism. Thus, Cisco proposed hybrid ICN (CCN and NDN) into 5G network [27], which encourage many researchers to leverage the deliverance of ICN into 5G network, [24], [28] and NDN-5G-SDN support [28], [29], IoT networking [30]–[32] and NDN vehicular ad-hoc networking [33], [34]. Consequently, further studies are motivated particularly for mobility support perspective of Named Data Networking.

II. Tracing-Based Mobility Support

Kite is a Tracing-Based (TB) mobility support scheme for the mobile producer, proposed by NDN research group to solve NDN producer mobility problem. The kite was proposed by Zhang et al., [35] uses two new features to support producer mobility; that is, locator-freeness and scenario-awareness. Moreover, Kite utilizes the state of PITs to trace and reached mobile nodes, also makes the new location of a mobile producer transparent through routable anchors [35]. This scheme is called a Pending Interest Table (PIT)-based approach [19] or anchor-based approach [36], where the mobile producer sends a traced Interest to the immobile anchor router, notifying it is a new location and establish trace in each PIT tables along with the intermediate routers to the immobile anchor. The corresponding node or immobile server requesting content sends an Interest to the immobile anchor that modified the Interest by adding a trace-name field and trace-only flag, which is called tracing Interest. The tracing Interest will be sent along the PIT trace to the current location of the producer. The transmission is shown in Figure 2.

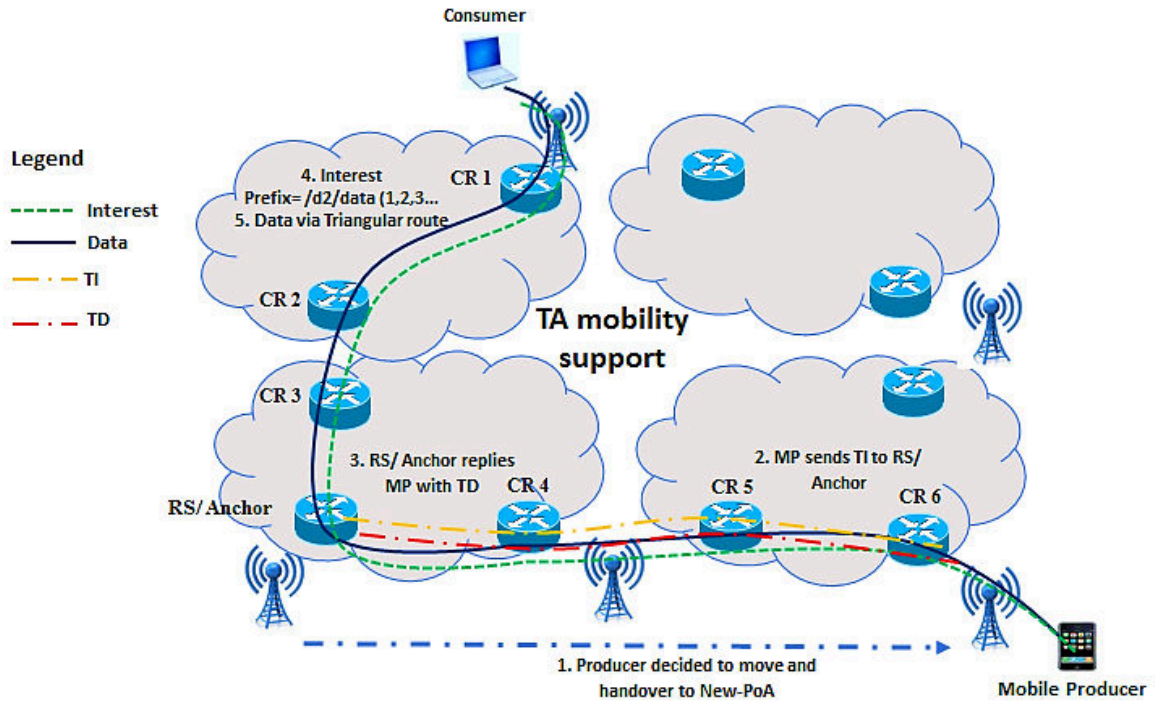


Fig. 2. Tracing Based Mobility Support

Recently, Kite scheme was improved by replacing immobile anchor with rendezvous server [37]. Still, the new Kite maintained the tracing-based and operates in the forwarding plane where the forwarding path is set up from immobile rendezvous to mobile producer. The rendezvous announces a routing prefix, and mobile producer issued a trace Interest with a special tag to the rendezvous. The rendezvous verifies the trace Interest and responds with

trace data to update the intermediate routers as trace setup processes [37]. The consumer sends Interest to the rendezvous and then forwarded along the trace to the mobile producer. The processes are similar to old Kite except for the change of tracing and traced Interest that is replaced with trace Interest and trace data in addition to consumer Interest.

However, the Kite scheme outperforms the mapping-based approach in terms of handoff delay with similar signaling overhead when mobile nodes frequently move [38], [39]. Also, the optimal path is not guaranteed in such a way that consumer's Interest packets always pass through the immobile anchor or rendezvous. Hence, Kite scheme falls into the category of indirection-based approach, that can make the forwarding path longer than the path computed in normal NDN routing plane [15]–[18].

III. Proposed Broadcasting Approach Mobility Support

The Broadcasting Approach (BA) proposed a new Mobility Interest (MI) packet. The MI design carries the producer's new location information to update the intermediate routers. Hence, the broadcasting strategy is designed in such a way that when content producer move to the new location, the new Content Router (CR) that accommodates a connected mobile producer will configure the new prefix and set the mobility flag tags the MI packets and broadcast to the available faces within the domain. Once the next router receives MI packets, the look-up is performed to check if there is exist any prefix records before handoff that need to be updated with the new prefix after. Otherwise, the router will record the new prefix in FIB and forward to available face.

Usually in NDN, if a content consumer sent an Interest requesting data when the Interest reached a content producer, the data is sent back immediately. The Interest can be dropped if the data and producer not found due to the change of location. In BA, when producer moved, and handoff process takes place by updating the intermediate routers about the new routing prefix that can be used to locate a mobile producer. The routing plane will create the best route for Interest to reach the new PoA of mobile producer once the handoff is completed. Then a pending Interest can be forwarded by the consumer via the optimal route towards the new location of the mobile producer, as shown in Figure 3. The Interest and data exchange processes take place with the help of three data structure mechanisms called Content Store (CS), PIT and FIB. The PIT mechanism has a table that can be used to records all Interest's name prefixes and incoming faces; the CS table records the name prefixes and corresponding data; the FIB table records the name prefixes and next or outgoing faces.

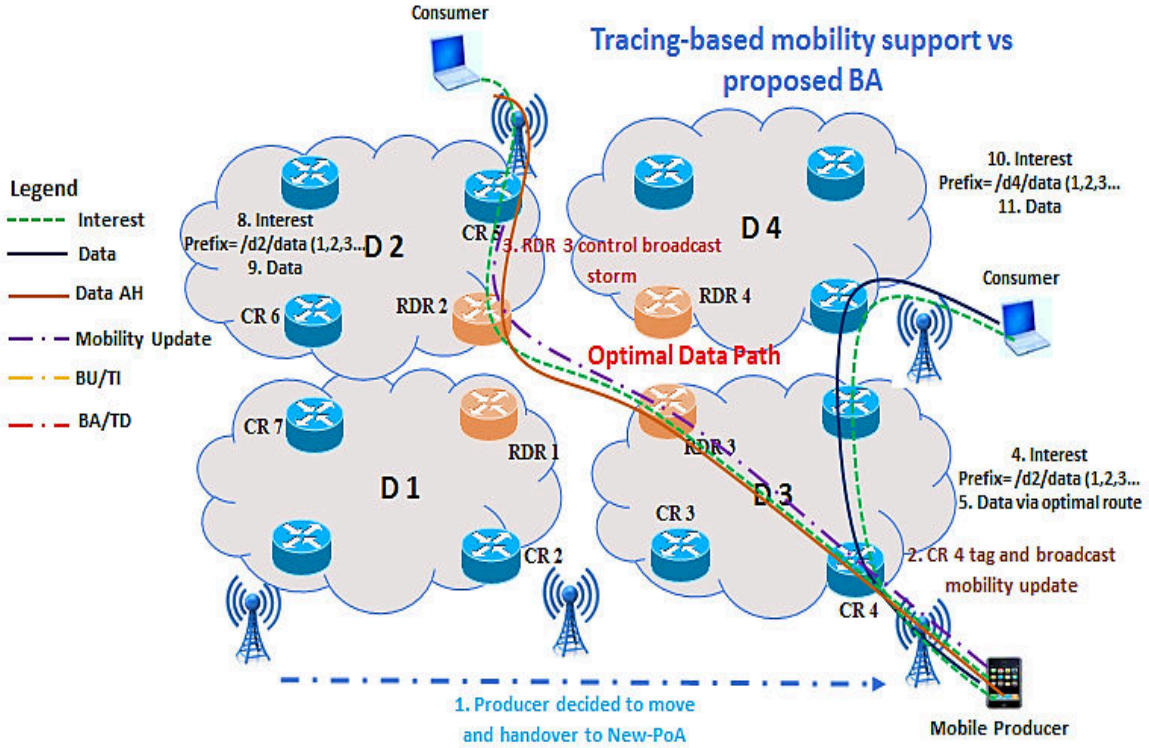


Fig. 3. Broadcasting-Based Mobility Support

III. Implementation

The BA solution is implemented in NDN Simulator called ndnSIM that operates and based on an open source NS-3. In particular, ndnSIM offers uppermost and reliable simulation results and allow the simple transfer of the simulation experiments to the real application. Also, ndnSIM is implemented using different C++classes in a modular passion for modeling the behavior of NDN entities such as Faces of communication with applications and other nodes, FIB, PIT and CS. It allows easy replacement or modification of any component without an impact on the other components. The design of ndnSIM contains the NDN protocol stack (ndn:: L3Protocol) as a core component of

the ndnSIM, ndn-cxx library and the NFD forwarding component for packet processing, which offered to the researcher flexibility to simulate different strategies with different scenarios and namespace. As a network-layer protocol model ndnSIM can run on any available link-layer protocol model such as point-to-point, CSMA, wireless, LTE, etc. Also, ndnSIM provides a collection of interfaces for face, application face abstractions, and network device face. The tracing of simulation output can be collected using helpers such as application, FIB, global routing, application, link control, strategy choice helpers and NDN stack to perform detailed tracing behavior of NDN traffic flow for every component.

A. Simulation Parameters

Simulations parameters used in this research contain the mobility model, speed, number of mobile producers, topologies and NDN set-up. Two different topologies Abilene and the 4x4 grid are used for two scenarios with different instances. The speeds of the mobile producer are ranged between 0m/s - 100m/s, depending on the scenario set-up. The summary of the parameters with their respected values are presented in Table I. The representation of 4x4 grid topology used is presented in Figure 3. The figure shows the position of the consumer node, mobile producer node, restricted domain routers, and rendezvous node position. Moreover, the figure demonstrates how a mobile producer moves along CR1, CR2, and CR4. It presents the data and Interest path, data path after handoff (Data AH), which is optimal path traced over mobility update information provided by MI packets, Trace Interest (TI) and Trace Data (TD).

TABLE VIII
SIMULATION PARAMETERS AND VALUES

Category	Parameters	Value
Mobility	Mobility Model	Random Waypoint Mobility
	Mobility Speed	1 – 100 ms
	Number of Mobile Nodes	1 – 5
	Network	802.11n
NDN	Forwarding Strategy	BestRoute
	Cache Replacement	LRU, 100 items on each node
	Cache size	1000 objects
	Interest rate	10/sec
	Data packets size	1024 bytes
	Interest packet size	24 bytes
	Application	Consumer and Producer
Topology	Abilene	1 Gbps link
	Grid 4x4	1 Gbps link
	Link Delay	10ms

B. Simulation Topologies

The topologies used in this research are 16-nodes-grid (4x4) and Abilene topology with the addition of mobile producers' nodes ranging from 1 to 5. The topologies are created in separate files and used ndnSIM annotated topology reader during the run of the simulation. In addition, the routers and consumers are linked using the point-to-point connection with 1Gbps data rate, 10ms channel delay, and 10 maximum packets for transmission queue per link on both directions. For each scenario, the program calls the specified topology file during the simulation run. The Abilene and nodes positioning is presented in Figure 4, while the 4x4 grid is in Figure 3. Both topologies were used precisely by Kite, a benchmark solution.

C. Simulation Scenarios

There are two simulation scenarios used for both Kite and BA solutions that are based on the two different types of topologies, for the performance evaluation of BA simulation model. The two scenarios are based on the 4x4 grids, and Abilene topology, as presented in Figure 3 and Figure 4, and the scenarios are described below:

Grid Scenario: In this scenario, the infrastructural topology is a 4x4 grid that consists of 19 to 23 nodes, with a distance of 100m between them. Among the nodes, there are three consumers and one mobile producer by default, while the simulation runs for 100 seconds. The mobile content producer moves randomly within the area covered by a different access point (AP) of 400x400 square area at varying the constant speed of 10m/s to 100 m/s. Also, the grid nodes serve as routers and APs. The scenario is run for the proposed BA, Kite upload, and pull scheme.

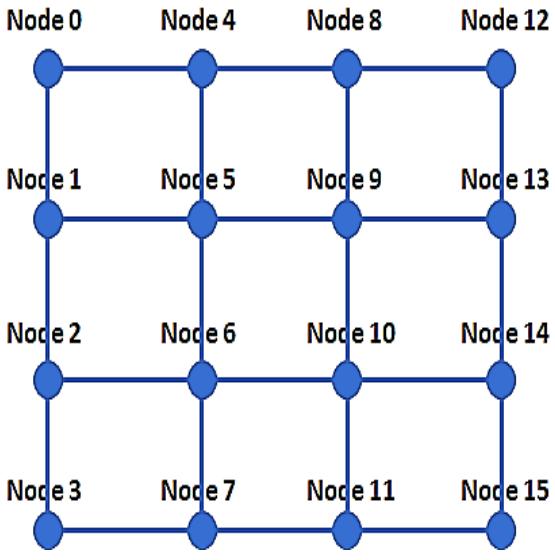


Fig. 3. 4x4 Grid Network Topology

Abilene Scenario: This scenario differs with Scenario I as it consists of 12 to 13 nodes in Abilene topology. Among the nodes, the mobile

producers' number is varied from 1 to 5 and the simulation time runs for 100 seconds, while producers move randomly at a constant speed that is varied from 10m/s to 100m/s. The scenarios are run for Kite and BA solution. Furthermore, to determine the data packets throughput, the simulation time is varied from 0sec to 100sec with the interval of 5sec for both Scenarios.

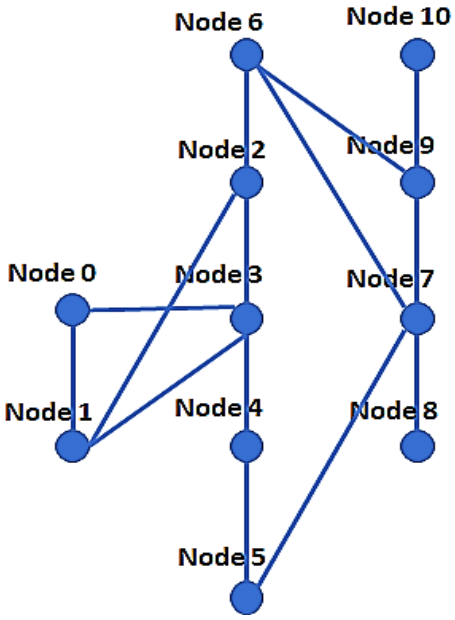


Fig. 4. Abilene Network Topology

D. Implementation

The proposed BA and TA are implemented into ndnSIM C++ programming environment for verification, validation, and evaluation preparation of proposed BA and benchmark solution. The programming codes are integrated with the ndnSIM modules that match with the specification of this study. Moreover, the codes of different scenarios are verified by configuring the ndnSIM module, compile and link the C++ headers included in the scenario's file. The verification is successful once the waf command is run without errors, and the displayed screen shows that the codes file is build finished successfully.

IV. VALIDATION OF PROPOSED BROADCASTING APPROACH MOBILITY SUPPORT

The validation process of the proposed BA is successful, and the validation output result is presented in Figure 5, 6, 7 and 8.

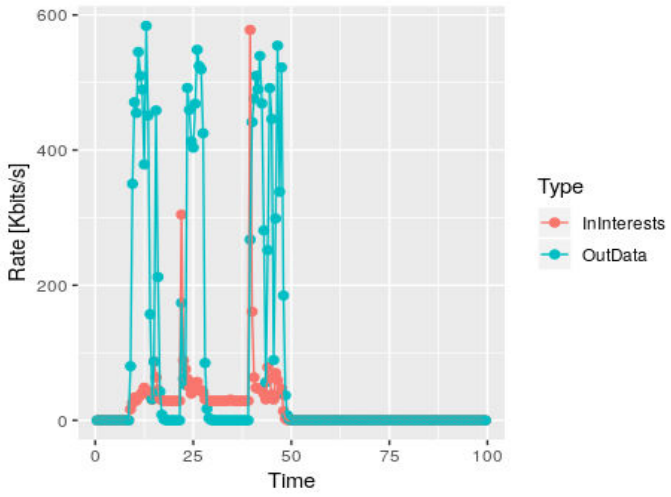


Fig. 5. Grid Scenario Data Rate Transmission for TA

The BA is run together with Kite solution for both scenarios, and the results are plotted as data rate on the y-axis against the simulation time on the y-axis. Figure 5 shows the behaviors of the output for Kite looks similar to that of BA in Figure 6 with regards to the In-Interest and Out-Data. The only difference is that in BA, the Out-Data is stabilized due to the improvement of path optimization and reduction of unnecessary Interest packets lost as seen in Figure 7 and Figure 8.

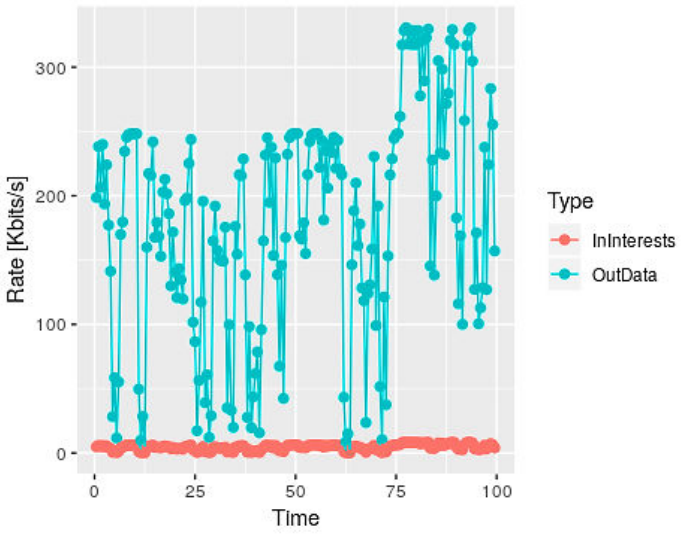


Fig. 6. Grid Scenario Data Rate Transmission for BA

The figure above (Figure 5 and 6) shows the output of producer mobility support after handoff. The rate shows how the proposed solution BA has better result compared to TB. The same result is shown in Figure 7 and 8 for the Abilene scenario. This is because of the nature of NDN that is designed as broadcasting friendly.

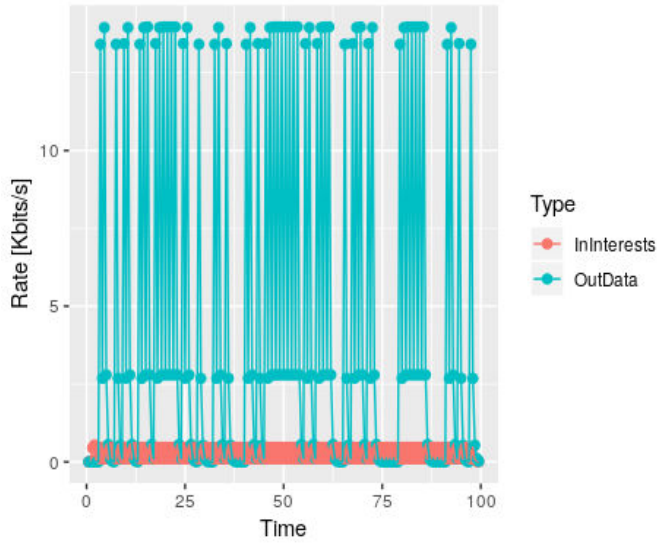


Fig. 7. Abilene Scenario Data Rate Transmission for TA

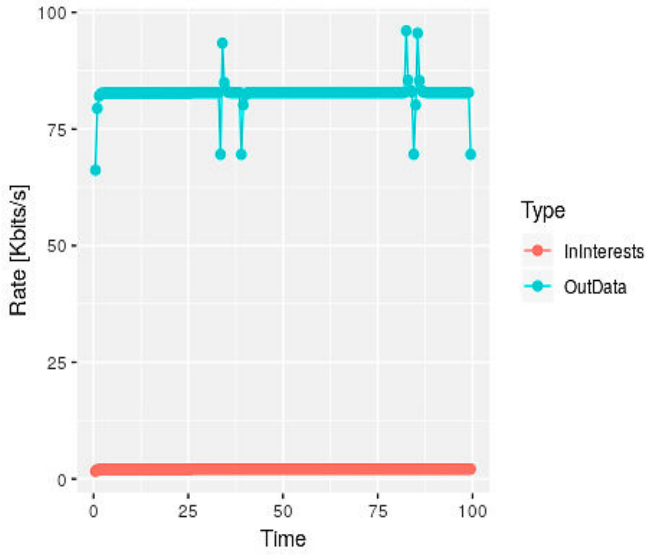


Fig. 8. Abilene Scenario Data Rate Transmission for BA

V. CONCLUSION

The proposed solution will support the content provider to relocate without disrupting consumer and intermediate routers for content (data) availability and its location, within minimal hand-off time by providing an update binding mobility information. The result will solve the inherited problems of tunneling and triangular routing in IP network mobility. Also, improve the handoff performance of mobile producers to curtail disruption delay, signaling overhead and Interest packet loss rate. Similarly, without special mechanism, 5/4G, Mobile Ad-hoc Networks (MANET), Delay-Tolerant Network and WiFi can be supported by NDN for mobile application.

The real-time data is considered a delay sensitive that may affect the performance of producer mobility. In this research, specific data is not given a priority, whereas real-time data such as audio and video content have a considerable influence and high percentage in the global Internet traffic. There is a need to further research by considering the impact of real-time application, especially with the research demand of integrating NDN with IoT, 5G, and wireless sensor networking.

REFERENCES

[1] A. Chaabane, E. De Cristofaro, M. A. Kaafar, and E. Uzun, "Privacy in content-oriented networking:Threats and contermeasures," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 3, p. 25, 2013.

[2] G. White and G. Rutz, "Content Delivery with Content- Centric Networking," *CableLabs, Strateg. Innov.*, pp. 1–26, 2016.

[3] M. Hussaini, S. A. Nor, and A. Ahmad, "Optimal broadcast strategy-based producer mobility support scheme for Named Data Networking," *Int. J. Interact. Mob. Technol.*, vol. 13, no. 04, pp. 4–19, Apr. 2019.

[4] M. Hussaini, S. A. Nor, and A. Ahmad, "Analytical modelling solution of producer mobility support scheme for named data networking," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 5, pp. 3850–3861, 2019.

[5] C. Perkins, "RFC 5944: IP Mobility Support for IPv4, Revised," *Ad Hoc Networks*, vol. 11, no. 5, pp. 1556–1570, 2013.

[6] C. J. Bernardos, "Proxy Mobile IPv6 Extensions to Support Flow Mobility," *Internet Engineering Task Force (IETF) Request for Comments: 7864*, pp. 1–19, 2016.

[7] B. Feng, H. Zhou, and Q. Xu, "Mobility support in Named Data Networking: a survey," *Eurasip J. Wirel. Commun. Netw.*, vol. 2016, no. 1, 2016.

[8] Z. Zhu, A. Afanasyev, and L. Zhang, "A new perspective on mobility support," *Named-Data Networking Project, Technical Report NDN-0013*, pp. 1–6, 2013.

[9] Y. Rao, H. Luo, D. Gao, H. Zhou, and H. Zhang, "LBMA: A novel Locator Based Mobility support Approach in Named Data Networking," *China Commun.*, vol. 11, no. 4, pp. 111–120, 2014.

[10] Y. Rao, D. Gao, and H. Luo, "NLBA: A novel provider mobility support approach in mobile NDN environment," in *2014 IEEE 11th Consumer Communications and Networking Conference, CCNC 2014*, 2014, no. CCNC, pp. 188–193.

[11] G. Tyson, N. Sastry, R. Cuevas, I. Rimac, and A. Mauthe, "A Survey of Mobility in Information-Centric Networks: Challenges and Research Directions," in *Proceedings of the 1st ACM workshop on Emerging Name-Oriented Mobile Networking Design-Architecture, Algorithms, and Applications*, 2012, pp. 1–6.

[12] G. Tyson, N. Sastry, R. Cuevas, I. Rimac, and A. Mauthe, "A survey of mobility in Information-Centric Networks," *Commun. ACM*, vol. 56, no. 12, pp. 90–98, 2013.

[13] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A Survey of Information-Centric Networking," *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 26–36, 2012.

[14] D. Saxena, V. Raychoudhury, N. Suri, C. Becker, and J. Cao, "Named Data Networking: A Survey," *Comput. Sci. Rev. Elsevier*, vol. 19, pp. 15--55, 2016.

[15] M. Hussaini, S. A. Nor, and A. Ahmad, "PMSS : Producer mobility support scheme optimization with RWP mobility model in Named Data Networking," *Int. J. Commun. Networks Inf. Secur.*, vol. 10, no. 2, pp. 329–339, 2018.

[16] M. Hussaini, S. A. Nor, and A. Ahmad, "Producer mobility support schemes for named data networking: A survey," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, pp. 31–42, 2018.

[17] M. Hussaini, S. A. Nor, and A. Ahmad, "Producer Mobility Support for Information Centric Networking Approaches: A Review," *Int. J. Appl. Eng. Res.*, vol. 13, no. 6, pp. 3272–3280, 2018.

[18] M. Hussaini, S. A. Nor, A. Ahmad, I. S. Maijama'a, A. Isah, and A. Aminu, "Mobility Support and Operation of Information Centric Networking Approach," *Int. J. Comput. Netw. Inf. Secur.*, vol. 10, pp. 18–25, 2018.

[19] T. X. Do and Y. Kim, "Optimal provider mobility in large-scale Named- Data Networking," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 10, pp. 4054–4071, 2015.

[20] M. Amadeo, C. Campolo, A. Iera, and A. Molinaro, "Named data networking for IoT: An architectural perspective," in *EuCNC 2014 - European Conference on Networks and Communications*, 2014, pp. 1–5.

[21] S. Ding, J. Cao, C. Li, K. Fan, and H. Li, "A Novel Attribute-Based Access Control Scheme Using Blockchain for IoT," *IEEE Access*, vol. 7, pp. 38431–38441, 2019.

[22] H. K. Yang, H. J. Cha, and Y. J. Song, "Secure identifier management based on blockchain technology in NDN environment," *IEEE Access*,

vol. 7, pp. 6262–6268, 2019.
[23] M. Amadeo, C. Campolo, and A. Molinaro, “Empowering 5G network softwarization through Information Centric Networking,” *Internet Technol. Lett.*, vol. 1, no. 2, p. e30, 2018.
[24] R. Ravindran, A. Chakraborti, S. O. Amin, A. Azgin, and G. Wang, “5G-ICN: Delivering ICN Services over 5G Using Network Slicing,” *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 101–107, 2017.
[25] H. Zhang, N. Liu, X. Chu, K. Long, A. H. Aghvami, and V. C. M. Leung, “Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges,” *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, 2017.
[26] J. Augé, G. Carofiglio, G. Grassi, L. Muscariello, G. Pau, and X. Zeng, “MAP-Me: Managing anchor-less producer mobility in Information-Centric Networks,” *IEEE Trans. Netw. Serv. Manag.*, vol. 15, no. 2, pp. 596–610, 2018.
[27] Cisco, “Mobile Video Delivery with Hybrid ICN IP-integrated ICN solution for 5G,” pp. 1–14, 2017.
[28] Z. Li, Y. Xu, K. Liu, X. Wang, and D. Liu, “5G with B-MaFIB Based Named Data Networking,” *IEEE Access*, vol. 6, no. 2018, pp. 30501–30507, 2018.
[29] S. Din, A. Paul, and A. Rehman, “5G-enabled hierarchical architecture for Software-Defined Intelligent Transportation System,” *Comput. Networks*, vol. 150, no. 2019, pp. 81–89, 2019.
[30] D. Mars, S. Mettali Gammar, A. Lahmadi, and L. Azouz Saidane, “Using Information Centric Networking in Internet of Things: A Survey,” *Wirel. Pers. Commun.*, no. 2019, pp. 1–17, 2019.
[31] S. Arshad, M. A. Azam, M. H. Rehmani, and J. Loo, “Recent advances in Information-Centric Networking based Internet of Things (ICN-IoT),” *IEEE Internet Things J.*, vol. 14, no. 8, pp. 1–31, 2018.
[32] M. Meddeb, A. Dhraief, A. Belghith, T. Monteil, K. Drira, and S. Gannouni, “AFIRM: Adaptive forwarding based link recovery for mobility support in NDN/IoT networks,” *Futur. Gener. Comput. Syst.*, vol. 87, pp. 351–363, 2018.
[33] J. A. Khan and Y. Ghamri-doudane, “ROVERS : Incentive-based recruitment of connected vehicles for urban big data collection,” *IEEE Trans. Veh. Technol.*, 2019.
[34] A. A. Prates, I. V. Bastos, and I. M. Moraes, “GeoZone : An interest-packet forwarding mechanism based on dissemination zone for content-centric vehicular networks R,” *Comput. Electr. Eng.*, vol. 73, no. 2019, pp. 155–166, 2019.
[35] Y. Zhang, H. Zhang, and L. Zhang, “Kite,” *NDN, Technical Report NDN-0020, 2014.*, no. 1, NDN Project, pp. 1–10, 2014.
[36] D. Kim and Y. Ko, “On-demand anchor-based mobility support method for Named Data Networking,” in *19th International Conference on Advanced Communication Technology (ICACT), 2017*, 2017, pp. 19–23.
[37] Y. Zhang, Z. Xia, S. Mastorakis, and L. Zhang, “KITE: Producer mobility support in Named Data Networking,” in *ICN ’18: 5th ACM Conference on Information-Centric Networking (ICN ’18)*, 2018, pp. 1–12.
[38] M. Hussaini, S. A. Nor, A. Ahmad, R. Mustapha, and A. F. Abdulateef, “A Conceptual Model of Producer Mobility Support for Named Data Networking using Design Research Methodology,” *IAENG Int. J. Comput. Sci.*, vol. 46, no. 3, pp. 1–11, 2019.
[39] M. Hussaini, M. A. Naeem, B.-S. Kim, and I. S. Maijama’a, “Efficient producer mobility management model in Information-Centric Networking,” *IEEE Access*, vol. 7, pp. 42032–42051, 2019.

Privacy Token Technique for Protecting User's PII in a Federated Identity Management System for the Cloud Environment

1st Maria M. Abur
Department of Computer Science,
Ahmadu Bello University, Zaria,Nigeria
mmabur1@gmail.com

2nd Sahalu B. Junaidu
Department of Computer Science,
Ahmadu Bello University, Zaria,Nigeria
abuyusra@gmail.com

4th Afolayan A. Obiniyi
Department of Computer Science,
Ahmadu Bello University, Zaria,Nigeria
aaobiniyi@gmail.com

3rd Saleh E. Abdullahi
Department of Computer Science,
Ahmadu Bello University, Zaria,Nigeria
seabdullahi@yahoo.com

***Abstract*—Researches have shown that, using the internet to carryout transactions and sharing of data on the cloud can expose the user to various security risks. User's Personal Identifiable Information (PII) is needed to access Services on the Cloud from different Service Providers (SPs). However, these SPs may sometimes violate user's privacy by making users PII available to others without the users' consent thereby causing damage to the users. Similarly, sensitive user's PII are received in their original form by the needed SPs in plaintext. As a result of these problems, user's privacy is being violated. This paper introduces a conceptual framework for the proposed user's PII privacy protection in a federated identity management system for the cloud. Also, on the proposed system, the use of pseudonymous technique called Privacy Token (PT) is employed. The composition of the PT makes it difficult for the User's PII to be revealed and further preventing the SPs from being able to keep them or reuse them in the future without the user's consent. The prototype was implemented with Java programming language and its performance tested on CloudAnalyst simulation and measured based on: Computational time, Response time and Cost. The behaviour of the proposed system presented lower computational time of 0.82ms for processing user's request than with the existing system of 0.9ms. Similarly, no significant difference between both systems in terms of response time and cost were recorded.**

Keywords—Cloud, Personal Identifiable Information, Privacy Token, Privacy, Services and user's privacy.

I. INTRODUCTION

Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (such as networks, applications, and services). These resources can be fast provision and released with negligible management effort or service provider collaboration. Cloud Computing is characterized by Security challenge, specifically privacy concern of user's PII; [1-5].

Once an individual employs the use of the Internet for; accessing information, carrying out transactions and sharing of data on the Cloud, they are connected to diverse computers on the network. As such, security of such transmitted data is most threatened and then potentially creating privacy risks on the federated identity management system (FIMS) in the Cloud for users' identity [6-8].

Federated identity management is an arrangement that can be made between several organizations to allow users apply the same identification data to acquire access to the networks of all the organizations in the group. Usually, User's Personal Identifiable Information (PII) are needed to access Services on the Cloud from Service Providers (SPs) [1-3, 6-8]. Researches have shown that security of transmitted data is most threatened as such potentially creating privacy risks for users on the federated identity management system in the Cloud. Usually, User's Personal Identifiable Information (PII) is needed to access Services on the Cloud from different Service Providers (SPs). However, these SPs may by themselves sometimes violate user's privacy [11]. This usually happens when the SPs reuses user's PII offered them for the purpose of releasing services to them. Without getting consent from the users, carrying out activities that may appear malicious and then causing damages to the users. Similarly, sensitive user's PII (e.g. first name, email, address and the likes) are received in their original form by the needed SPs in plaintext. As a result of these problems, user's privacy is being violated. Since these SPs may reuse them or connive with other SPs to expose a user's identity in the cloud environment [6-11].

The objective of the paper is to provide an effective solution that should preserve user's PII whenever they try to access services on the cloud through the use of privacy token. This research introduces a protective and novel approach that shall no longer release original user's PII to SPs but pseudonyms that shall prevent the SPs from violating user's privacy through connivance. The paper introduces a conceptual framework for the proposed user's PII privacy protection in a federated identity management system for the cloud. On the proposed system, the use of Privacy Token (PT) is employed. The PT ensures users' original PII values are not sent directly to the SP but auto generated pseudo PII values. It is composed of: pseudo PII values, timestamp and SP_ID. The composition of the PT makes it difficult for the User's PII to be revealed and further preventing the SPs from being able to keep them or reuse them in the future without the user's consent for any purpose. Another important feature of the PT is its ability to forestall collusion among several collaborating service providers. This is due to the fact that what the SPs receive are just randomly generated pseudo values and do not contain the real user's PII values. The prototype was implemented with Java programming language and its performance tested on CloudAnalyst simulation.

The subsequent sections are organized as follows: Section II introduces some preliminary concept and technologies relevant to the research; Section III describes related work; Section IV illustrates the design of the conceptual framework for the proposed system; Section V describes the formalization of PII dissemination amongst federated parties; Section VI demonstrates the Simulation and the performance evaluation of the proposed system against the existing system. Finally, section VII presents conclusion of the paper.

II. PRELIMINARY CONCEPT AND TECHNOLOGIES

This section deals with definition of concepts that reflect ideas that are essential to the understanding of the research in this paper.

A. Federated identity management system (FIMS)

FIMS is an arrangement that can be made between two or more trust domains that allows users to access applications and services using the same digital identity [6-8]. This kind of identity is referred to as federated identity and the application of such is called identity federation. Federated identity management is built upon the basis of trust between two or more domains [6-11]. In federated identity management, the following parties are identified: 1) Identity Provider is the party of a federated identity management system that creates, maintains and manages identity information for the users and provides users' authentication to other service providers within a federation. 2) Service Providers is the party of a federated identity management system that denotes organizations that provide services or resources desired by a user, by requesting for the submission of valid credentials from the user's IdP. 3)User is the party of a federated identity management system that depicts persons whose PII are kept on the Identity Provider (IdP). Other important FIMS concepts are described as: 1) User's PII or Personally Identifiable Information (PII) which refer to any information usually used to uniquely identify a person whether alone or by combination with other public data that could be connected to a particular person [1-3, 6-14]. It is also called Personal Identifiable Information. In this paper, User's PII and PII are used interchangeably. 2) Services/resources are presented to users on request via the Web from a Service Provider (SP). 3) Privacy is the claim of persons, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others. Technologies that provide support to federated identity include: Shibboleth, OpenID, OpenAm, CardSpace, Liberty and OAuth, Chadwick [3], [6-8]. This research is focused on the Shibboleth systems for it pervasive acceptance in academia.

B.Shibboleth

Shibboleth is a joint project of Internet2 and IBM. It is open-source based systeBm that supports inter-institutional resource sharing with access. Shibboleth enables exchange of interoperability of services [6-14]. It employs the idea of federated identity and Single Sign-On (SSO) authentication [1-3, 6-14]. As a federated Identity Management system, Shibboleth is liable for forming the identity of a user, i.e. creating, maintaining and managing identity information for the user, and also managing access to services by the user [6, 7, 9, 10 &11]. It encourages interaction between the IdP, SP and user. The entire system of Shibboleth and the proposed system is built on mutual trust between the IdP and SPs i.e. whatever is sent by the IdP is accepted after authentication to represent user's information Typically, Shibboleth system comprises majorly of two parts: Identity Provider and Service Provider as shown on the Shibboleth Architecture of [9] on Fig. 1 with the flow of operations explained immediately below it.

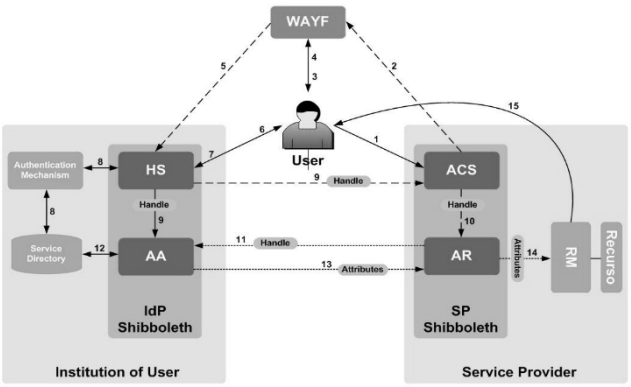


Fig. 1: Architecture of Shibboleth system [9]

The flow of operation of Shibboleth architecture as represented on Fig. 1 are:

In Step 1, the user routes to the SP to access a secure resource. In Steps 2 and 3, Shibboleth passes on the user to the Where Are You From (WAYF) page, where he should notify his IdP. In Step 4, the user goes into his IdP, and Step 5 passes on the user to the site, which is the component Handle Service (HS) of the IdP. In Steps 6 and 7, the user enters his verification data and in Step 8 the HS authenticate the user. The HS creates a handle to identify the user and sends it also to the Attribute Authority (AA). Step 9 sends that user authentication handle to AA and to Assertion Consumer Service (ACS). The handle is verified by the ACS and transferred to the Attribute Requester (AR), and in Step 10 a session is established. In Step 11 the AR uses the handle to request user PII to the IdP. Step 12 checks whether the IdP can release the PII and in Step 13 the AA responds with the attribute values. In Step 14 the SP receives the PII and sends them to the Resource Manager (RM), which loads the resource in Step 15 and gives to the user.

In the Shibboleth architecture of Fig. 1, it can be noticed that the flow of information from user to SP, then to IdP and then back to the user is lengthy and thereby exposing users to security threat. Examples of the security threats are: data leakage such as sniffing, spoofing, eavesdropping and malicious attack on the network, [12-14]. Secondly, there is delay in WAYF populating the IdP for User's selection. Thirdly, during dissemination of user's PII from the IdP to the SP, the original user's PII values are sent to the SP. The PII are then kept or further reused again by the SPs without the consent of the user for malicious purposes and thus, violating the user's privacy.

Also, it should be noted that no performance metrics were used on the existing systems to evaluate the performance of their system. Furthermore, it is worth noting that as Shibboleth has been discussed for almost a decade, the lack of extensive discussion on previously

published work in the area with similar objectives is a problem.

This research is motivated to provide a protective and novel solution that shall no longer release original user's PII to SPs but privacy token. This shall preserve user's privacy whenever they try to access services on the cloud.

III. RELATED WORK

Shibboleth is an open-source based system that supports inter-institutional resource sharing with access [1-3, 6-14]. Shibboleth enables exchange of interoperability of services [1-3, 6-14]. It employs the idea of federated identity and Single Sign-On (SSO) authentication where there is interaction between the IdP, SP and User. However, there are still limited features and functionality that threatens user's privacy and identity if they store and process personal information with inadequate protective measures [15]. Also, from the Shibboleth architecture in [9, 10, 11 & 16], it can be noted that the flow of information from user to SP, then to IdP and then back to the user is lengthy and thereby exposing users to security threat on the network. Secondly, there is delay in WAYF populating the IdP for User's selection. Thirdly, during dissemination of user's PII from the IdP to the SP, the original user's PII values are sent to the SP. The PII are then kept or further reused again by the SPs without the consent of the user for malicious purposes and thus, violating the user's privacy, [11].

The research in [16] added uApprove plugin – a user consent module for shibboleth identity providers. The uApprove plugin displays to the user, the PII needed by the requesting SP on behalf of the user through the IdP. It also offers awareness of data release when accessing some services. As a limitation, users cannot make a choice of data that should be divulged to the Service Provider. Similarly, the client has no option assenting/dissenting with His/her PII disclosure. During the dissemination of user's PII to SPs on the network, data leakage is envisaged as user's PII are usually exposed to security threat such as spoofing, sniffing, eavesdropping and malicious attack, [12-14] and the SPs having received these PII use them maliciously either directly or indirectly against the users without their consent and then leading to collusion [11]. This way, the users' privacy is being threatened.

Similarly, the research in [10] added "uApprove.jp, a user consent acquisition system (UCAS) with an attribute-filter mechanism for a Shibboleth based SSO system". uApprove.jp request the "user's consent and enable the user control the release of his/her original PII" values or PII values whether mandatory/optional from the IdP to the SP and then allows the user to determine which of his/her original PII values are meant to be sent to the SP in order to access services provided by Service Providers. The uApprove.jp is meant to strengthen uApprove in [16]. However, the user's control of his/her PII is ineffective making them vulnerable as there is still data leakage. Also original users' PII values are released to the SP by the IdP are sometimes maliciously shared among other SPs or even used without the user's consent to either harm the user directly or indirectly hence leading to violation of the user's privacy and causing collusion problem.

Also, the work in [11] improved on the research of [10] by adding two objects namely: Template Data Dissemination (TDD) and Cryptography Encryption Key Technique on Shibboleth/uApprove.jp framework. The former object helps users manage their PII's release during the course of dissemination of their PII from the IdP to the SPs; help them manage their PII release. Also, Users contact SP directly for resources, during which user's PII can be attacked by man in middle (MIM) attack and DoS causing leakage of the user's PII [12 & 14]. Furthermore, TDD allows users to send their PII in a plaintext to the required SP in order to release resources to them. Thus exposing those users' PII to data leakage (Spoofing, Sniffing attack and malicious insider attacks) on the network [12 & 14]. while the later enable users to encrypt their PII before sending them to IdPs with Key I, and during a transaction when some PII data is needed users would be entreated to open that data with key II in order to disseminate it to a SP. The Cryptography Encryption Key Technique used by [11] is the Rivest–Shamir–Adleman (RSA). As limitation, three basic weaknesses were identified on this system. The first is centred on the encryption techniques used i.e. the RSA which have been compromised and user's sensitive information leaked [17-19]. The second weakness identified is the exposure of the transmitted PII on the communication medium, which attracts potential hackers due to data leakage, [12 & 14]. Thirdly, SPs having received these PII may keep them and use them without the user's consent or even maliciously use them against users in the future. This leads to collusion.

The work in [9] illustrated the flow of operations of the Shibboleth architecture and discussed its main components in details. It was observed that the flow of information from user to SP, then to IdP and then back to the user is lengthy and thereby exposing users to security threat such as data leakage on the network [12 & 14]. Secondly, there is delay in WAYF populating the IdP for User's selection. Thirdly, during dissemination of user's PII from the IdP to the SP, the original user's attribute values are sent to the SP. This PII are then kept or further reused again by the SPs without the consent of the user for malicious purposes and thus, violating the user's privacy.

Also, it should be noted that no performance metrics were used on the existing systems to evaluate the performance of their systems. Furthermore, it is worth noting that as Shibboleth has been discussed for almost a decade, the lack of extensive discussion on previously published work in the area with similar objective is a problem.

However, the new approach is aimed at providing an effective solution that should preserve user's PII whenever they try to access services on the cloud through the use of privacy token from security threats such as data leakage, malicious SPs that seek to connive in order to expose user's identity for self-gain reasons.

IV. DESIGN OF THE CONCEPTUAL FRAMEWORK FOR THE PROPOSED SYSTEM

This section gives general overview of activities for the Design of the conceptual framework for the proposed system; In section IV A the framework for the proposed solution is explained; Section IV B illustrates operations of the proposed privacy protection in a federated identity system; Section IV C discusses the major components of

the proposed system; while Section IV D presents the proposed dissemination algorithm between federated parties and finally Section IV E describes the pseudonymous technique or privacy token.

A. Conceptual Framework for the Proposed Solution

The framework in the proposed solution adopted, modified and improves the Shibboleth architecture in [9]. It is adopted due to the fact that some components of the architecture of [9] are present in the proposed framework. However, the functionalities of the components were modified to be in consonance with the reality of the conceptual framework for the proposed privacy protection in a federated identity system. As an improvement, new components such as Privacy Token Channel, Muapprove.jp and proposed PII Privacy protection model were added to the adopted version of the Shibboleth architecture. The flow of activities in the proposed framework is different from the existing system. The entire system of Shibboleth and the proposed system is built on mutual trust between the IdP and SPs i.e. whatever is sent by the IdP is accepted after authentication to represent user's information and this is why pseudo values could be accepted by SPs since it comes from a recognised IdP. It is therefore safer and simplifier to initiate a request from the user through His/her IdP to the SP than going to the SP to make a request before coming to authenticate the user from his/her IdP through the (Where are you from (WAYF)) service.

The framework for the proposed solution consists of the following components: Identity Provider (IdP) with the PPPPM and M_uapprove.jp, Privacy Token Channel (PTC) and Service Provider (SP). These components are described in Section IV C of this paper. Fig. 2 displays the conceptual framework of the proposed system.

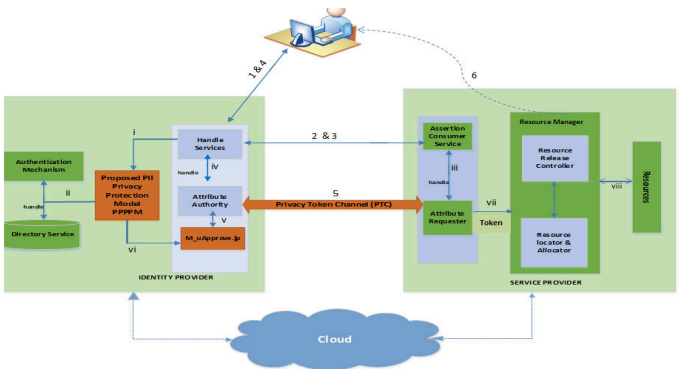


Fig. 2: Conceptual framework for the proposed privacy protection in a federated identity system

From the Fig. 2, it should be noted that solid lines are used to represent direct communication between two components of the system while dashed lines are used to represent indirect communication. Also numbers are used to label communication between the three different entities i.e. the user, IdP and SP whereas roman numeral are used to label communication within each entity. The dotted lines at the bottom of the Fig.2 indicated that every activity happens within the cloud environment.

B. Operations of the proposed privacy protection in a federated identity system.

1. User logins into his/her IdP through the Handle Service (HS) which then communicates with the Authenticate Mechanism in order to authenticate the user through the Proposed PII Protection Model (PPPPM).

2. The HS connect the user to the SP's Attribute Consumer Service (ACS) in which the user needs to access resource from.

3. The SP's ACS relates with the SP's Attribute Requester (AR) to request for the PII required for that resource and then this goes back to the HS.

4. The IdP through the HS component gets user's consent for the release of the pseudo value.

5. Then HS interact with Attribute Authority (AA) which then communicates that SP's request for PII to the M_uapprove.jp where mapping and auto generation of user's pseudo values are done. These pseudo values are sent to the AR through the Privacy Token Channel (PTC). Upon receipt of the Privacy Token (PT) by the AR, it then sends PT to the Resource Release Controller (RRC) of the Resource Manager (RM) where the PT is acknowledged and verified and passed to the Resource Locator and Allocator (RLA). The resource is allocated and sent to the user.

6. The resource is released to the user via IdP.

C. Major Components of the proposed framework

The components of the proposed conceptual framework as represented in Fig. 2 are:

1) Identity Provider (IdP) of the Proposed System.

The IdP of the proposed system is the component in charge of authenticating and verifying users. It upholds users' PII as in the IdP of the Shibboleth Architecture of [9]. However, the IdP of the proposed system is responsible for generating and disseminating user's information (i.e. Pseudonyms) to SPs that request them for the release of resources. This is usually based on the User's consent. In this research, pseudonyms are the pseudo values that are auto generated to replace the original attribute values of the user. This is meant for the purpose of preserving the user's identity when a SP request for them in order to release resources to the user. It comprises of the following parts:

Handle Service (HS): When a user visits an IdP, the IdP requests user's login details of which the user must provide in order to be logged unto the IdP. HS communicates with Authentication Mechanism (AM) in order to authenticate the user through the PPPPM. Also upon authenticating the user on the IdP, the IdP communicates through HS with the SP's Assertion Consumer Service (ACS) to connect the SP to the user. The SP terms of usage is sent back from ACS to the user through the HS. However, the HS of the Shibboleth Architecture of [9] communicates with AM and allows an organization to choose the authentication mechanism which is not so in the HS of the proposed framework.

Attribute Authority (AA): On the proposed system, AA performs attribute resolving, PII filtering along with handling requests for Service Providers. And also communicating with the SP requesting for PII to the M_uapprove.jp components where mapping and auto generation of the user's pseudo values are done. Again the

AA receives user's generated pseudonyms from M_uapprove.jp and sends them in SAML assertion form to the AR of the SP requesting for them through the privacy token channel (PTC). More so the AA directly communicates with the PT channel on behalf of the IdP. In summary AA serves as the gateway between the IdP and the SP.

Directory Service (DS): On the proposed framework, the DS help users locate network resources to their respective network addresses. It is used for managing and organizing every item and network resources. DS also communicates with the AM. However, the DS of the Shibboleth Architecture of [11] serves as a reservoir for keeping users PII while in the proposed model the user's PII are kept on the PPPPM in more secured manner.

Authentication Mechanism: This permits users to authenticate with the IdP using a single login/password pair. AM communicates with the HS via the PPPPM in order to authenticate the user when the user visits the IdP. However, the AM of the Shibboleth Architecture of [11] is external to Shibboleth and admit users to authenticate with central services.

M_uApprove.jp: M_uApprove.jp of the prototype system communicates with AA and PPPPM to map and auto generate user's pseudo values when request is made for them. It is modified based on the shortcomings of uApprove.jp of the existing system of [10 & 11] as highlighted in section III of this paper. Although the goal of uApprove.jp in the existing system is to request the user's agreement and enable the user control the release of his/her original PII values whether mandatory/optional from the IdP to the SP and then allows the user to determine which of his/her original PII values are meant to be sent to the SP in order to access resources/services provided by the SPs.

2) Proposed PII Privacy Protection Model (PPPPM):

The *PPPPM* is situated between HS and AA of the proposed Identity Provider. This is to help preserve users attribute in a more secured manner. The proposed system builds a more secured system which will make up for the weaknesses of the existing systems. The processes involved in the proposed PII privacy model for the effective control of user's attribute is described in [3]. This is shown in Fig. 3.

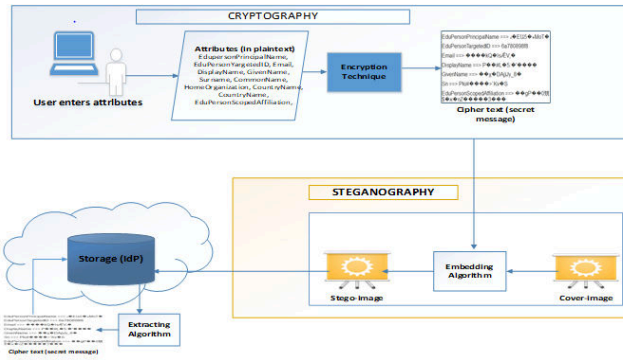


Fig. 3: Prototype PII Privacy Protection Model (PPPPM) [3]

3) Privacy Token Channel (PTC)

The prototype model introduces use of Privacy Tokens (PT) generated and disseminated by IdP to SPs on behalf of users to release resources to them when users request for services/resources from SPs. The PT is further explained in Section IV E, of the paper. The PTC acts as a conveyor which serves the purpose of transmitting user's privacy tokens from the Identity Provider to the Service Provider.

4) Service Provider (SP) of the Proposed System:

The Service Provider of the Proposed System is the place resources are kept and can be made available to the user on request just like the existing system. It applies access regulation on resources based on confirmation sent by the IdP. A particular SP may be consisting of numerous applications, but shall still be treated as a single element by the IdP. Also similar to the SP on the Shibboleth architecture of [9], the proposed SP has three main components and maintains additional functionalities. The three main components are: Assertion Consumer Service (ACS) responsible for receiving messages to establish a secure environment as in the case of the ACS in the Shibboleth Architecture of [9]. ACS connects the SP to the user through HS on the IdP. The SP terms of usage is sent from ACS to the user via the HS. Also the ACS informs the Attribute Requester about the user's request for resource. Another component of SP is the Attribute Requester (AR) that receives information of user's request for resources from ACS as in the same way the AR in the Shibboleth Architecture of [9] is done. It is also in charge of requesting user's PII, receiving PII response (i.e. the Privacy Token (PT)) and passes this user's pseudo PII values received from IdP to the Resource Manager. However, in the Shibboleth Architecture of [9]. AR receives the original user's PII value and not pseudo values as in the proposed AR. Another component of SP is Resource Manager (RM) responsible for receiving, acknowledging, verifying and using the received Privacy Token from the AR and then using it to grant the user access to resource. PT gives no chance for malicious intention of SPs assembling user's attribute to harm them in future. Conversely, the RM on the Shibboleth Architecture of [9] receives original user's PII values from the IdP and uses them to release resources to user. In addition, the RM of the proposed SP has two (2) components that assist in the release of resources requested by users: Resource Release Controller (RRC) and Resource Locator & Allocator (RLA).

D. Prototype Dissemination Algorithm between federated parties

The prototype algorithm for the effective dissemination of users' PII between the IdP and SP is illustrated on the effective disseminator algorithm with it corresponding SP authentication algorithm on Algorithm 1 and Algorithm 2 respectively. This is used to acknowledge the receipt of the Attribute and verifying the user's token before releasing the resources to the user. The algorithm is to aid meeting the needs of cloud users by preserving their privacy during dissemination of finding secure and reliable services and not exposing their PII for any dubious intention.

Algorithm 1: Effective disseminator

Algorithm PII_Dissemination():

1. Input: sp_list //list of SPs;
2. authenUsername //user whose PII are to be disseminated
3. Output:service page

```
4.  PII[] : Array
5.  agree: boolean
6.  acknowledge: boolean
7. BEGIN
8. display sp_list
9. sp=selected from sp_list
10. PII:= list of PII required by sp
11. agree:=display consent page with required PII
12.  IF agree THEN
13.          tokenValue:="";
14.          delimeter:= "@";
15.         FOREACH PII as PIIName=>attributeValue
16.                      pseudoPII = generatePseudo(PIIValue)
17.         token = IdP.Rand(pseudo PII,Timestamp, sp_Id)
18.         tokenValue=tokenValue+ delimeter + token
19.         tokenName=PIIName+ delimeter
20.          END-FOREACH
21.         MappedPIIValues= add(tokenName, tokenValue)
22.         acknowledge= IdP.send(sp, mappedPIIValues,
23. authenUsername)
24.  IF acknowledge
25.         release services/resources
26.         ELSE
27.               wrong sp selection
28. END-IF
29. END
```

Now, Rand ( ) definition on the IdP for the Effective disseminator in Algorithm 1 is described by:

1) Pseudo PII values
2) Timestamp
3) SP_ID

Therefore, the composition of the description in 1) – 3). constitutes the Privacy Token.

Algorithm 2: spAuthentication

```
1. Algorithm spAuthentication():
2.     Input: sp, mappedPIIValues, authenUsername
3.    Output:acknowledge, and services/resources
4.     attributes:= list of PII required by sp
5. BEGIN
6.      delimeter:= "@";
7.     WAITING_REQUEST_LIST := read all authenicated users for the
sp
8.     IF authenUsername exists in WAITING_REQUEST_LIST   THEN
9.        unmappedPII=key(mappedPIIValues.split(delimeter)) =>
10.       value(mappedPIIValues.split(delimeter))
11.          FOREACH unmappedPII as aName=>aValue
12.                      IF ! contains(PII, aName=>aValue) THEN
13.                                  return false
14.          END-FOREACH
15.                                  release services/resources
16.                                  return true
17.     ELSE
18.              return false
19.  END
```

E. *Pseudonymous Technique or Privacy Token (PT)*

In this paper, the Privacy Token is referred to as the auto generated user's attribute that do not contain original user's PII value but pseudo values as required by each Service Provider (SP) in a secure-manner without exposing the identity of the user. The PT is composed of: Pseudo Attribute values, Timestamp and SP_ID. The user pseudo PII being transmitted from the IdP to the SP is secured using a timestamp and SP_ID with these security features in place, the Privacy Token that is generated by the IdP has a specified usage timeframe for which it becomes invalid at it expiration. This is put in place to forestall reuse of the Token by the SP. Also, the PT is generated for specific SP and as such the PT is not usable by other Service Providers. In the event of data leakage, only the Token could be accessed because it contains randomly generated pseudo values and not the original user's PII values. Similarly, collusion as observed in the existing system of [11] has been prevented.

V. FORMALIZATION OF PII DISSEMINATION AMONGST FEDERATED PARTIES

A. *Formalizing Shibboleth dissemination of PII:*

Shibboleth dissemination of PII deals with the way users' PII are collected on the Cloud and disclosed to Organizations (Service Providers (SPs)) that require them for the release of resources to the user. It has become necessary to ensure that user's PII be protected in a privacy preserving manner during and after dissemination. On the existing system, user's PII are endangered to data leakage (sniffing, spoofing and malicious attack) on the network making their PII to be exposed during and after dissemination to SPs [13 & 14].

1) *Existing System*

Assuming a user U requests for a resource R from the SP, Si where 1<=i<=n and G represents an entity IdP. Si requests the IdP G, for the users PII A_i, where $A_i = (a_{j_i})$ is the set of requested PII of U and $A_i \subseteq A$. In the existing system, the IdP requests the user's consent for the release of the requested PII to be sent to the SP as shown on the dissemination relation between U, G and Si on fig 4.

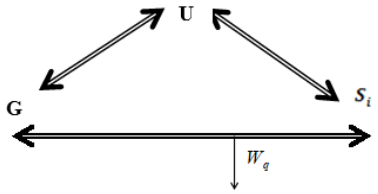


Fig 4: Dissemination relation between U, G and S_i

The user U communicates with SP, S_i requesting for resources R and S_i requested for PII A_i from the IdP, G. The G seeks user's consent to release the requested PII Ai to S_i after which the IdP releases the requested PII in plaintext to the SP that made. Consequent upon this, the requested resources R, are released to the user.

It can be noted on this system that in the process of transmitting the requested PII in plaintext from G to the S_i, data leakage is possible. Secondly, after the release of the requested resources, the PII in plaintext still remain with the SP S_i and this could be maliciously used to collude to profile a user identity in a cloud environment (Collusion). This usually happen without the consent of the user. Also, the illicit use of data, as well as distributing with other parties, has become a communal pattern among Service Providers thereby violating user's privacy [10-11].

2) *Proposed model*

The data leakage and unauthorized usage of users PII by the Service Providers resulting to Collusion is of concern to this research and as such, special interest is

taken to prevent the occurrence of such attacks. To achieve this, the following modifications are introduced:

a) Modification of flow of activities:

In the proposed model the user logs on to the IdP to access available Service Providers, SPs, selects the SP with the required resources. The SP prepares a list of PII that are required to access the resource and sends to the IdP. The IdP then presents this list of PII to the user to give consent for the release of the PII to the requesting SP. Upon obtaining the consent from the user, the IdP generates a Privacy Token (in place of the original PII) and send it to the requesting SP for the release of the requested resource.

b) Security of the User PII.

The user pseudo PII being transmitted from the IdP to the SP are secured using a timestamp and SP_ID. With these security features in place, the Privacy Token that is generated by the IdP has a specified usage timeframe for which it becomes invalid at it expiration. This is put in place to forestall reuse of the Token by the SP. Also, the PT is generated for specific SP and as such the PT is not usable by other Service Providers. In the event of data leakage, only the Token could be accessed which do not contain original user's PII values. Collusion as noted in the existing systems have been prevented due to the fact that what the SPs receive are just randomly generated pseudo values and does not contain the real user's PII. The proposed dissemination relation between the user U, IdP G and SP S_i is presented in fig.5.

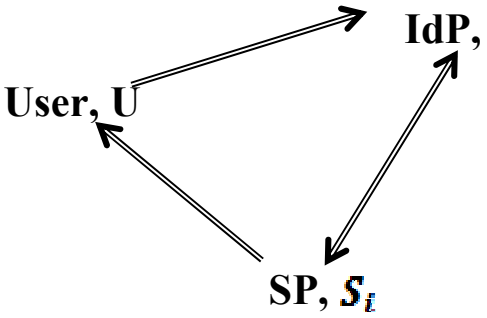


Fig.5: Proposed dissemination relation between the user U, IdP G and SP S_i

VI. Simulation and Performance Evaluation of the Proposed System Against the Existing System.

This section deals with the implementation of the Prototype system and further testing of it performance against the existing system. Section VI A demonstrates the implementation of the proposed system; Section VI B illustrates the security analysis of the privacy token while Section VI C explains the performance of the proposed verses existing system.

A. Implementation of the Proposed system

The proposed system was implemented with Java programming language on a Windows 10 system with 8GB RAM size, and the processor of Intel(R) Core(TM), i3-5005U CPU @ 2.00GHz of 64-bit operating system and x64-based processor.

The prototype developed incorporates the entities mainly involved in the federated Cloud environment namely: the PII, user, Identity provider and Service Provider. A layer of security was presented at the Identity provider to enhance user's privacy and then the Privacy Token (PT) was also introduced to ensure that no original PII of a user was released to a Service provider. This is to ensure for further protection of the user's privacy during dissemination on the Cloud.

On prototype system, upon logging on to the IdP, the user is presented with a list of service providers he/she can access then he/she can select the service(s) he/she wants to access from the service provider (SP). Subsequently, the SP presents the user with it Term of Usage (ToU) and the IdP presents the user with consent page on how their PII are to be disseminated and if the user agrees the PT is generated and sent to the SP in order to release resources to the user. The Identity Provider login page is illustrated on Fig 6.

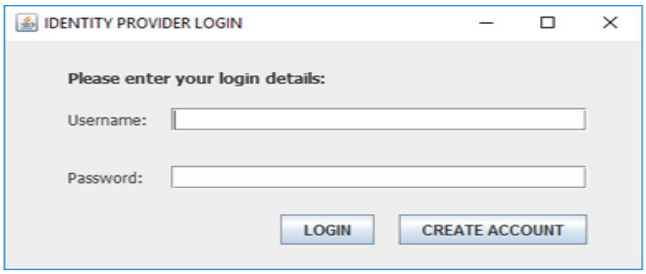


Fig. 6: Identity Provider login page

Similarly, Fig. 7 shows list of Service Providers considered in the research case study.

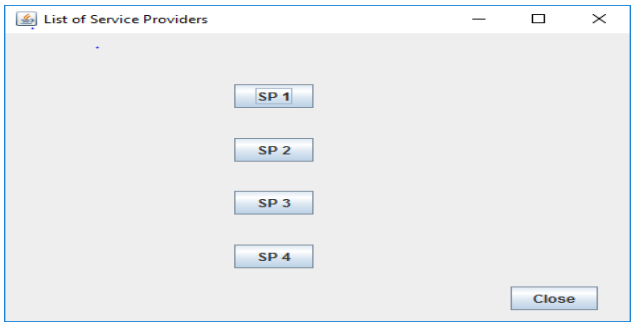
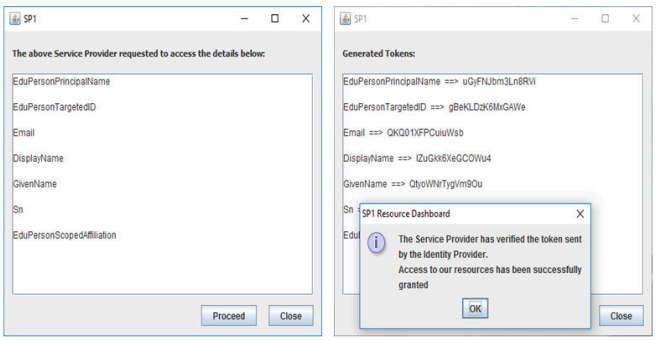


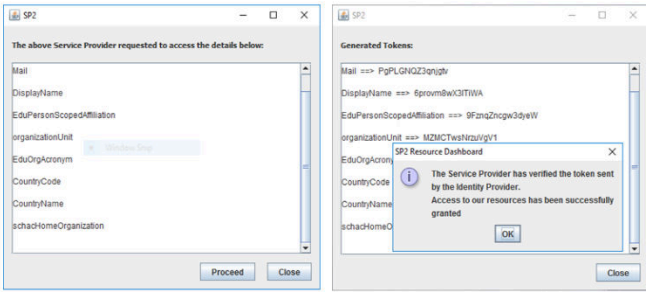
Fig. 7: List of Service Providers

Moreover, Figs 8 & 9 demonstrates (i): List of PII required by SP1 & SP2. (ii) Generated Tokens sent to SP1 & SP2 on behalf of a user respectively.



(i) (ii)

Fig. 8 (i): List of PII required by SP1. (ii) Generated Tokens sent to SP1 on behalf of a user



(i) (ii)

Fig. 9 (i): List of PII required by SP2. (ii) Generated Tokens sent to SP2 on behalf of a user

From the implementation as demonstrated on Fig. 8(ii) and Fig. 9(ii) respectively, it can be observed that the auto generated PT for a user sent to one SP is different from auto generated privacy token (PT) for another SP. By this novel approach, SPs can no longer keep plaintext of original users PII again for any malicious purpose or future use. Similarly avoiding connivance of SPs to profile user's identity for any dubious reasons. The composition of the PT are: pseudo attribute values, timestamp and SP_ID. The user pseudo PII being transmitted from the IdP to the SP is secured using a timestamp and SP_ID with these security features in place, the Privacy Token that is generated by the IdP has a specified usage timeframe for which it becomes invalid at it expiration. This is put in place to forestall reuse of the PT by the SP. Also, the PT is generated for specific SP and as such the PT is not usable by other Service Providers. In the event of data leakage, only the Token could be accessed which has no direct bearing with the Original user's PII. Collusion as noted in the existing system has been prevented due to the fact that what the SPs receive are just randomly generated pseudo values and not the original users' PII. These pseudo values are being used in place of the original user's PII.

Considering a scenario if a user needs to access services from SP2, by clicking on the generate token button, the Privacy Token is generated with timestamp as shown on fig. 10. The user selects the send button and still given the opportunity to select the right SP within the timestamp as modeled in fig. 11. If mistakenly the privacy token goes to another SP e.g. SP3 instead of SP2 the Token is rendered invalid and Token expires within that given timestamp as illustrated on figs. 12 and 13 respectively. If the generated PT goes to the right SP, the SP acknowledges and verifies the token and then processes the users request and finally releases resource to the user.

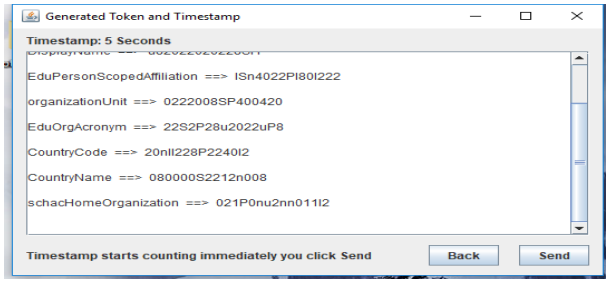


Fig. 10: privacy token is generated with timestamp as shown

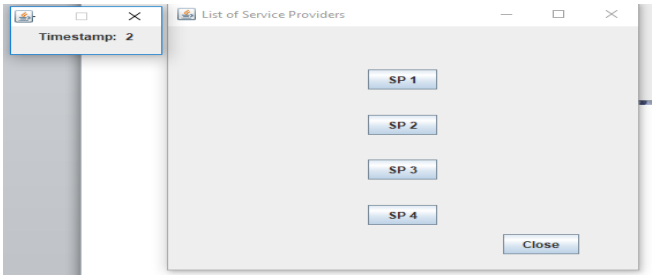


Fig. 11: Opportunity given to user to select Correct SP within a given
timestamp

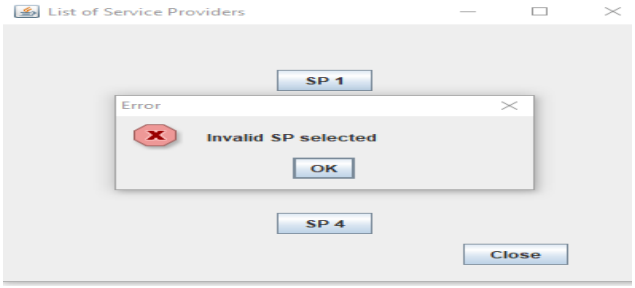


Fig 12: Privacy Token is invalid with wrong SP

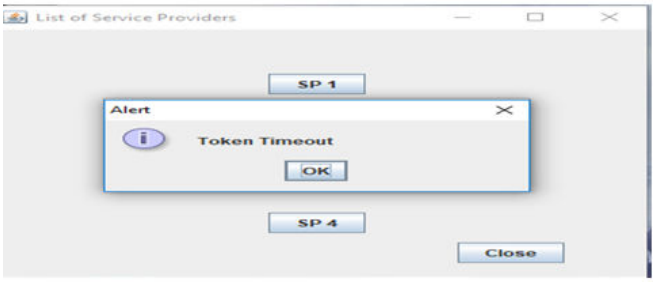


Fig. 13: privacy token expires

A. Security analysis of the Privacy Token

The PT is composed of: Pseudo Attribute values, Timestamp and SP_ID. The user's pseudo PII being transmitted from the IdP to the SP is secured using a timestamp and SP_ID with these security features in place, the Privacy Token that is generated by the IdP has a specified usage timeframe for which it becomes invalid at it expiration. This is put in place to forestall reuse of the Token by the SP. Also, the PT is generated for specific SP and as such the PT is not usable by other Service Providers. In the event of data leakage, only the Token could be accessed which does not contain the user's original PII. Collusion as observed in the existing system of [11] has been prevented. This is due to the fact that what the SPs receive are just randomly generated pseudo values and not the real users' PII. These pseudo values are being used in place of the original user's PII. This way, user's PII is preserved on an Identity management system.

B. Performance evaluation of the proposed verses existing system

The performance of the proposed system was further tested with CloudAnalyst and compared with that of the existing system of [11] based on the following metrics: Response time, Cost and Computational time. During the experiments, the following group of users were plugged on

to the simulator (i.e. 5, 10, 15, 20, 25, 30 and 35) for a single Service Provider on both the existing and proposed system along with detailed parameter settings (such as User growth factor varied from 100 -10000000, Request growth factor of 10 and Execution instruction per length of 100 for the experiment. The experimental result is tabulated on Table 1.

The simulation experiments were executed on a computer with the following configuration: Windows 10 with 8GB RAM size, and the processor of Intel(R) Core(TM), i3-5005U CPU @ 2.00GHz. Furthermore, the computer is a 64-bit operating system and x64-based processor.

TABLE 1: SETTINGS FOR VARIED NUMBER OF USERS ON THE EXISTING AND PROPOSED SYSTEMS WITH CORRESPONDING RESULTS AFTER THE SIMULATION FOR RESPONSE TIME, COST AND COMPUTATIONAL TIMES.

User	No. of SP	Response Time (ms)		Cost (₦)		SP Computational Time (ms)					
						Average		Minimum		Maximum	
		Ext. Sys.	Pro. Sys.	Ext. Sys.	Pro. Sys.	Ext. Sys.	Prop. Sys.	Ext. Sys.	Prop. Sys.	Ext. Sys.	Prop. Sys.
5	1	300.12	300.12	0.82	0.82	0.35	0.35	0.02	0.02	0.65	0.64
10	1	300.04	300.04	1.14	1.14	0.35	0.35	0.01	0.02	0.80	0.80
15	1	299.89	299.89	1.46	1.46	0.35	0.35	0.02	0.02	0.83	0.82
20	1	299.97	299.97	1.78	1.78	0.35	0.35	0.01	0.01	1.00	0.87
25	1	299.93	299.93	2.10	2.10	0.35	0.35	0.01	0.01	1.01	1.02
30	1	299.98	299.97	2.42	2.42	0.35	0.35	0.01	0.02	1.02	0.99
35	1	301.01	300.01	2.73	2.73	0.36	0.35	0.01	0.01	1.10	1.07

From the result of the simulation, it was observed from the experiment in terms of response time, that the behaviour of the proposed system when compare with existing system showed no significance difference. Even though, the proposed system has more protective provision for user's privacy than the existing system. It should be noted that the integration of privacy provision on the proposed system did not slow the response time in any way on the simulator.

Also for the computational time, it can be observed that from the simulated result presented on Table 1: average values of the existing system is slightly dissimilar to the prototype system while the Minimum values of the existing system is lower than that of the proposed system. Maximum values of the existing system are higher than the proposed system. In general, from the results presented, despite the fact that the proposed system has better provision for user's privacy than the existing system, it still presented average lower computational time of 0.8miliseconds (ms) for processing users' request than with the existing system with (0.9 milliseconds (ms)) which lacks effective user's privacy protection and which takes longer time for the computation of users' request. Moreover, the graph on fig. 14 shows the plot of the SPs computational time against the number of users. This is to verify that the behaviour of the proposed system is better when placed side by side with the existing system. Fig. 15 illustrating the comparison of the existing and proposed systems with respect to cost. The cost is remained the same in both cases.

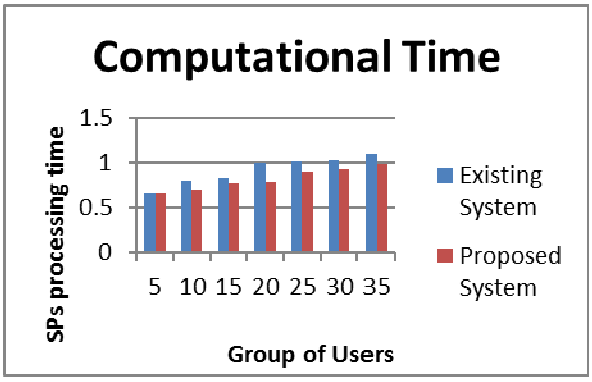


Fig. 14: SPs computational time of existing system against the prototype system.

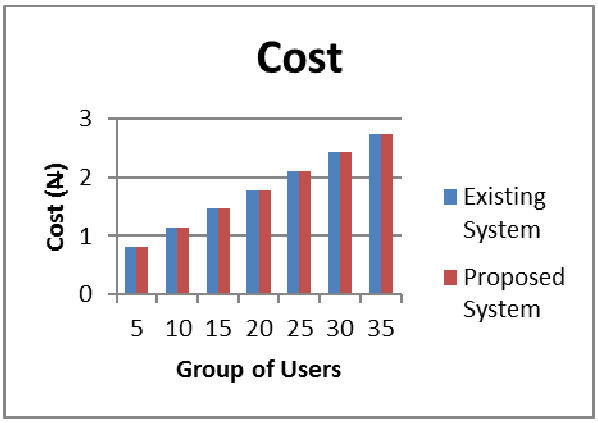


Fig. 15: Cost for proposed verses the existing system.

VII. CONCLUSION

The paper presented a novel approach known as a Privacy Token Technique for Protecting User's PII in a Federated Identity Management System for the Cloud Environment. Privacy Token (PT) technique auto generates user's PII that do not contain original user's information but pseudo values as required by each Service Provider (SP) in a secure-manner. And not exposing the identity of the user. The constituent of the PT makes it difficult for the exposure of user's PII to security threat. Similarly, collusion problems by SPs to violate user's privacy for their self-gain as observed on the existing systems have been tackled.

Also, a conceptual framework of the proposed system was presented. Formalization of PII dissemination amongst federated parties was illustrated. The proposed dissemination algorithm between federated entities were demonstrated.

More so, the prototype was implemented with Java programming language and its performance tested on CloudAnalyst simulation against existing system based on: Computational time, Response time and Cost. The behaviour of the proposed system presented lower computational time of 0.82ms for processing user's request than with the existing system of 0.9ms. Similarly, no significant difference between both systems in terms of response time and cost were recorded. From the above discussions, the proposed system presents a more effective and secures solution to the user than the existing system.

In this research, user's PII and Services were focused on the academic domain only. As future work, researchers should investigate into how user's PII and Services can be modeled and integrated into the Prototype system to cater for other domain such as industries and businesses.

REFERENCES

[1] Abur M. M., Junaidu S. B., Danjuma S., Arlis S., Ritonga R., Herawan T. (2018): Towards a Privacy Mechanism for Preventing Malicious Collusion of Multiple Service Providers (SPs) on the Cloud. In: Bhateja V., Nguyen B., Nguyen N., Satapathy S., Le DN. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, Singapore: Springer, vol 672.

[2] Abur M. M., Adewale O. S. & Junaidu S. B., (2015): Cloud Computing Challenges: A review on Security and Privacy issues. Proceedings of the ACM International Conference on Computer Science Research and Innovations (CoSRI), Ibadan pp. 89-92.

[3] Abur M. M., Junaidu S. B., Obiniyi A. A. and Abdullahi S. E. (2018) "Privacy Protection and Collusion Avoidance Solution for Cloud Computing Users", 1st International Conference on Education and Development (ITED 2018), Base University, Abuja.

[4] Zhou M., Zhang R., Xie W., Quian W. & Zhou A., (2010) Security and Privacy in cloud: Survey. In Proc. Of the 6Th International Conference on Semantics, Knowledge and Grids, IEEE. Pages 105-112.

[5] Chadwick D. W. (n. d.). Federated Identity Management: Computer Laboratory, University of Kent, Canterbury, CT2, &NF, UK.

[6] Chen D. & Zhao H. (2012): Data Security and Privacy Protection Issues in Cloud Computing. Proc. of the 1st International conference on Computer Science and Electronics Engineering, Hangzhou China. Doi: 10.1036/0071393722.

[7] Suriadi S., Foo E., Josang A. (2007). A User-Centric Federated Single-On System. IFIP International Conference on Network and Parallel Computing Workshops.

[8] Teena A. M. and Aaramuthan M. (2017): Federated Cloud Identity Management: A Study on Privacy Tactics, Tools and Technologies. Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, 19(6), 34-40

[9] Leandro M. A.P., Nascimento T. J., Santos D., Westphall C. M., & Westphall C. B. (2014). Multi-Tenancy Authorization System with Federated Identity for Cloud-Based Environments Using Shibboleth. In proceeding of the Eleventh International Conference on Networks (NetWare2014), Lisbon, Portugal. pp. 42-67.

[10] Orawiwattanakul T., Yamaji K., Nakamura M., Kataoka T. & Sonehara N. (2010): "User-controlled privacy protection with attribute-filther mechanism for a Federated SSO environment using Shibboleth," in P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), International Conference on IEEE, pp.243-249.

[11] Weingartner R., Westphall C. M., (2014) "Enhancing privacy on identity providers", Emerging Security Information Systems and Technologies (SECURWARE). The Eighth International Conference Lisbon, Portugal pp. 1-7.

[12] Asha P. N., Mahalakshmi T., Archana S and Lingareddy S. C., (2016): Wireless Sensor Networks: A Survey on Security Threats Issues and Challenges. International Journal of Computer Science and Mobile Computing, 5(5), 249-267.

[13] Cloud Security Alliance (CSA). 2013: The Nine Notorious Threats. Top threats working group.

[14] Cloud Security Alliance (CSA). 2016: The Treacherous 12 - Cloud Computing Top Threats

[15] Aldeen Y. A., Salleh M. & Abdur Razzaque M., (2015): "A Survey Paper on Privacy Issue in Cloud Computing". Research Journal of Applied Sciences, Engineering and Technology, 10 (3): 328-337.

[16] SWITCH, (2010). "uapprove - user consent module for shibboleth identity providers," retrieved: [Online]. Retrieved from: https://www.switch.ch/aai/support/tools/uApprove.html/03/03/2016

[17] Kumar Y., Munjal R. and Sharma H., (2011) Comparison of Symmetric and Asymmetric Cryptography with Existing Vulnerabilities and Countermeasures. International Journal of Computer Science and Management Studies (IJCSMS) 11(3), 60-63.

[18] Tripathi R. & Agrawal S. (2014). Comparative Study of Symmetric and Asymmetric Cryptography Techniques. International Journal of Advance Foundation and Research in Computer (IJAFRC) 1(6), 68-76.

[19] Hacker News (2017): Researchers Crack 1024-bit RSA Encryption in GnuPG Crypto Library. Retrieved July 3, 2017 from wiki: https://thehackernews.com/2017/07/gnupg-libgcrypt-rsa-encryption.html

[20] Lokhande U. and Gulve A. K. (2014): Steganography using Cryptography and Pseudo random numbers. International Journal of Computer Applications, 96 (19), 41-45.

[21] Aleisa N. (2015): A comparison of the 3DES and AES encryption standards. International Journal of Security and its Applications 9(7):241-246 http://dx.doi.org/10.14257/ijsia.2015.9.7.21.

[22] Song J., Lee K., and Lee H., (2014). Biclique Cryptanalysis on the Full Crypton-256 and mCrypton-128. Journal of Applied Mathematics. 2014, 1-10, http://dx.doi.org/10.1155/2014/529736.

DELAY AWARE POWER SAVING SCHEME (DAPSS) BASE ON TRAFFIC LOAD IN IEEE 802.16e NETWORKS.

[1]D. D Wisdom, [2]A. Y Tambuwal, [3]A. Mohammed, [4]A. Audu [5]M. B Soroyewun, [6]S. Isaac
[5]Usmanu Danfodiyo University Sokoto (UDUS), Faculty of Sciences, Department of Mathematics, Computer Unit, Sokoto State, Nigeria.
[1]danieldaudawisdom1@gmail.com,[2]ahmed_tambuwal@yahoo.com,[3]maminuus@yahoo.com,[4]ahmed.audu@udusok.edu.ng,[5]delemike@gmail.com
[5]Department of Computer Science, Faculty of Physical Sciences, Ahmadu Bello University, Zaria.
[2,] Department of ICT UDUS, [6]Department of Computer Science, Kaduna State University [6]samson.isaac@kasu.edu.ng

ABSTRACT
IEEE 802.16 Standard also known as Worldwide interoperability for Microwave Access (WiMAX) is designed to support wider coverage, higher bandwidth, less cost of deployment with different traffic classes support for power savings. The IEEE added mobility characteristics which made battery-life of Mobile Station (MS) a critical challenge, since MS are battery powered with an impose rechargeable life. An Efficient Battery Lifetime Aware Power Saving Scheme was proposed. The Scheme minimizes frequent transition of MS in order-to reduce power consumption but increases average response delay due to a longer sleep interval used. Thus, a Delay Aware Power Saving Scheme (DAPSS) Based on Traffic Load is proposed to reduce the excessive response delay. The Scheme introduces a modified minimum and maximum sleep interval in order to reduce the longer sleep time of a MS, and dynamically tunes the sleep parameters more appropriately according to the traffic load. It employed a delay aware algorithm to save power. The Scheme was evaluated using discrete event simulator, the results showed that the proposed DAPSS achieves superior performance compared to the existing Scheme in terms of the average power consumption and response delay.
KEYWORDS: Battery-Life, Delay-Aware, IEEE, Power-Saving-Scheme.

1. Introduction

The IEEE 802.16 also known as Worldwide interoperability for Microwave Access (WiMAX) is designed to support wider coverage, higher bandwidth, less cost of deployment, quality of service (QoS), with vast traffic classes support to users as well as smaller scale business. As one of the emerging broadband wireless access systems for mobile stations (MS). Formally, the IEEE 802.16 is designed for a fixed MS [1], while subsequent version of the IEEE 802.16e is an extension of the former standard with mobile features so that MS could be move-able (Mobile) [2].
And because of the significance of the mobility characteristics added in the subsequent 802.16e standard, efficiency subsequently became a critical challenge for battery-powered devices since MS are battery powered with a rechargeable supper impose lifetime.

Power Saving Classes (PSCs) of type I, II, and III are designed to address the challenges mentioned above.
Type I is designed for best effort (BE) and non-real-time variable rate (NRT-VR) traffics, it consists of listening intervals as well as sleep intervals which are interleaved respectively (Figure 1). The length of the listening intervals in this power saving class is fixed. A MS with PSC I subsequently checks if there are some buffered packets for it in the listening intervals (Figure 4 and 5). If there are buffered packets, the MS will revert to normal operation mode to receive the packet (s). Else, the sleep window is activated in order to further save power. This procedure is repeated and the length of the sleep intervals is doubled until it reaches the maximum length of the sleep window called Tmax and maintains Power Savings [18]. PSC of Type II is for unsolicited grant service (UGS) and real time variable rate (RT-VR) traffics, similarly type II consists of listening intervals and sleep intervals.
Unlike type I, the length of listening and sleep intervals are both fixed for PSC of type II and the sum of the sleep windows is called, the sleep cycle. PSC II is also capable of transmitting data packets without returning to normal operation. Thus, the length of listening intervals is long enough to receive all packets arriving during a single sleep cycle in PSC II [19] (Figure 1). PSC of type III is use for managing operations and multicast connections. These three PSC differ from each other by their parameter sets, methods of activation/deactivation, as well as the policies of MS availability for data transmission [2]. Unlike PSC of type I and II, PSC III comprises of a single sleep intervals and is mainly use for multicast connection as well as management of operations as seen in Figure 1 below. By activating the PSC, a single sleep interval with defined length in WiMAX Networks begins and subsequently the MS reverts to normal operation mode [20]. Note that sleep window and sleep intervals are same in this paper.

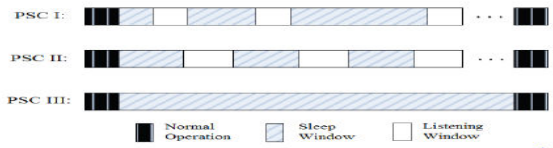


Figure1: Types of PCS in WiMAX Networks
These PSCs used three parameters to improve on power savings, namely, idle threshold, initial sleep interval and final sleep interval [12] [13]. The idle threshold is the

time interval that the MS is in a waiting state, it has no messages to send or receive before moving to inactive state. The MS before moving to inactive state negotiates with it BS for approval in order to switch to a period of inactivity. The BS allocates the sleep parameters namely: initial sleep interval (Tmin), final sleep interval (Tmax) and listening interval (L) to the MS, the MS transmits to it period of inactivity after it receives these parameters [13]. The Tmin is the range of the first-sleep session (T1) as seen in Figure 4 below, that an MS will go to sleep. After which it wakes up for the first T to listen to the traffic indication messages from the BS within the duration of the L. When the traffic indication messages indicate negative, the MS continues to sleep mode after the L duration. Else, the traffic indication message is positive, and the MS return to an active session. The T together with it L is the (T+L) Sleep Cycle. Whenever MS remain in a period of inactivity, the next sleep cycle start as well as the T is doubled. These procedure is repeated till the Tmax is achieved which is the extreme length of the sleep intervals. When a MS achieves Tmax the sleep time is maintained until a Positive MOB-TRF-IND messages arrives from the BS where the MS wakes up to receive/transmit intending packets as seen in Figure 4 below in the third sleep interval (T3).

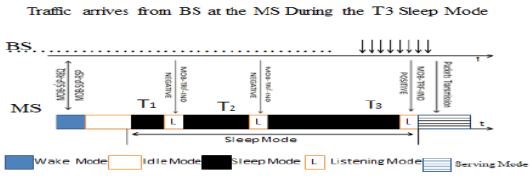


Figure4: IEEE 802.16e Sleep Parameters with variant modes.

Note that at T3 the MS is experience the third sleep window which is larger/longer than the previous windows, and each sleep window is to double the previous one in order to effectively save power. However, it is also observed in this paper as seen in figure 4 above that, in cases of a higher traffic load arrival, the longer sleep window will also incur response delay which will results to congestion, as well as packets loss/buffer overrun respectively (Figure 4).

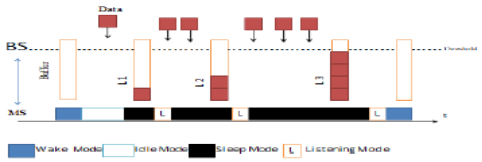


Figure 5: shows the effects of longer sleep period as seen above in the buffer L3 where the traffic load has reached the threshold of the buffer with likely experience of congestion, packet loss or buffer overrun.

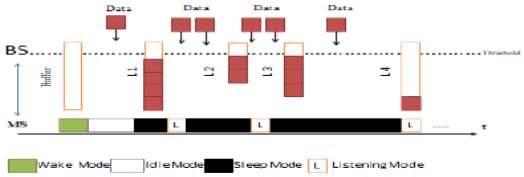


Figure 6: above shows the actual variant buffer sizes of the intermediate devices as in the case of TCP where in between the MS and the BS these devices exist. it can be seen that L2 and L3 have less buffer size compared to L1 and L4 which can take in more data. In situations where there are higher traffics (Load) arrival, congestion and buffer overrun is likely to be experience at L2 and L3 since both have reached their threshold and the MS at this time is in sleep mode and cannot process data until at the listening mode stage where a MON-TRF-IND message is received as Positive as seen in Figure 4 which subsequently results to packet loss as well as performance degradation of the network. The proposed DAPSS scheme have significantly minimize this excessive response delay accordingly, by appropriately adjusting the sleep intervals (time) using the three sleep parameters and employed a delay aware algorithm as seen in algorithm 1 below to improve power savings.

Several Schemes have been proposed in order to improve on power efficiency of MS in [3][4][5] [13] [6][7] [12]. However, the schemes in [3], [4] wastes energy due to their excessive listening operations and or frequent switching frequency from sleep/wake mode while [7][13] use half of it last sleep interval, to adjust the Tmin when it exits from the preceding sleep mode operation as the initiates sleep interval in the next sleep mode operation to reduce the excessive listening operations of mobile station (MS). However, the scheme has excessive response delay due to its longer sleep interval, which also results to packet congestion, buffer overflow as well as overall performance degradation of a MS. Thus, this paper proposes a new Scheme called a Delay Aware Power Saving Scheme(DAPSS) Based on Traffic load for mobile broadband Network Services in order to enhance the parameters of the existing Scheme as well as resolve the critical challenge of the excessive response delays of packets, which subsequently affect performance of a mobile device.

The performance of the propose scheme was evaluated and compared with that of the existing Scheme using discrete event simulator. The rest of this paper is organized as follows: Section II presents Related Literatures, Section III the Propose Scheme, Section IV Procedure of Parameter Adjustments, Section V Performance Evaluation, Section VI Concludes this research.

2. Literature Review

This section presents a related work on existing schemes. These schemes are review by highlighting their Operational, Strength and Weaknesses of each scheme as follows: Energy-Saving Mechanism (ESM) was proposed by [21], to extend the battery-life of mobile stations (MS). The ESM considers the MS to be in sleep-mode during listening intervals but it is active. It increases the sleep interval exponentially when there is no traffic arrival. The mechanism significantly minimized the frame response time and energy consumption at the expense of excessive listening operations, which may lead to waste of energy. An Enhanced Energy Saving Mechanism (EESM) was proposed in [22], to reduce the excessive listening operations of Mobile Station (MS) in the ESM. The mechanism uses half of the last sleep interval in the next sleep-mode operation. When the initiate sleep interval is less than Tmin, then the initiate sleep interval is set to

Tmin. The Base Station (BS) is notified of the initiate sleep interval request message sent by the MS. When the traffic is low the inter-Service Data Unit (SDU) arrival interval is large enabling the EESM to effectively decrease the number of listening intervals in one sleep-mode operation. The mechanism improved energy conservation by extending the lifetime of MSS. However, the mechanism has higher response delay due to the longer sleep interval.

A Delay-Aware Auto Sleep Mode Operation was proposed in [23], to minimize delay and conserve energy of mobile station. The algorithm dynamically tunes the initial sleep window based on the traffic load and the delay requirements, after serving all buffered packets; the MS reverts to the initial-sleep window which depends on the number of packets served. It successfully bound the delay to a certain range. However, it suffers little increase in energy-consumption. Remaining Energy-Aware Power Management Mechanism (REAPM) was proposed by [24], to extend the battery-life of mobile stations (MS) and minimize the response delay.

The REAPM updates the sleep parameters dynamically taking into consideration the remaining energy and the inter-arrival of each frame. The Tmax is updated using smoothing technique with current inter-arrival time of a MAC SDU at each frame arrival and adjust Tmin by considering the energy remaining and the Tmax. More so, after the mechanism initialize parameters such as, the Tmin and Tmax and the current inter-arrival time of MAC SDU. It commences from normal operation mode and terminates this mode when its receives request message to enter sleep mode operation. This mechanism can achieve low-response delay if there is sufficient energy and prolong the battery-life. However, it the scheme has an average increase in energy consumption due to the switching frequency and a constant Listening interval (L) that is between the sleep intervals. A Real-Time Heuristic Algorithm was proposed in [25], to minimize the switching frequency of mobile station (MS). The algorithm operates based on three criteria, the probability of buffer overflow, expected delay violation, and battery lifetime expiry. It uses the probability of the finite buffer overflow to ensures that when the packet arrives during the next sleeping window the expected delay may not exceed the delay violation tolerable bound, and checks the battery lifetime in order to ensure that the power is adequate to extend the sleep time by at least a period, and that the MS still has enough resources to handle the transmission so as to obtain all the packets coming to the buffer during the expected period of time. The algorithm minimizes energy consumption with an increase in the average waiting time.

A Dynamic Traffic Load-Aware Sleep Mode Operation Algorithm was proposed in [26], to enhance the performance of battery powered mobile devices. The algorithm employs a dynamic scheme to tune the idle check time of a MS which is the waiting period after the packet arrival in the wake-mode before entering the next sleep-mode. The idle check time is adjusted dynamically after the entire buffered packet is served based on the number of packets served and the previous sleep window interval. The waiting time is then set to zero when the last sleep window is bigger than the Tmin which makes MS to go to sleep immediately but when the last sleep window is the same as Tmin it is set to a certain value, which makes MS to wait for some time before transiting to sleep-mode. The algorithm improves power savings with a small increase in the algorithm complexity, ignore aligning both the downlink and uplink traffics.

A downlink (DL) and uplink (UL) Alignment (DUAL) Scheme was proposed in [27], to address the asynchronous problem between UL and DL (mobile station and base station connections. The DUAL scheme used the mean packet arrival of UL (λ u) and a relatively safer threshold of buffer size QT as the parameters to determine the maximum waiting interval to support the two opposite connections in order to have effective sleep intervals as possible. The Scheme enhances the energy conservation greatly when UL traffic is greater than DL traffic, it however wastes energy when the maximum threshold value is exceeded which might further lead to poor performance of the mobile device. A Power Saving Mechanism with periodic traffic indications was proposed in [3], to minimize power utilization and delay of mobile station (MS). The mechanism uses traffic indication (TRFI D) messages to initiate transmission at every constant time. The TRF-IND messages consist of a listening interval, awake interval and a sleep interval. During listening intervals, a MS synchronizes with the current base station (BS) and decides whether to switch to awake-mode or remain in a sleep-mode. If there are data traffics in the buffer for the marked MS, the BS sends a positive TRF-IND message and the MS switch to awake-mode. The BS sends information during the awake-mode and the awake-mode terminates if no traffic arrives during a time-out/fixed time of a constant length T. If any data traffic arrives during inactive time **T,** the MS switches to awake mode and transmits the data. Otherwise, goes to a sleep-mode from the awake-mode without switching MOB-SLP-REQ/RSP messages. The mechanism reduced average response delay because of its frequent switching from sleep mode or awake mode at the expense of an increase in energy-consumption. EBLAPS is recommended to solve the above mentioned. An Efficient Battery Life-Aware Power Saving Scheme (EBLAPS) was recommended in ref [4] to reduce average delay and also minimize the energy consumption of mobile station (MS).

The EBLAPS adaptively amend the three parameters:

idle threshold, initial sleep interval, and final sleep interval based on traffic arrival pattern. It employs an upgraded sleep mode control algorithm to consider type I (Non Real Time Services) power saving classes (PSC) in the downlink Operation of the 802.16e in order to minimize the frequent transition to listening mode under low traffic arrival. The scheme reduces the average response delay and the average energy utilization. However, it ignores to consider type II (Real-Time Services) PSC which courses an effect in energy saving, as well as a little increase in the average waiting delay due to it longer sleep period.

3.0 Proposed DAPSS Scheme

A Delay Aware Power-Saving Scheme (DAPSS) for mobile broadband network services (MBNS) Based on Traffic Load is propose, which is a modification of the Efficient Battery Lifetime-Aware Power Saving Scheme described in ref [4] with the corresponding shortcomings of the scheme as at the first presentation in view. The existing Scheme successfully minimized the frequent transition of MS to sleep mode, at the expense of an increase in a longer sleeping interval (session) which has an intolerable delay bound. In the existing scheme a MS trades off power savings at the expense of longer sleep interval in order to minimize its power consumption, but this also causes an increase in the average feedback delay due to the longer sleep interval called response delay (time). The longer sleep time also resulted in an increase in both delay and power consumption due to the switching time taken for a Mobile device to revert (return)from sleep to active mode respectively.

Hence, the existing scheme delay bound has a severe effect on the battery life and the overall performance of a Mobile device as well. Therefore, the existing scheme control-sleep algorithm frequently restrains MS from receiving/transmitting, intending packets within an appropriate time or just in time.

Whenever packets arrive to the buffer queue; they are not receive/transmitted appropriately within their life time (Figure 5 and 6). Due to the impose control sleep bound, as such arriving packets are being delayed/Queued unnecessarily, resulting to packets congestion which further result to loss of packets due to the longer delay (sleep)bound called response delay and the short lived lifetime of packets on transits. The longer response delay is due to the larger sleep intervals used during adjustment of the sleep parameters. The longer response delay subsequently has effects on the quality of service (QoS) of MS and may result to user dissatisfaction of an interactive session. Thus, improving on these will be a significant edge in Power Savings. Hence, we developed a scheme that addressed these challenges highlighted aptly.

In this paper a new Scheme called a Delay Aware Power Saving Scheme (DAPSS) Base on traffic load is proposed. The Scheme proffers an adequate solution by minimizing the larger sleep window used called intervals by the existing scheme. The longer sleep time of the existing scheme is minimized, analytically and dynamically adjusting the sleep time interval of the sleep parameters used in PSC more appropriately or just in time as well as minimizing the response time and power savings efficiency as well. The propose DAPSS dynamically captures the traffics arrival pattern more appropriately by adaptively tuning the three operating parameters namely: idle time, initial (Tmin) and final (Tmax) sleep window. A modified sleep mode algorithm called a Delay Aware algorithm is employed minimizing the average response delay and improving performance of a mobile device accordingly.

More-so, Figure 6 above shows the actual variant buffer sizes of the intermediate devices as in the case of TCP where in between the MS and the BS there are devices called the intermediate devices that are of different buffers sizes. And from the Figure 6 above we can see that L2 and L3 have less buffer size compared to L1 and L4 which can take in more data. In cases where there is higher traffic Load arrival, congestion and buffer overrun is likely to be experience at L2 and L3 since both have reached their threshold and the MS at this time is in sleep mode and cannot process data until at the listening mode stage where a MON-TRF-IND message is received as Positive (Figure 4, T3) which subsequently results to packet loss as well as performance degradation of the network. The propose DAPSS scheme have minimize the longer delay accordingly, by appropriately adjusting the sleep time rightly or just in time and have also proposed an algorithm called a delay aware algorithm (algorithm 1) that have successfully minimized the excess longer delay of the existing scheme. We also validated our results analytically to justify the proposed DAPSS scheme.

Finally, we used, discrete event simulator for the simulation; the Simulation results evaluated showed that the proposed scheme outperforms the existing one in terms of both average response delay and power savings.

The main difference between the existing Scheme and the propose DAPSS Scheme is the modification of Tmin and Tmax sleep intervals (Equation 1 and 4), the appropriate choice of sleep parameters set used in the proposed DAPSS and the way of adjustment of the sleep parameters respectively. The Propose DAPSS Scheme is analytically modified as follows:

Firstly, we define the sleep parameters which are namely: sleep mode, listening mode, wake mode, idle mode, serving mode. Minimum sleep interval (Tmin), maximum sleep interval (Tmax).

Note that sleep window and sleep interval are same in this paper.

We call the duration of first sleep interval T1=Tmin, then the duration of kth sleep interval is given below:

3.1 A DELAY AWARE POWER SAVING SCHEME (DAPSS) BASED ON TRAFFIC LOAD

To address the problems highlighted above; a DAPSS Scheme Base on Traffic load is proposed. Unlike the existing scheme where the Tmin and Tmax is fixed in Equation 2 of the existing scheme in ref [4].

First, in the proposed DAPSS Scheme the sleep interval, Tmin and Tmax are modified as follows:

$$T_k = \begin{cases} \left(1+\frac{K}{\lambda}\right)2^{k-1}T_{\min}, & if \quad \left(1+\frac{K}{\lambda}\right)2^{k-1}T_{\min} < T_{\max} \\ \frac{T_{k-1}+T_{\max}}{2}, & Otherwise \end{cases} \quad (1)$$

Where Tk is the kth sleep window, T_{min} is the minimum sleep window, T_{max} is the maximum sleep window, k is a positive integer. T_{min} is determined by examining the inter arrival time of a downlink frame(s) in order to reduce the average response delay the downlink frames may had incurred while waiting for the MS to wake up. The minimum sleep window is T_{min} is give in Equation 4 below:

$$T_{average} = (1-\gamma)T_{\min} + \gamma T_d \quad (2)$$

Where Taverage is the weighted average inter arrival time in between the downlink frame from the BS for the MS.

$$\sigma_n = (1-\gamma)\sigma_n - 1 + \gamma\left|T_{average} - T_d\right| \quad (3)$$

σ_n is the weighted average variance of the inter arrival time of the downlink frame

$$T_{\min} = T_{average} - k\sigma_n \quad (4)$$

Finally, the Tsmin which is the minimum sleep interval is derived as seen in Equation 4 above. From the above Equations γ and k are constants given as $0 < \gamma < 1$ and k>1, Td is the time taken after which the downlink (DL) frames arrive at the base station (BS) for a mobile Station (MS) since it went into sleep mode last. Unlike the existing scheme which has a Tmin value that is fixed and the sleep window is also made to be constant to Tmax respectively.

The existing scheme Tmax sleep interval is maintained, therefore, when the sleep interval approaches Tmax as shown in Equation (2) of the existing Scheme in ref [4] the sleep interval becomes larger resulting to a longer sleep window, which is a key concern as the traffic arrival increases with an exponential increase sleep intervals seen in Figure 4 above. In the propose DAPSS Scheme we have modified the Tmin and Tmax sleep parameters as seen above in Equation 1 are used dynamically to adjust the sleep intervals aptly, as the sleep window increase the sleep parameters take a suitable average value so as to significantly minimize the existing schemes longer sleep intervals. More so, the sleep parameters are examined based on traffic load arrival pattern dynamically in order to predict the next actual arrival of the downlink frames which significantly improved performance as well as minimize the average number of listening intervals in the sleep window as well as power consumption. In addition, when the sleep window subsequently approaches Tmax, the sleep window is increased incrementally as an average of kth sleep window and Tmax as an average sleep interval as seen in Equation (1) to (4) respectively, thereby minimizing the response delay the downlink frame may had incurred in the process while waiting for the MS to switch to wake mode.

Unlike the existing scheme in ref [4], that take the full length of the of the Tmin and Tmax sleep intervals that is constantly increasing and subsequently resulst to a longer sleep window. The proposed DAPSS Tmin and Tmax sleep intervals are modified and given as seen in Equation 4 above for Tmin and the Tmax as $\frac{T_{k-1}+T_{\max}}{2}$ thus as the MS approaches Tmax where a longer sleep intervals is subsequently experienced in the existing scheme, the propose DAPSS considers an average suitable sleep intervals in order to quickly reverts to active mode in case of positive MOB-TRF-IND Messages as well as higher traffic load arrival from BS.

Let (n and P) denote the number of sleep intervals and a period of successive series of sleep intervals in sleep mode. (L) is the listening interval, while power consumption in sleep mode is (PS), power consumption in the listening interval is (PL). D represents the frame response delay that a MS requires to successfully deliver packets. We assume that MS follow a Poisson distribution with arrival rate as (λ). This implies that, the inter arrival time is distributed according to an exponential law with parameter (1/ λ). Let ek denote the event that there is at least one (1) packet arrival during the monitoring period.

$$\Pr\left[e_k = true\right] = -e^{-\lambda(T_k+L)} \quad (5)$$

The term Pr[n=k] represents the probability of success in the exact k-th iteration, which is also the probability of failure in iteration 1 to k -1 and success in the k-th. The number of sleep cycles is an independent random variable.

$$\Pr(n=1) = \Pr(e_1 = true) = 1 - e^{-\lambda(T_k+L)} \quad (6)$$

$$for\ n \geq 2, (n=k)$$

$$\Pr\left(e_1 = false; \ldots..; e_{k-1} = false : e_k = true\right)$$

$$= \prod_{i=1}^{k-1} \Pr\left(e_1 = false\right)\Pr\left(e_k = true\right)$$

$$= e^{-\lambda\sum_{i=1}^{k-1}[T_i+L]}\left(1 - e^{-\lambda(T_k+L)}\right) \quad (7)$$

$$= \sum_{k=1}^{\infty} e^{-\lambda\sum_{i=1}^{k-1}(T_k+L)} - \sum_{k-1}^{\infty} e^{-\lambda\sum_{i=1}^{k}(T_i+L)} \quad (8)$$

From Equation (6) and (8), the average numbers of sleep intervals are calculated. We present n as the number of sleep intervals and the probable value of n is represented by P[n]. Note that the value of n is between 0 to ∞. Thus, the possible value of n is given as follows:

$$\Pr[n] = \sum_{k=1}^{\infty} \Pr(n=k)$$

$$\Pr[n] = \sum_{k=1}^{\infty} e^{-\lambda\sum_{i=1}^{k-1}(T_i+L)} - \sum_{k=1}^{\infty} e^{-\lambda\sum_{i=1}^{k}(T_i+L)} \quad (9)$$

Individual sleep cycle has a length of Tk+L. Thus, the possible duration of a sequence of sleep cycles is dynamically calculated as follows:

$$E[D] = \sum_{k=1}^{\infty} \Pr(n = k)(k - th\ cycle\ duration) \quad (10)$$

$$\sum_{k=1}^{\infty} \Pr(n = k) \sum_{j=1}^{k} (T_j + L)$$

$$= (\sum_{k=1}^{\infty} e^{-\lambda \sum_{i=1}^{k-1}(T_i+L)} - \sum_{k=1}^{\infty} e^{-\lambda \sum_{i=1}^{k}(T_i+L)}) \sum_{j=1}^{k}(T_j+L) \quad (11)$$

$$k - th\ sleep\ cycle\ is\ P_k = \sum_{j=1}^{k} \left(T_j P_s + PL_L\right) \quad (12)$$

Assuming the frame response delay resulting to the outflow/overrun from the sequence of sleep cycles will arrive at any moment during the last sleep cycle with uniform probability. The length of k-th cycle is (Tk +L). The possible response Delay of frame is presented as follows:

$$E[D] = \sum_{k=1}^{\infty} \Pr(n = k) \sum_{j=1}^{k} (T_k + L) \frac{1}{2} \quad (13)$$

$$\Pr[D] = \frac{1}{2} \sum_{k=1}^{\infty} \left(e^{-\lambda (T_k + L)} - e^{-\lambda \sum_{i=1}^{k} (T_i + L)} \right) \sum_{j=1}^{k} (T_k + L)$$

Finally, the expected frame response delay is expressed as

$$E[D] = \frac{1}{2} \sum_{k=1}^{\infty} e^{-\lambda (T_i + L) \sum_{i=1}^{k} (T_k + L)} - \frac{1}{2} e^{-\lambda \sum_{i=1}^{\infty} (T_i + L) \sum_{j=1}^{k} (T_k + L)} \quad (14)$$

Hence, the average power consumption during the sleep, listening and idle mode are also presented as follows:

$$P_{sleep} = \sum_{s=1}^{S} P_s^{w_s + \sum_{i=1}^{L} \left(P_{Downlink-subframe} + \sum_{d=1}^{D} \alpha \beta_{d,1} + P_{\alpha\beta \to T\beta} \right)} \quad (15)$$

$$P_{awake} = \sum_{f=1}^{F} \left(P_{Downlink-subframe} + \sum_{d=1}^{D} P_{\alpha\beta_{df}} + P_{\alpha\beta \to T\beta} \right) \quad (16)$$

$$P_{iddle} = \sum_{i=1}^{I} P_{iddle} T_i + \sum_{p=1}^{P} \left(P_{Downlink-subframe} + \sum_{d=1}^{D} \alpha \beta_{dp} + P_{\alpha\beta \to T\beta} \right) \quad (17)$$

Therefore, the sum of the average power consumption A(P) of the proposed DAPSS Scheme is also expressed as follows:

$$A(P) = T_{iddle} + P_{iddle} \sum_{k=0}^{\infty} P_k \sum_{L=0}^{k} \left(T_{iddle} + T_{awake} + P_{sleep} \right) \quad (18)$$

Where $T_{iddle} + T_{awake} + P_{sleep}$ is the idle, sleep, wake mode (state).

3.2 Procedure of Parameters Adjustment of the DAPSS Scheme

The procedures of parameters adjustment of DAPSS scheme: The MS begins in a normal mode operation. And subsequently request for sleep if the mobile sleep request (MOB-SLP-REQ) is granted, the MS transits to sleep mode else a positive (+) mobile traffic indication message (MOB-TRF-IND) is sent to the MS from the BS and the MS wakes up and process data packets on the queue. This process is repeated until a Negative (-) Mobile Sleep Response(MOB-SLP-RSP)) is granted by the BS as seen in figure 2 above. Otherwise the MS reverts to normal mode operation and continue the process else end the process. To be more precise see Figure 7 below.

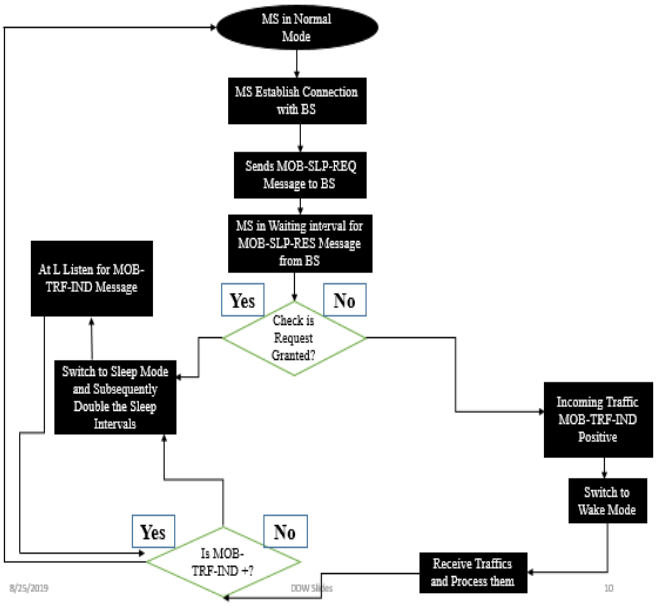


Figure: 7 Procedure of parameters adjustment of the DAPSS Scheme.

Algorithm 1: Delay Aware Power Saving (DAPSS) Scheme Based on Traffic load

Input: $T_{s\min\_\lambda} = k,\ T_{s\max=1024},\ k = 1$

1. $Calculate\ T_{s\min} = 2^k T_{s\min\_\lambda}$
2. $if(T_{s\min} < T_{s\max})$;
3. $k = k+1\ and\ repeat\ step\ 1$
4. **Else**
5. $calculate\ T_{s\max} = 2^k T_{s\min\_\lambda}$
6. $if(Power(T_{s\min}, T_{s\max,\lambda}) < \min\_Power\ Consumption \& (T_{s\min}, T_{s\max,\lambda}) < \min\_response\ Delay)$
7. $T_{s\min\_\lambda} = T_{s\min}, T_{s\max\_\lambda} = T_{s\max}$
8. $\min\_power\ Consumption = Power(T_{s\min}, T_{s\max,\lambda})$
9. $\min\_responseDelay = Delay(T_{s\min}, T_{s\max,\lambda})$
10. **Else**
11. $Input\ k$
12. $do\ step\ 1$
13. **End**

4. Performance Evaluation

This section presents the performance evaluation of the propose DAPSS against that of the EBLAPS using a Network simulator 3 (NS3). The evaluation is based on the average energy savings and average response delay. The simulation topology consists of a base station (BS) with Mobile Station connected around it as seen in Figure 8 below. The simulation parameters are also presented in table 1 below:

Table 1: Simulation Parameters

Pcsm	power consumption in sleep mode	1
Pclm	power consumption in listening mode	30
L	Listening mode	1
Tmin	Minimum Sleep Intervals	1
Tmax	Maximum Sleep Interval	1024
Ar	Arrival Rates	0.001; 0.4
K	Constant	

Id	Idle Time	0.5-4

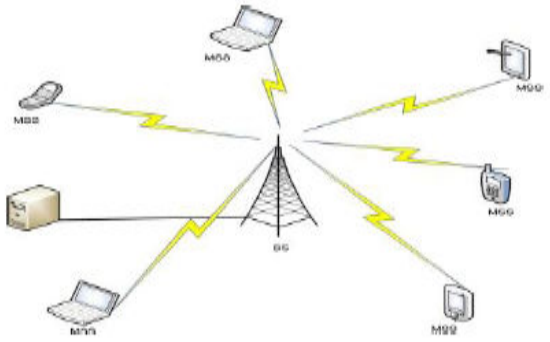


Figure 8. Simulation Topology

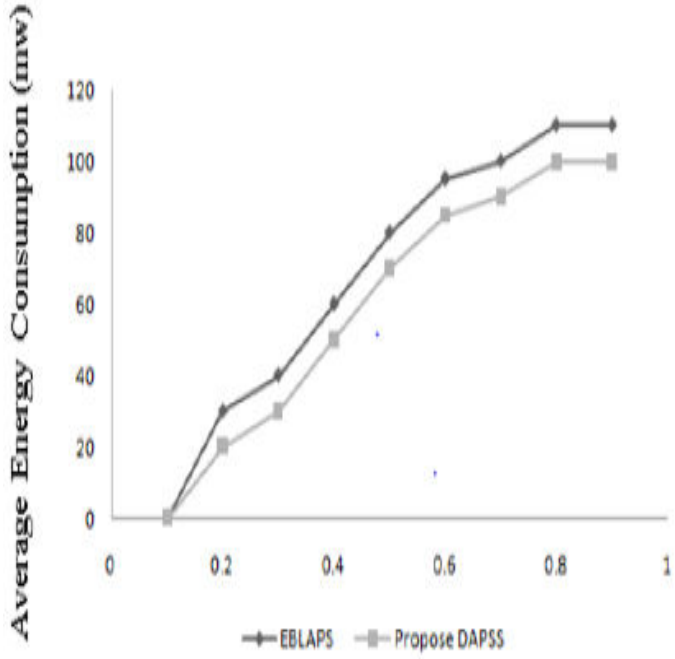


Figure 2: shows the Average Energy Consumption against the Mean Arrival Rate of packets.

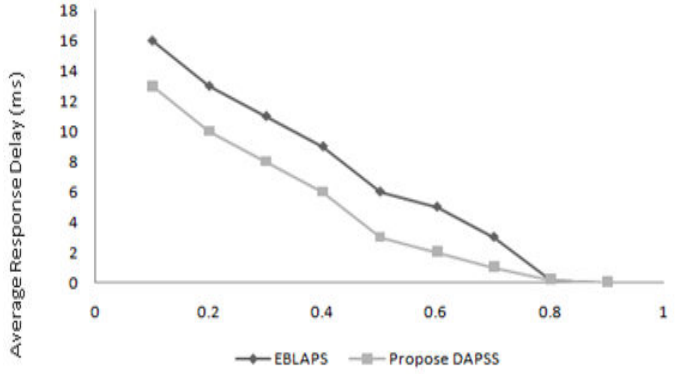


Figure 3. *Mean Packet Arrival Rate* (λ)

The above Figure 3: shows the Average Response Delay against the Mean Packet Arrival Rate.

5. Conclusion

We have proposed a new Scheme called a Delay Aware Power Saving Scheme (DAPSS) in the downlink operation of the IEEE 802.6e WiMAX Networks for MBNS based on Traffic Load. The DAPSS Scheme minimized the excessive longer sleep interval of the existing scheme by appropriately adjusting the sleep parameters, with the aid of a delay aware algorithm and resolve the response delay of MS. In addition, we analytically modified the Tmin and Tmax sleep parameters and conducted a simulation experiment using a discrete event Network Simulator. Our simulation results proved that the proposed DAPSS Scheme has a superior performance significantly in terms of both the average response delay and power savings.

Acknowledgements

The Authors wish to thank anonymous reviewers who took out their time and constructively made comments that have improved on this manuscript, and also thank the Usmanu Danfodiyo University for their supports as at the time of this research studies.

References

[1] IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE 802.16 Working Group and others", IEEE Std, 802.16-2004.

[2] IEEE Std. 802.16e-2005, IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, and IEEE Std. 802.16-2004/Cor1-2005, Corrigendum 1, December 2005.

[3] H. Eunju, J.K. Kyung, J.S. Jung and D. C. Bong, "Power Saving Mechanism with Periodic Traffic Indications: A New Sleep Mode Scheme in the IEEE 802.16e". Proceedings of the Third Korea-Netherlands Conference on Queueing Theoryand its Applications to Telecommunication Systems,pp 319-334, 2007.

[4] I. Saidu, S. Shamala, J. Azmi, A. Zuriati and Zukarnain, "An efficient battery lifetime aware power saving (EBLAPS) mechanism in IEEE 802.16e networks", Wireless Pers. Commun., No 80, Vol 1, pp. 29-24, 2015.

[5] L.-D. Chou, D. C. Li, and W.-Y. Hong, "Improvin g energy-efficient communications with a battery lifetime-aware mechanism in IEEE 802.16e wireless networks", in Concurrency and computation: Practice and experience, vol. 25, pp. 94-111, 2013.

[6] K-T. Feng, W-C. Su, and C-Y. Chen, "Comprehensive Performance Analysis and Sleep Window Determination for IEEE 802.16 Broadband Wireless Networks", IEEE Transaction on Mobile Computing, pp. 1536-1233, 2015.

[7] O. J. Vatsa, M. Raj, R. K. Kumar, D. Panigrahy, and D. Das, "Adaptive power saving algorithm for mobile subscriber station in 802.16e", in the 2nd IEEE International conference on communication systems software and middleware, (COMSWARE) pp. 1–7, 2007.

[8] M.-G. Kim, M. Kang, and J. Y. Choi, "Remaining energy-aware power management mechanism in the 802.16e MAC", In the 5th IEEE-consumer communications and networking conference

(CCNC2008) pp. 222–235, 2008.
[9] J. Xue, Z.Yuan, Q-Y. Zhang, "Traffic Load-Aware Power-saving Mechanism for IEEE 802.16e Sleep Mode", College of Computer and Communication Lanzhou University of Technology Lanzhou, China, 2008.
[10] C. C. Yang, C. H. Fang, and J. R. Lin, "Adaptive Power Saving Strategy Based on Traffic Load in the IEEE 802.16e Network", in Proc. International Conference on Information and Communication Technologies (ICICT 2010), pp. 26-28, Tokyo, Japan, 2010.
[11] C. C. Yang, Y. T. Mai, J. Y. Chen, Y.S. Shen, and Y. C. Kuo, "LBPS: Load-Based Power Saving in the IEEE 802.16e Network". Computers and Electrical Engineering, PP. 891–905, 2012.
[12] S. Mehta, N. Seth, N. Snigdha, "A Novel Approach for Minimizing the Delay and Load in Wireless Network", Int. Journal of Engineering Research and Applications, www.ijera.com ISSN: 2248-9622, Vol. 3, Issue 6, Nov-Dec. pp.1344-1350, 2013.
[13] S-R. Yang, "Dynamic Power saving mechanism for3G UMTS Systems", Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.in Mobile Networks and Applications, February 2007.
[14] J-R. Lee and D-H. Cho, "Performance Evaluation of Energy-Saving Mechanism Based on Probabilistic Sleep Interval Decision Algorithm in IEEE 802.16e", IEEE Transactions on Vehicular Technology, Vol. 56, No. 4, July 2007.
[15] J. C-R. Bennett, and H. Zhang, "Worst Case Fair Waited Fair Queuing", School of computer Science, Carnegie Mellon University,1997.
[16] M. Shreedhar and G. Varghese, "Efficient fair Queuing using Deficit Round Robin" , IEEE/ACM Transactions on Networking, Vol. 4, No. 3, 1996.
[17] J. Feng, and R. Xia, "Load-Aware Network Entry in WiMAX Mesh Mode", International Conference on Computer Science and Software Engineering, 2008.
[18] J. Jang, K. Han and S. Choi, "Adaptive Power Saving Strategies for IEEE 802.16e Mobile Broadband Wireless Access", ‖ IEEE Communications, APCC '06. Asia-Pacific Conference, 2006.
[19] W.-H. Liao, and W.-M. Yen,"Power-Saving scheduling with a QoS guarantee in a mobile WiMAX system", ‖ Journal of Network and Computer Applications, vol. 32, no. 6, pp. 1144–1152, 2009.
[20] M-G Kim, J.Y. Choi and M Kang, "Enhanced Power-Saving Mechanism to Maximize Operational Efficiency in IEEE 802.16e Systems" , IEEE Transactions on Wireless Communications, Vol. 8, pp no. 9, September 2009.
[21] Y. Xiao, "Energy saving mechanism in the IEEE 802.16e wireless MAN", IEEE Communications Letters, vol. 9, pp. 595–597, 2005.
[22] J. Xiao, S. Zou, B. Ren, and S. Cheng, "an enhanced energy saving mechanism in IEEE 802.16e", in IEEE global telecommunications conference, 2006. (GLOBECOM'06) pp. 1-5.
[23] Z. Shengqing, M. Xiaoyu and W. Lujian,"A delay-aware auto sleeps mode operation for power saving WiMAX", in Proceedings of 16th international conference on computer communications and networks (ICCCN 2007), pp. 997-1001 2007.
[24] K. Min-Gon, K. Minho, and C. Jungyul, "Remaining Energy-Aware Power Management Mechanism in the 802.16e MAC", IEEE Communications Society subject matter experts for publication in the IEEE (CCNC) Proceedings of the winter simulation conference, pp. 4244-1457, 2008.
[25] G. Wong, Q. Zhang, H. Danny and K. Tsang, "Switching Cost Minimization in the IEEE 802.16e Mobile WiMAX Sleep Mode Operation" , Wireless communications and mobile computing, pp. 1576-1588. 2009.
[26] X. Jianbin, Y. Zhanting, and Z. QiuYu, "Traffic Load-Aware Power-saving Mechanism for IEEE 802.16e Sleep Mode", in College of Computer and Communication Lanzhou University of Technology Lanzhou, China., pp. 735-745, 2008.
[27] C. Jenhui, T. Woei-Hwa, and J. Der "A Downlink and Uplink Alignment Scheme for Power Saving in IEEE 802.16 Protocol", Scientific World Journal, Volume Article ID 217973, pp. 11, doi.org/10.1155/2014/217973, 2014.

An Approach for Journal Summarization using Clustering Based Microsummary Generation

Hammed Adeleye MOJEED
Department of Computer Science
Univerisity Of Ilorin
Ilorin, Nigeria
mojeed.ha@unilorin.edu.ng

Shakirat Aderonke SALIHU
Department of Computer Science
Univerisity Of Ilorin
Ilorin, Nigeria
salihusa1980@gmail.com

Ummu SANOH
Department of Computer Science line
Univerisity Of Ilorin
Ilorin, Nigeria
sanohumu67@gmail.com

***Abstract*— The major challenge in summarization lies in distinguishing the more informative parts of a document from the less ones. Previous works proposed techniques focusing on essay, news articles and web documents could not work effectively for summarizing journal articles due to its size and structure. This work presents a new approach for summarization of journal articles based on micro summary generation to aid article review process. The approach focused on segmenting the journal articles into sections based on its peculiar structure and applies a information retrieval sentence similarity based clustering approach using K-means to produce micro summaries for each section. The generated summaries are combined systematically to generate the final summary. Results from evaluation of the proposed approach on seven selected journals proved promising with good precision value close to 0.8, recall of 1.00 and F-measure of 0.9. The proposed approach would aid the review process of researchers.**

Keywords*— Text Summarization, Information Retrieval, Sentence Similarity, Clustering.

I. INTRODUCTION

Summarization is the process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document. According to Kaynar, Isik, and Gormez [1] the main idea of summarization is to find a representative subset of the data, which contains the information of the entire set. The explosive growth of the world-wide web has dramatically increased the speed and the scale of information dissemination and this has led to awakening interest in summarization. In the field of automatic text summarization it is customary to differentiate between extraction based summaries where the summary is composed of more or less edited fragments from the source text and abstraction based summaries where the source text is transcribed into some formal representation and from this its regenerated in a shorter more concise form [2]. Abstraction involves paraphrasing sections of the source document. In general, abstraction can condense a text more strongly than extraction, but the programs that can do this are harder to develop as they requires use of natural language generation technology, which itself is a growing field [3]. The Extraction based summarization works by selecting a subset of existing words, phrases, or sentences in the original text to form summary. In this summarization task, the automatic system extract object from the entire collection without modifying the objects themselves.

Some applications of extraction include key phrase extraction, where the goal is to select individual words or phrases to "tag" a document, and document summarization, where the goal is to select sentences (without modifying them) to create a short paragraph summary. Similarly, in image collection summarization, the system extracts images from the collection without modifying the images themselves. The major challenge in summarization lies in distinguishing the more informative parts of a document from the less ones. Though there have been instances of research describing the automatic creation of abstracts, most works presented in the literature focused on direct extraction of sentences to address the problem of text summarization [4].

Jayashree, Srikanta Murthy and Sunny [5] created a keyword extraction based summarization of categorized Kannada text documents which gives the result of a manual evaluation of the summarizer with three different human summaries across various categories limiting it to manual=10.The number of sentences common between the two summaries gives the relevance score; a data set consisting of 87 documents across different categories is input into the summarizer. This human reference summary was compared to the summary generated by the machine. The computer generated scores gave a competitive results with the human generated results in terms of relevance. Their work is focused on essay articles rather than journals.

Gupta et al. [6] presented a new statistical approach to automatic summarization based on the Kernel of the source text called KernelSum (KERNEL SUMMarizer). Using simple statistical measures, Kernel is identified as the most significant passage of the source text. It serves as the guideline to choose the other sentences for summary. The procedure proposed is based on summarizing online data in html form. Scoring algorithm estimate the score of each sentence by using the TF-ISF (Term Frequency - Inverse Sentence Frequency) then ranking algorithm counts rank of every sentence according to the scores, location, length and heading sentence etc. Summarizing part picked the sentences from the ranked list and concatenated to produce the expected brief of the input document. Final extract have been evaluated under the light of Kernel preservation and textuality and found 90% of the extracts have been judged to totally or partially preserving the gist, textuality was also highly graded: 85% of them were totally or partially coherent and cohesive . Agrawal and Gupta [7] proposed an extraction-based summarization technique using k-means

clustering algorithm, which is an unsupervised technique. This paper describes an algorithm that incorporates k-means clustering, term-frequency inverse-document-frequency and tokenization to perform extraction based text summarization. Each sentence is given an important score. The scores can be used to order sentences and pick most important sentences. The probability of a sentence to be present in the summary is proportional to its score. Each sentence is represented by a set of features. To evaluate the approach, some six (6) text samples were simulated and were compared with other existing summarizer which uses an extraction based technique. The average result obtained were: 44.11% of size of text in summary by their approach was retained while 19.1% of size of text in summary was retained by another summarizer (autosummarizer.com)

Aparıcio et al. [8] proposed generic text summarization algorithms to films and documentaries, using extracts from news articles produced by reference models of extractive summarization. The authors used three datasets: (i) news articles, (ii) film scripts and subtitles, and (iii) documentary subtitles. Standard ROUGE metrics are used for comparing generated summaries against news abstracts, plot summaries, and synopses. The authors showed that the best performing algorithms are LSA, for news articles and documentaries, and LexRank and Support Sets, for films. Despite the different nature of films and documentaries, their relative behavior is in accordance with that obtained for news articles. Overall, LSA performed consistently better for news articles and documentaries, one possible reason for these results is that LSA tries to capture the relation between words in sentences. Yadav and Shah [9] proposed a text summarization with fuzzy approach which summarizes reports of a particular event automatically and comprehensively. To speed up summarization, a pre summarization approach is introduced to condense each report to a sub-summary, which can reduce the scale of subsequent processing. As an event should be told in chronological order, a timeline is introduced to organize and aggregate event-relevant sub-summaries. With each day's subsummaries, extraction algorithm is used to cluster them into topics and generate a meaningful label for each topic. Finally, a selection criterion is introduced to select relevant and novel sentences for each topic

Previous works have focused on essay and news articles with techniques which cannot work effectively for summarizing journal articles due to its large size and structure. Efforts towards designing a summarization technique specifically for journal is very scarce in the literature. Therefore this work focuses designing a model for journal summarization based on micro-summaries generation using clustering based summarization technique.

II. METHODOLOGY

This section presents the proposed approach for Journal summarization based on micro-summary generation. It involves the stages or steps in the actualization of the system. The steps begin with preprocessing the journal, then splitting it into various subsections, each of which is viewed as a single document on its own. The idea is to take a single document which is the journal, split it into subsection to derive multi-documents which the summarization process shall be performed on. After the document has been split into subsections, Similarity measures using information retrieval technique is carried out to score each sentence in each subsection. After that, clustering based summarization algorithm with K-Means is applied to obtain the micro-summary after candidate selection. The micro-summaries are then combined to generate the actual summary. Fig. 1 presents the flowchart for the proposed approach.

The first step is to prepare the journal input for summarization through preprocessing. In preprocessing, the input text is broken down into sentences followed by terms/word extraction. As the preprocessing stage commences, the input/text document undergoes several processes namely stop word removals and word net stemming. Removal of stop word is executed based on stop word list which constitutes of a list of terms to be excluded for consideration in generating summary. These words do not contribute in understanding the main ideas present in the input text. Examples of stop words are “him”, “are”, “the”, and “it”. Word stemming which is done based on word net database. The purpose of having word stemming process is to have only one root word which could be written more than once in different format, that is, different format of a word will result in different meaning. However, the root word would be the same for example, the words “gone”, going and goes all have the same root word which is go. All in all, the word stemming process deals with the prefixes and suffixes of each word. This process ultimately contributes to accuracy of the similarity measures and consequently the micro-summary generated.

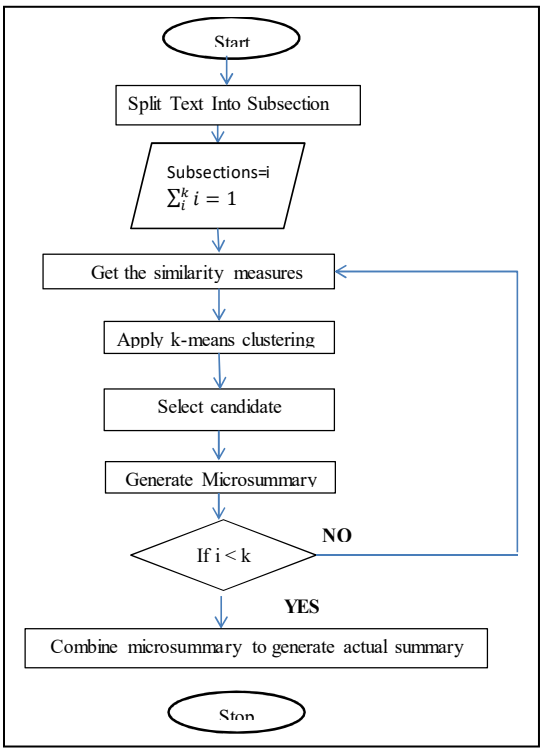


Fig. 1: Flow Chart of the Proposed Journal Summarization Approach

A. Text Splitting

After preprocessing the text is split into subsection based on the general structure of journals. All journals are categorized into section using section number. This feature is leverage on to partition the whole journal text into small

sections that can be handled efficiently. The text is scanned from the beginning to the end and anytime the carriage return followed by a number is encountered the text read so far is separated. The process continues until the whole text is partitioned.

B. *Information retrieval based Sentence Similarity Measure*

Information retrieval, as the name implies, concerns retrieving of relevant information from databases. It involves techniques that can be employed to generate similarities between document in a corpus. In this work information retrieval is used to calculate the similarities between sentences in a subsection of a journal. The IR methods Vector Space Model and cosine similarities are used.

Vector Space Model (VSM): *Vector space model* or term *vector model* is an algebraic *model* for representing text documents (and any objects, in general) as *vectors* of identifiers, such as, for example, index terms. It is used in information filtering, information retrieval, indexing and relevancy rankings. The VSM is generated for each sentence in a subsection in this work using TF-IDF. TF-IDF contains two components: term frequency (TF) and inverse document frequency (IDF). TF is the frequency with which a term occurs in a sentence. The rationale is that the more frequently a term occurs, the more likely this term describes a document's content or user's information need. IDF reflects the importance of the term by computing the inverse frequency of sentence containing the term within the entire subsection to be summarized. The basic assumption is that a term should be given a higher weight if few other sentences also contain that term, because rare terms will likely be more representative of a document's content or user's interests. While TF-IDF was originally developed for classic search, TF-IDF is also one of the most popular term-weighting schemes for user modeling and recommender systems. For instance, TF-IDF is used by 83% of surveyed text-based research-paper recommender systems [10]..

For example, when a 30-word sentence in a section of journal contains the term "software" 5 times, the TF for the word 'cat' is

TF(software) = 5/30 = 0.16

The IDF (inverse document frequency) of a word is the measure of how significant that term is in the whole section. For example in our case, Let's assume there are 15 sentences that contain the term "software" out of 45 sentences in the section, then the IDF (i.e. log {DF}) is given by logarithm of the total number of sentences (45) divided by the number of sentences containing the term "software" (15).

IDF (software) = log (45/15) = 0.47

Now the TF-IDF of the term software is given as:

TF*IDF (software) = 0.16 * 1.47 = 0.075

The TF-IDF of each term in the sentence is calculated and this forms the vector representation of the sentence.

Cosine Similarity: Cosine similarity is a widely implemented metric in information retrieval and related studies. This metric models a text document as a vector of terms. Similarly, the similarity between two sentences can be derived by calculating cosine value between term vectors. One advantage of cosine similarity is its low-complexity and especially for sparse vectors: only the non-zero dimensions need to be considered.

The cosine of two non-zero vectors can be derived by using the Euclidean dot product formula:

$$A.B = \|A\|\|B\|\cos\theta \tag{1}$$

Given two vectors of attributes, A and B, the cosine similarity, cos(θ), is represented using a dot product and magnitude as:

$$Similarity = \cos(\theta) = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2}$$

Where A_i and B_i are components of vector A and B respectively.

After this process every sentence in a subsection are given score according to their relevance and importance in the section based on their total similarity scores with reference to cosine similarity. The numerical scores are used to cluster the sentences into two groups based on the scores: relevant and non-relevant.

C. *Clustering and Micro-summary Generation*

Clustering group data object based only on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better and more distinct the clustering. The clustering algorithm used in this work is k-means algorithm. According to the basic K-mean clustering algorithm, clusters are fully dependent on the selection of the initial clusters centroids. K data elements are selected as initial centers; then distances of all data elements are calculated by Euclidean distance formula. In this work K is selected to be 2 as relevant and non-relevant groups. Data elements having less distance to centroids are moved to the appropriate cluster. The process is continued until no more changes occur in clusters. First it needs to define initial clusters which makes subsets (or groups) of nearest points (from centroid) inside the data set and these subsets (or groups) called clusters technique [11]. Secondly, it finds means value for each cluster and define new centroid to allocate data points to this new centroid and this iterative process will continue until centroid does not change [12]. K-means is used in this work to group the sentence into clusters so that related sentences are grouped together. Algorithm 1 presents the procedures for K-Means.

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values

Step 2: Initialize the first K clusters
- Take first k instances or
- Take a Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset.

Step 4: K-means assigns each record in the dataset to only one of the initial clusters.
- Each record is assigned to the nearest cluster using a measure of distance (e.g. Euclidean distance, Minkowski distance and so on).

Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-

Algorithm 1: Procedure for K-Means Algorithm [13]

- Next is to select candidate sentences for the micro-summary generation. The goal of the selection procedure is to identify a set of sentences that contain important information. The task of content selection is to identify which sentences in the source section are worth taking into the micro-summary. In this work, we used centrality-based algorithms where sentences are ranked based on their similarity with other sentences in the same section as processed by K-means. The intuition behind this approach is that the core information will be repeated or referred to in many sentences, thus the sentences that are similar to many other sentences are likely to contain this information and should be included in the extract. The clusters with higher centroid weight from the clustering output should contain sentences that are more relevant in the section than the one with lower centroid value. Consequently all sentences in the cluster with higher centroid value are selected.

A micro-summary is obtained after candidate sentences have been selected. Each sentence from a subsection selected to form a micro-summary have higher scores in comparison to other sentences in the same subsection. This process is iterated for all sections or partitions generated from the source journal and all sections have their own micro-summary.

D. Final Summary Generation

The micro-summaries generated are combined together to form a reduced document. The summarization process is applied again to remove redundancies and generates the final summary. To cater for the coherence of the final summary, sentences a placed according to the order in which they appear in the original journal text. This way the summary flows normally like the original journal.

E. Performance Measures Used

Since summarization is an information retrieval process, standard information retrieval performance measures precision, recall and F-measure are used to evaluate the performance of the proposed approach.

Precision: precision is the fraction of relevant sentences among the retrieved sentences in the summary. Precision related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results. It is a refinement in a measurement, calculation, or specification, especially as represented by the number of digits given. Precision it is measured in this work as:

$$\Pr ecision = \frac{\mathrm{Re}levantSentences \bigcap \mathrm{Re}trievedSentences}{\mathrm{Re}trievedSentences} \quad (3)$$

Precision is also called Positive Predictive Value.

Recall: Recall is the fraction of relevant sentences that have been retrieved over total relevant sentences in the summary. It is also known as sensitivity

$$\mathrm{Re}call = \frac{\mathrm{Re}levantSentences \bigcap \mathrm{Re}trievedSentences}{\mathrm{Re}levantSentences} \quad (4)$$

Recall can be viewed as a probability that relevant sentences are retrieved by the summary.

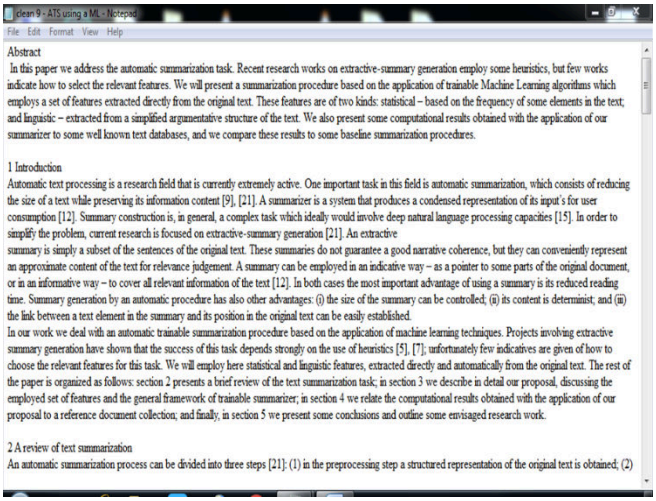
F-Measure: The F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test. This measure is approximately the average of the two when they are close and is more generally the harmonic mean, which, for the case of two numbers, coincides with the square of the geometric mean divided by the arithmetic mean.

$$F - measure = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (5)$$

III. RESULTS AND DICUSSION

The proposed approach was implemented in Python using the libraries NLTK, Scikit-Learn, NumPy, SciPy and Sympy. Experiments on selected journals is performed on an intel core i5 processors of 2.26 GHz speed each and a RAM of 4GB size PC running windows 10. Fig. 2 and 3 shows the screenshots of the original journal in text format and the resulting summary generated from the proposed approach.

.


clean 9 - ATS using a ML - Notepad

File Edit Format View Help

Abstract

In this paper we address the automatic summarization task. Recent research works on extractive-summary generation employ some heuristics, but few works indicate how to select the relevant features. We will present a summarization procedure based on the application of trainable Machine Learning algorithms which employs a set of features extracted directly from the original text. These features are of two kinds: statistical – based on the frequency of some elements in the text; and linguistic – extracted from a simplified argumentative structure of the text. We also present some computational results obtained with the application of our summarizer to some well known text databases, and we compare these results to some baseline summarization procedures.

1 Introduction

Automatic text processing is a research field that is currently extremely active. One important task in this field is automatic summarization, which consists of reducing the size of a text while preserving its information content [9], [21]. A summarizer is a system that produces a condensed representation of its input's for user consumption [12]. Summary construction is, in general, a complex task which ideally would involve deep natural language processing capacities [15]. In order to simplify the problem, current research is focused on extractive-summary generation [21]. An extractive summary is simply a subset of the sentences of the original text. These summaries do not guarantee a good narrative coherence, but they can conveniently represent an approximate content of the text for relevance judgement. A summary can be employed in an indicative way – as a pointer to some parts of the original document, or in an informative way – to cover all relevant information of the text [12]. In both cases the most important advantage of using a summary is its reduced reading time. Summary generation by an automatic procedure has also other advantages: (i) the size of the summary can be controlled; (ii) its content is determinist; and (iii) the link between a text element in the summary and its position in the original text can be easily established.

In our work we deal with an automatic trainable summarization procedure based on the application of machine learning techniques. Projects involving extractive summary generation have shown that the success of this task depends strongly on the use of heuristics [5], [7]; unfortunately few indicatives are given of how to choose the relevant features for this task. We will employ here statistical and linguistic features, extracted directly and automatically from the original text. The rest of the paper is organized as follows: section 2 presents a brief review of the text summarization task; in section 3 we describe in detail our proposal, discussing the employed set of features and the general framework of trainable summarizer; in section 4 we relate the computational results obtained with the application of our proposal to a reference document collection; and finally, in section 5 we present some conclusions and outline some envisaged research work.

2 A review of text summarization

An automatic summarization process can be divided into three steps [21]: (1) in the preprocessing step a structured representation of the original text is obtained; (2)

Fig. 2: Input Journal Text

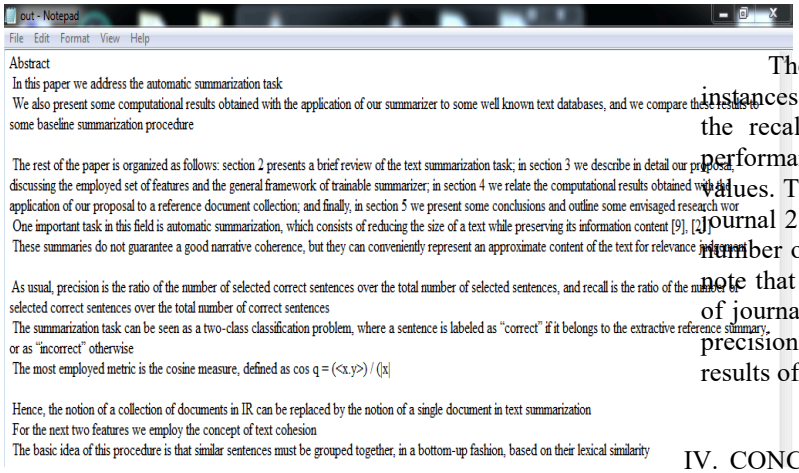


Fig. 3: Summary Output

To evaluate the performance of the applied summarization method, the summary obtained is compared with manual extracts which are provided by three different human summarizers on seven different journals selected from different fields in computing based on majority count. Table 1 shows the comparison between the number of sentences obtained by human summarizers using majority count and the number of sentences obtained by the automatic summarizer**.**

Table 1: Number of Sentences Obtained by Human and Automatic Summarizer

TEXT	No. of Sentences in original text	No. of sentences Human Summarizer	No. of sentences Automatic Summarizer
JN1	120	38	35
JN2	97	33	24
JN3	105	19	22
JN4	126	37	28
JN5	165	38	21

It can be seen from the results that the summarizer was able to produce good number of sentences in the summary generated with high compression ratio as expected for effective summary. It was able to generate lower number of relevant sentences in 5 out of 7 considered journals giving 71.4% performance rate. As a test of efficacy the precision, recall and F-measure are calculated based on the seven considered journals. Table 2 presents these results. From the results, it is obvious that the precision of the proposed approach is good with highest value reaching 0.825. This shows that the proposed approach was able to produce relevant sentences from the journal. The performance in journals 2 and 3 is low with 0.58 and 0.59 precision values. This may be due to the fact that the journals are relatively shorter in size.

The performance in terms of recall is high in all instances reaching up to 1.00 values. It can be observed that the recall compensate for the shortfall in the precision performance. This is clearly manifested in the f-measure values. The lowest F-measure value of 0.638 is recorded for journal 2, this is due to its short length as manifested in the number of sentences it comprised of. It is also important to note that the proposed approach will improve the relevance of journal summaries if applied on long journals with good precision, recall and F- measure values as shown in the results of journal 5.

IV. CONCLUSION

This work presents a new extractive journal summarization technique based on micro-summary generation. It proposed splitting of the journal text into sections and applied IR sentence similarity based clustering technique to generate microsummaries for each section. The generated summaries are then combined and resummarized using the same approach to produce the final summary. The approach was evaluated on carefully selected journals from different fields on computing using standard IR measures; precision, recall and F-measure. It was observed from the results that the proposed algorithm ranks some sentences that are relevant higher than human extracts. With performance measures, the proposed approach recorded good precision values rising close to 0.8, recall rising up to 1.00 and F-measure of 0.9. This shows the ability and effectiveness of the proposed approach in handling large sized text which are difficult to handle by the present summarization approaches.

In the future, we plan to compare the performance of the proposed algorithm empirically with other state of the art .

Table 2: Performance Measures Results

S/N	PRECISION	RECALL	F-MEASURE
JN1	0.825	1.00	0.904
JN2	0.580	0.711	0.638
JN3	0.591	0.867	0.702
JN4	0.714	0.869	0.783
JN5	0.857	1.00	0.922
JN6	0.643	0.90	0.750
JN7	0.680	0.894	0.772

REFERENCES

[1] O. Kaynar, Y.E Isik, and Y. Gormez. Graph based automatic document summarization with different similarity methods. *In Proceedings of IEEE conference on Signal Processing and Communications Applications* . 2017 doi: 10.1109/SIU.2017.7960613

[2] M. Hassel. Exploitation of Named Entities in Automatic Text Summarization for Swedish. In *Proceedings of NODALIDA'03 - 14th Nordic Conference on Computational Linguistics*, Reykjavik, Iceland. 2004.

[3] A. Zadbuke, S. Pimenta, D. Padwal and V. Wangikar. Automatic Summarization of News Articles using TextRank. In *International Conference on Electronics & Computing Technologies*. 6, 2016, p.3

[4] Zhang, J., Zhou, Y. and Zong, C., 2016. Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(10), pp.1842-1853.

[5] R. Jayashree, S. K. Murthy and K.. Sunny. Keyword extraction based summarization of categorized Kannada text documents. *International Journal on Soft Computing*, *2*(4), 2011, p. 81.

[6] V. Gupta, P. Chauhan, S. Garg, A. Borude and S. Krishnan. An statistical tool for multi-document summarization. *International Journal of Scientific and Research Publications*, *2*(5) , 2012.

[7] A. Agrawal and U. Gupta. Extraction based approach for text summarization using k-means clustering. *International Journal of Scientific and Research Publications*, *4*(11) , 2014.

[8] M. Aparício, P. Figueiredo, F. Raposo, D. M. de Matos, R. Ribeiro and L. Marujo. Summarization of films and documentaries based on subtitles and scripts. *Pattern Recognition Letters*, *73*, 2016, pp. 7-12.

[9] S., Yadav and D. Shah. News Summarization using Text Mining. *International Research Journal of Engineering and Technology (IRJET)* 05 (11), 2018, pp.202 – 206

[10] J. Beel, B. Gipp, S. Langer and C. Breitinger. Paper Recommender Systems: A Literature Survey. *International Journal on Digital Libraries*, *17*(4), 2016, pp. 305-338.

[11] H. Jiawei, M. Kamber and J. Pei. Data Mining: Concepts and Techniques, San Francisco California, Morgan Kaufmann Publishers, 2012 .

[12] T. S. Madhulatha. *An overview on clustering methods*. arXiv preprint 2012, arXiv:1205.1117

[13] R. Marapareddy and S. W. Saripalli. Runway Detection Using K-Means Clustering Method Using UAVSAR Data. In *Proceedings of the International Conference on Artificial Intelligence and Applications (AIAPP), Geneva, Switzardland*, 2017 pp. 25-2

A PC-Based Multifunctional Virtual Oscilloscope

Abdussamad U Jibia
Department of Mechatronics Engineering
Bayero University Kano
Kano, Nigeria
ajumar@buk.edu.ng

Ndubuisi D. Robinson
Department of Electrical Engineering
Bayero University Kano
Kano, Nigeria
ndubisirobinson@gmail.co

abstract— A cost effective and flexible DAQ (Data AQuisition) card and PC sound card based multifunctional virtual oscilloscope with time domain analysis, frequency domain analysis, joint time-frequency domain analysis and data analysis function is presented. The NI Labview based virtual instrument was designed and developed with the capability to acquire , display and analyze signals in time- frequency domain using the short-time Fourier transform analysis (STFT) and wavelet transform analysis tool kits in Labview and also carry out data analysis functions such as curve fitting (regression analysis) and statistics measurement (mean, median, and mode). The PC sound card used as an alternate signal acquisition in the proposed virtual oscilloscope makes the system flexible for school laboratory and the time-frequency analysis functions, data analysis functions of the virtual instrument makes the system a compact multifunctional virtual oscilloscope for signal (stationary and nonstationary signals) and data analysis in the laboratory.

Keywords— Labview, multifunctional, oscilloscope, time-frequency, virtual.

I. INTRODUCTION

DISPLAY and analysis of signals commonly done with traditional oscilloscope is only easy when only either time or frequency information of the signal is of interest. For joint time and frequency analysis, the traditional table-top oscilloscope will not be helpful. In addition to this, there are other disadvantages. For example, the traditional oscilloscope is standalone. Since it is not computer-based, it has limited connectivity to networks, peripherals and applications. It is also expensive in terms of development and maintenance. There is also lack of flexibility since it is hardware-based.

The rapid development of integrated circuits and computer technology gives birth to a new kind of instrument; virtual instrument which is a computer-based instrument system. This instrument has more functions, higher processing speed, wider bandwidth, friendlier interface, small size and better expandability [1]. These instruments are based on the PC platform which not only makes it convenient to use software and hardware resources of the PC but also makes data analysis, data representation and report presentation flexible and convenient.

Digital oscilloscopes have been implemented over the years with added features of pc connectivity such that signals can be stored in the PC for later analysis [2]. However, this commercially available oscilloscope with such advanced features are expensive are often underutilized in the low power and low frequency requirement of school laboratories hence this has paved way for a Virtual oscilloscope as an alternative system for instrumentation and measurement.

Chanden et al [2] developed a low cost pc-based virtual oscilloscope providing a low cost yet simple, effective integrated system with multichannel data acquisition, display and printing operation for undergraduate laboratories. The virtual oscilloscope designed uses a conventional pc parallel port interfacing circuitry and a graphical user interface developed using program languages such as Turbo C++ and Visual BASIC. However, the system has limited signal processing modules and expanding the functionality requires some level of expertise on Turbo C++ and Visual BASIC. Wie and Fang [3] developed a virtual oscilloscope on the platform of virtual instrument labview using the PCI 6024 data acquisition card. The virtual oscilloscope has signal processing functions limited to waveform display, data record and playback function, voltage measurement, time measurement, spectrum analysis function and in addition the oscilloscope could realize measurement over long distance through the data transfer module. Its functions are limited to analysis of stationary signals in time domain and frequency domain.

Ping and wie [4] designed a multifunctional oscilloscope using a USB data acquisition card and Labview based on virtual technology with time domain analysis, frequency analysis and voltage measurement functions. The USB data acquisition card makes it portable and easy to install and carry. Its functions are however limited to display and analysis of signals in time domain and frequency domain.

Payal and Chauhan [5] designed a lab view version of an oscilloscope named digital phosphor oscilloscope TDS 5104B holding the features of the TDS 5104B oscilloscope, resulting in a compact and cost effective system. Its functions however

are limited to data acquisition, basic signal measurement and waveform display in time domain. Gozde and Cuniyt [6] designed a virtual oscilloscope based on Labview that could measure the amplitude, maximum voltage, rise time and fall time of signals from a pulse generator and background radiation signals from NaI (T1) scintillation detector. Chen and Liu [7] designed a virtual oscilloscope used in experiment teaching based on a pc sound card. The pc sound card used made the system to be of low cost. The results obtained showed that the pc sound card could correctly collect voice/audio signals and achieve oscilloscope measurement and spectrum analysis functions. Yi Deng [8] also designed a low cost virtual oscilloscope based on a pc sound card for time domain and frequency domain analysis. The pc sound card used makes the system cost effective. Other designed virtual oscilloscope reviewed [9, 10, 11] used labview as a programming platform for the design of a virtual oscilloscope for time domain and frequency domain display and analysis of signals.

The virtual oscilloscope presented in this paper overcomes the disadvantages of the existing oscilloscope used for the time and frequency domain analysis. It is a cost effective and flexible DAQ card and PC sound card based multifunctional virtual oscilloscope with time domain analysis, frequency domain analysis, joint time-frequency domain analysis and data analysis functions. The labview based virtual instrument was designed and developed with the capability to acquire , display and analyze signals in time- frequency domain using the short-time Fourier transform analysis (STFT) and wavelet transform analysis tool kits in labview and also carry out data analysis functions such as curve fitting (regression analysis) and statistics measurement(mean, median, and mode).

II. DESIGN METHODOLOGY

Some of the basic functions deployed in a hardware oscilloscope are acquiring a live signal, maintaining its horizontal and vertical aspects in order to fit the display in the best possible way, changing the amplitude and the frequency of the signal, analyzing the signal by measuring the waveform parameters. Oscilloscope using labview is a form of virtual instrument that is not present in real scenarios but software is built that forms the functions of the real instrument.

The designed virtual oscilloscope consists of hardware and software components. The hardware components include the general computer PCI 6251 data acquisition card and EDAS 0.25 interface console that could realize the function of data acquisition, also for a more cost effective and flexible system the PC sound card is used for audio/voice signal acquisition and recording. The software included driver programs and the user interface that controls the data acquisition and realize the functions of data collection, display and analysis. The front panel consist of the display graph and knobs and controls similar to the traditional scopes.

The system design comprises of basic layout of the program and the layman architecture for developing the device software. The multifunctional virtual oscilloscope is designed using Labview functions and control commands. Fig. 1 shows a flow chart depicting the process of system design. The conceptual block diagram of the system is shown in Fig. 2.

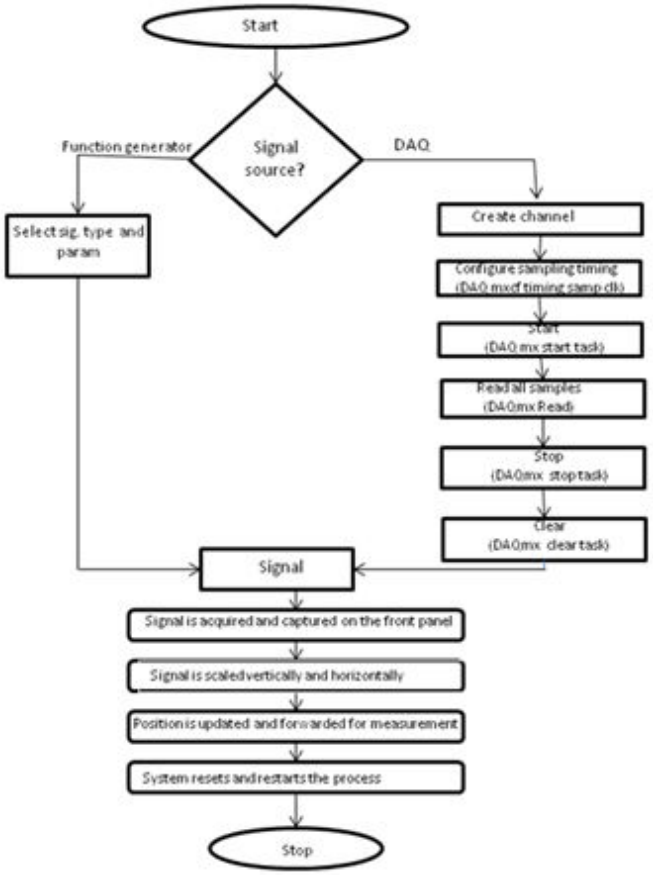


Fig. 1. Algorithm for data acquisition

The data acquisition subprogram is shown in Fig. 3. The maximum and minimum voltage level of signals is set at ±10V, the sample rate, the number of samples per channel and the physical channel can be set and adjusted from the front panel of the oscilloscope.

The NI DAQmx Create Virtual Channel (VI) and it is configured to measure analogue voltage. The NI DAQmx Read (VI) reads the waveforms from a task that contains analog input channels. It is configured to wait for the task to acquire all requested samples, then read those samples. Time out is set at 10 seconds.

The signal generator program is shown in Fig. 4. It consists of a case structure with subdiagrams of sine, triangular, sawtooth and square waveform VIs. The frequencies, amplitude, sampling rate and number of samples can be controlled. Also white noise can be added to the signals.

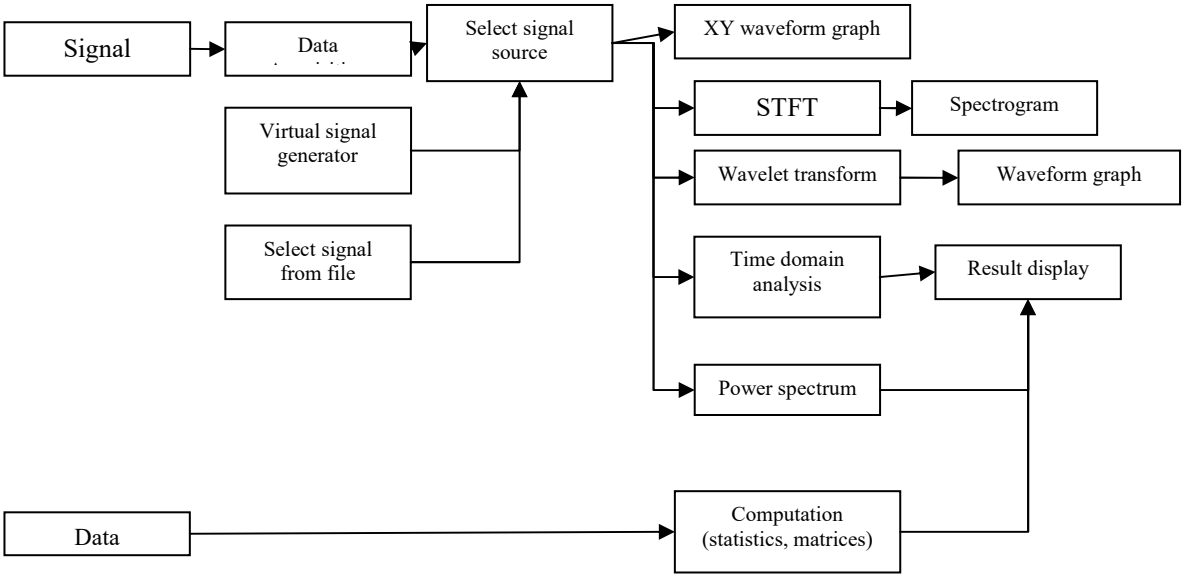


Fig. 2. System block diagram

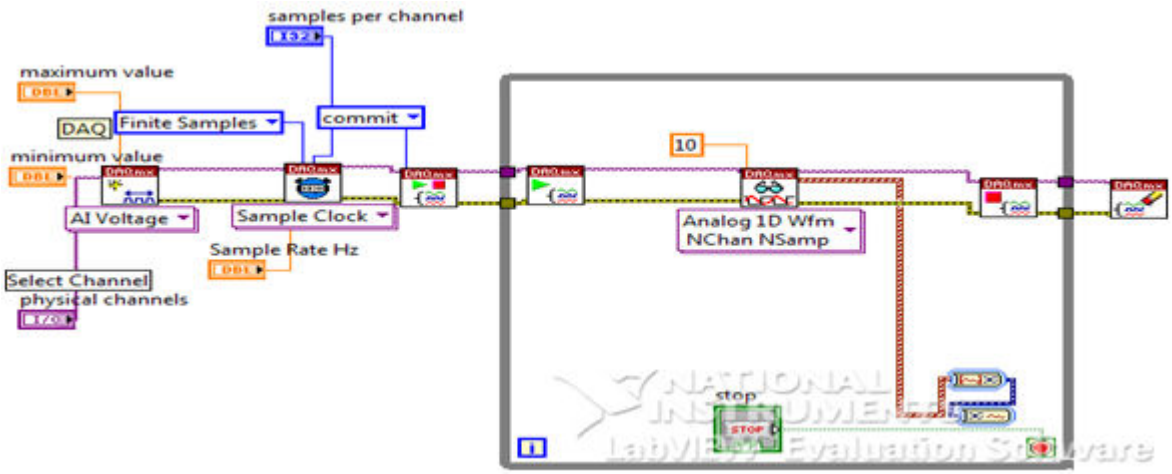


Fig. 3. Data acquisition subprogram

In addition to these, block diagrams were developed for time-domain analysis, spectral analysis, short-time Fourier Transforms Spectrogram, data analysis and wavelet transform analysis. Fig. 5 shows the block diagram program for the wavelet transform analysis module which displays a scalogram and a waveform plot of the cumulative of the coefficient of wavelet transform. One application of this is in detection of breakdown point or abrupt discontinuity in a signal. It uses the continuous wavelet transform VI which computes the wavelet transform of a signal with real valued wavelets. The type of wavelets, the scale and time steps can be controlled from the front panel. The module is also programmed to read signals from file as well as audio signals from microphone. The breakdown point locations are also displayed. The scale information is maintained on all plots and results displayed.

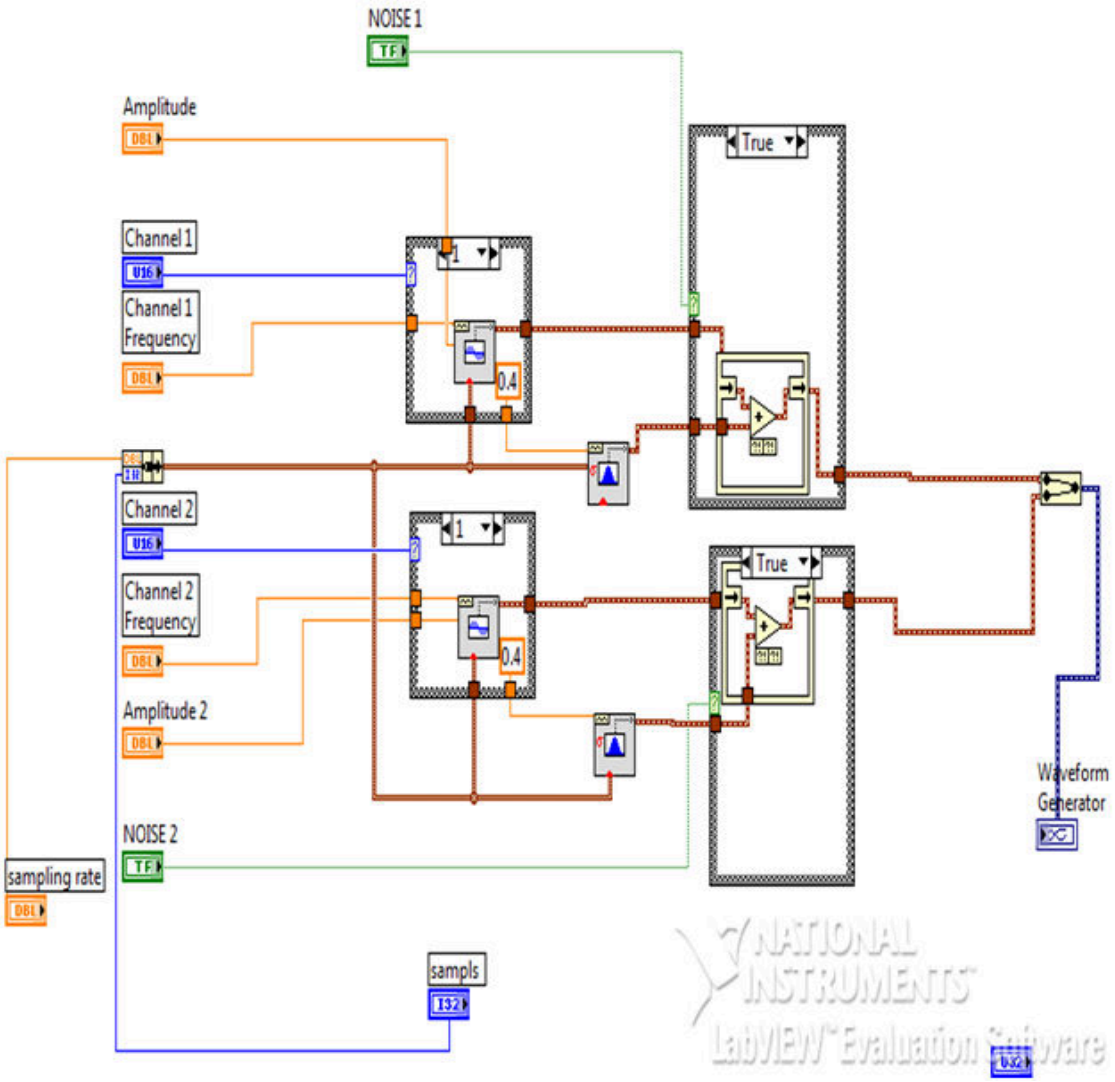


Fig. 4. Signal generator.

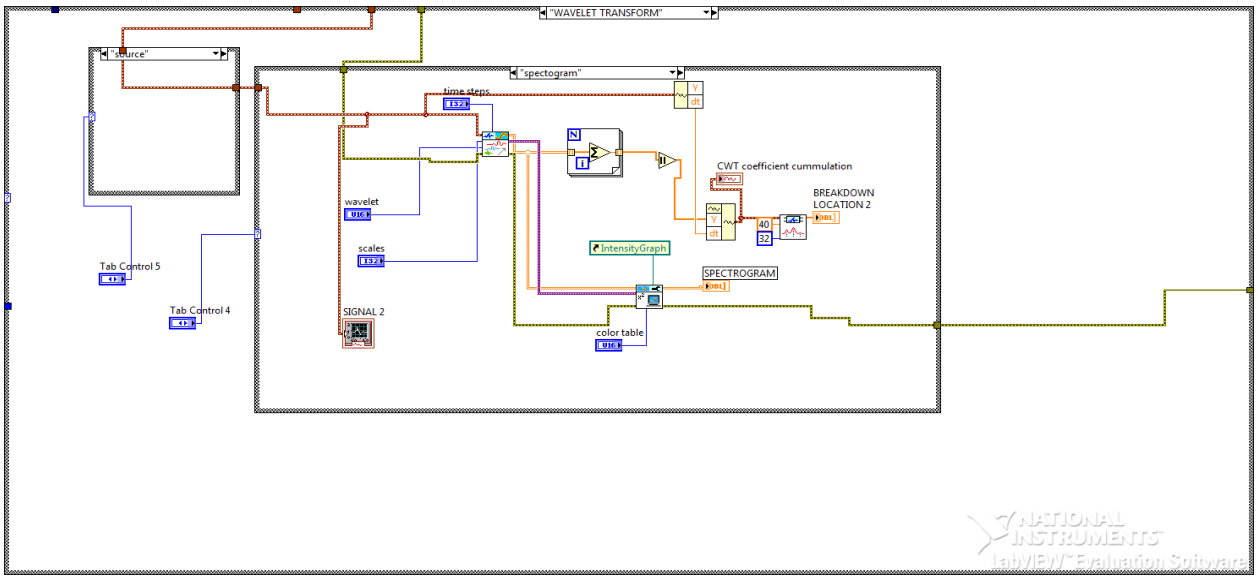


Fig. 5. Block diagram for wavelet transform analysis.

The front panel is the interface between the user and the instrument, it could be used to set, adjust and select measurements to be carried out on the oscilloscope using the panel's knobs and control buttons. The front panel consists of knobs for scale adjustment on the oscilloscope, a channel selector which has the options of selecting the source of the signal either from the virtual oscilloscope or the data acquisition system to display or measure. The front panel also has a window for various signal analysis and measurement functions. The front panel of the designed oscilloscope is shown in Fig. 6.

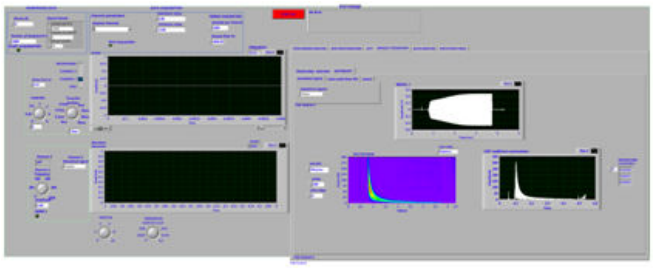


Fig. 6. Virtual oscilloscope front panel.

The front panel of the virtual oscilloscope is designed with the following functionalities

- **Command Control axis**. It allows changing the volts in every division in the amplitude axis (ordinates axis) and the milliseconds in every division in the time axis (abscissa axis)
- **Channel Selector**. It allows activating or deactivating each one of the Oscilloscope channels. Channel 1&2 from the function generator and the DAQ channel for real-time signal acquisition
- **Run/Stop**: The button ***RUN/STOP*** allows to hold the concrete signal in the graph, until the same button is pressed. This function allows working more Precisely with the cursors.
- Measurement window: it has a window for carrying out signal analysis, measurements and displaying results such as time domain analysis, spectrum analysis, short-time Fourier transform analysis wavelet analysis windows, statistics and math function windows. Each of the windows displays once the button is clicked.

III. RESULTS AND DISCUSSION

Test results obtained by acquiring sweep tones (chirp signals) through the microphone using the PC sound card as a data acquisition device are presented.

Fig. 7c shows the time-frequency plot of a sweep tone (10Hz – 2000Hz) displayed on an STFT spectrogram. Figure 7b is the standard Fourier spectrum. From this spectrum alone, one cannot tell how the frequencies have changed over

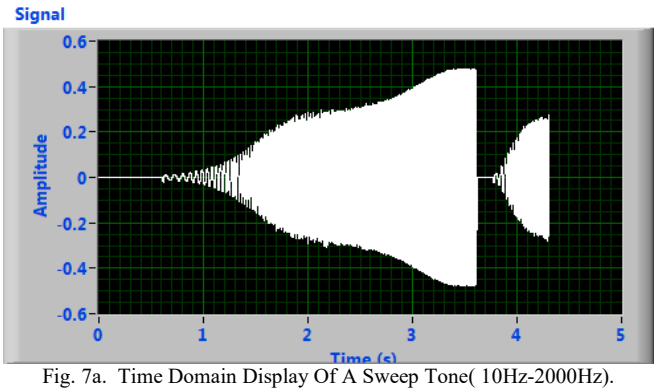


Fig. 7a. Time Domain Display Of A Sweep Tone(10Hz-2000Hz).

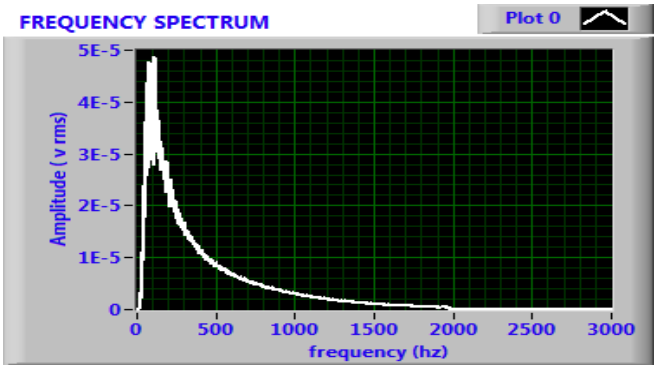


Fig. 7b. Frequency Spectrum of a Sweep Tone(10Hz-2000Hz).

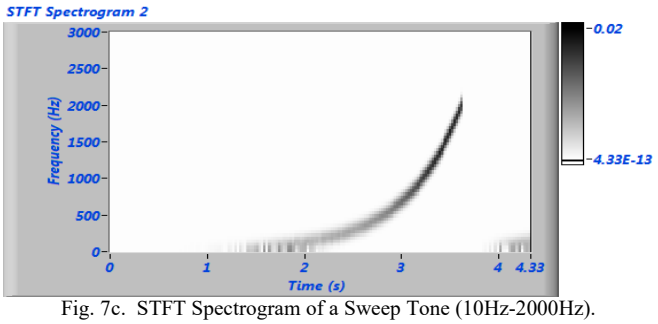


Fig. 7c. STFT Spectrogram of a Sweep Tone (10Hz-2000Hz).

time. Figure 7a is the plot of the time waveform of the chirp signal which shows only how the sound level changes as a function of time. The advantage of having a time frequency plot is that not only can one tell the range of frequencies of the signal, but also how these frequencies changed as a function of time. The spectrogram clearly shows that the signal is a sweep tone with a low frequency at the beginning which then increases as indicated on the graph,

Furthermore, in the time-frequency plot, not only can you see how the frequency changed in time, but one can also see the intensity of the frequency, indicated by the relative brightness levels of the plot.

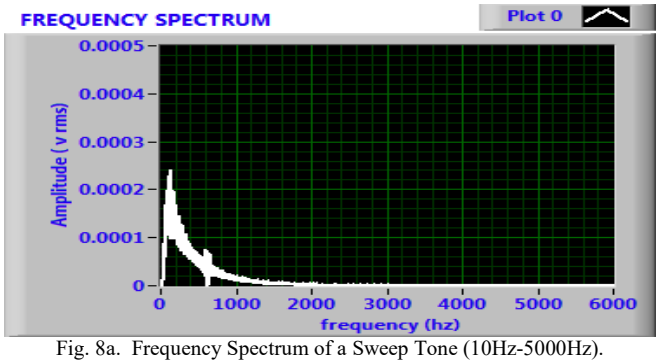


Fig. 8a. Frequency Spectrum of a Sweep Tone (10Hz-5000Hz).

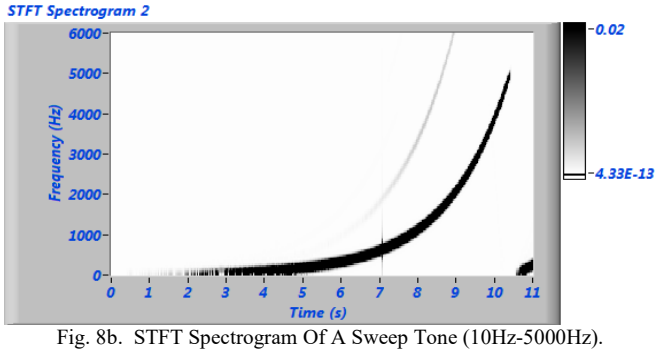


Fig. 8b. STFT Spectrogram Of A Sweep Tone (10Hz-5000Hz).

From the result shown it can be seen that the time frequency analysis function of the designed oscilloscope on the spectrogram offers a better understanding of the nature of non stationary signals compared with the FFT based spectrum analyzer.

One practical application of the spectrogram is in the detection and estimation of a signal severely corrupted by random noise hence making the detection of this signal via standard Fourier transform techniques and time waveform difficult because of its low signal to noise ratio. This is demonstrated by simulating a non linear chirp signal with added random noise and analyzing it with the spectrogram module of the oscilloscope as shown in figure 9.

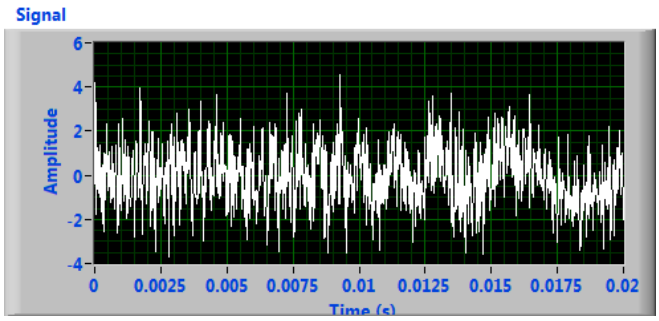


Fig. 9a. Corrupted signal.

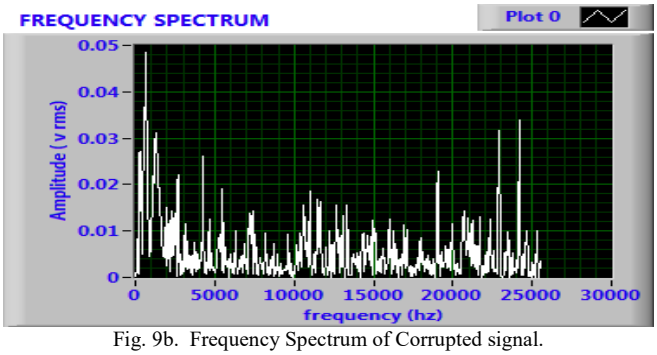


Fig. 9b. Frequency Spectrum of Corrupted signal.

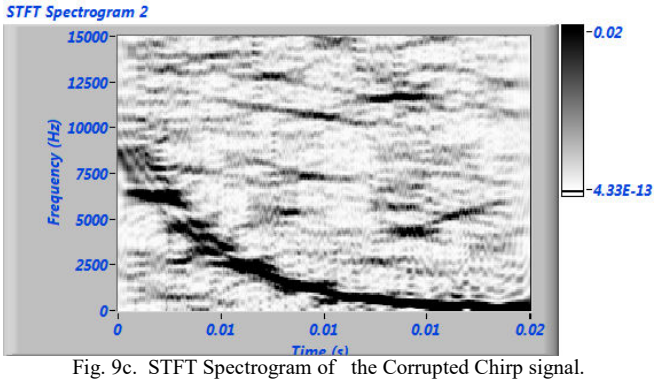


Fig. 9c. STFT Spectrogram of the Corrupted Chirp signal.

Because of the low signal to noise ratio one can hardly see the chirp signal in the time waveform as it is completely hidden by the noise . the frequency spectrum only presents the range of frequencies present in the spectrum.However, from the spectrogram which is a function of both time and frequency one can immediately identify the presence of a chirp signal arching across the joint time frequency domain. With such representation one can now do post processing to mask the desired signal.

A. Wavelet Transform

This module computes the continuous wavelet transform of a signal with real valued wavelets. Wavelet transform is the convolution of a signal with a set of functions which are generated by translation and dilation of a main function [12]. The main function is known as the mother wavelets and the translated or dilated function is known as wavelets. Mathematically the coefficient of wavelet transform (CWT) is given by

$$W(a,b) = \frac{1}{\sqrt{a}} \int S(t) \psi \left(\frac{t-b}{a} \right) dt \qquad (1)$$

where a is the dilation of the wavelets and b is the time translation.

The square magnitude of the CWT $|W(a,b)|^2$ is equivalent to the power spectrum as a function of time offset b.

This module computes the Continuous wavelet transform by specifying a set of integer values or arbitrary real positive

values for the scales and a set of equal-increment values for the shifts. The output of the CWT is the CWT coefficients, which reflect the similarity between the analyzed signal and the wavelets. A computation of the squares of the cwt coefficient forms a scalogram, which is analogous to the spectrogram in time-frequency analysis. The scalogram presents a time-scale view of the signal. In signal processing, scalograms are useful in pattern-matching applications and discontinuity detections [8].

Fig. 10 shows the front panel of the wavelet transforms modules of the virtual oscilloscope. The wavelet transform module has the following functionality on the front panel

- **TIME STEPS:** specifies the number of samples to translate, or shift, the wavelet in the continuous wavelet transform (CWT).
- **SCALES**: Specifies the number of scales of the dilated wavelet.
- **WAVELET:** Specifies the wavelet type to use to compute the continuous wavelet coefficients. The options include the following types: **Mexican Hat**, **Meyer**, **Morlet**, orthogonal (**Haar**, Daubechies (**db***xx*), Coiflets (**coif***x*), Symmlets (**sym***x*)) and biorthogonal (Biorthogonal (**bior***x_x*), including FBI (**bior4_4 (FBI)**)), where *x* indicates the order of the wavelet.

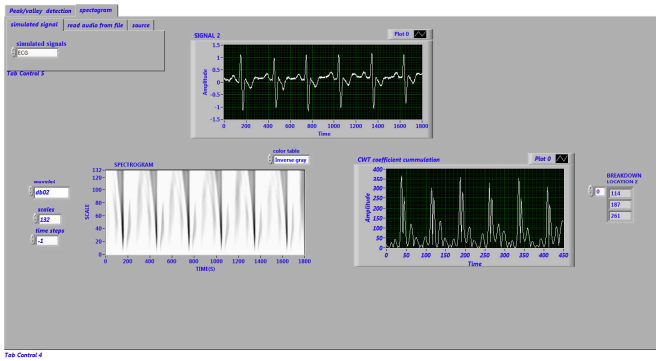


Fig. 10. Wavelet transform module.

One useful CWT application is the detection of abrupt discontinuities or breakdown points in a signal. Breakdown points and noise can generate large values in the resulting coefficients. Breakdown points generate large positive or negative coefficients at all scales while noise generates positive coefficients at some scales and negative coefficients at other scales. Thus if the coefficients at all scales are accumulated, the coefficients of breakdown points are enlarged while the coefficients of noise at different scales counteract one another.

Therefore, the peaks in the **CWT Coefficients Cummulation** graph correspond only to the breakdown points. Fig. 11 shows the plot on the scalogram from a simulated signal with two breakdown points.

The signal is a sinusoid with two breakdown points—one at 0.307 and the other at 0.737. The CWT precisely shows the positions of the two breakdown points.

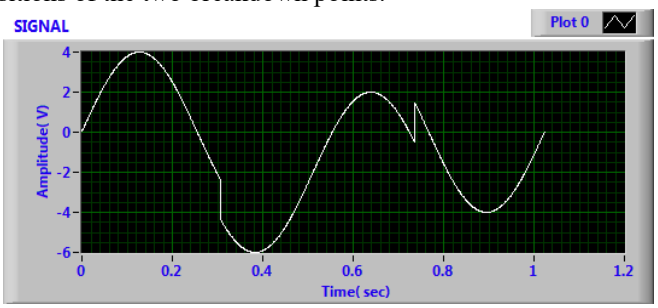


Fig. 11a. Signal with Discontinuity

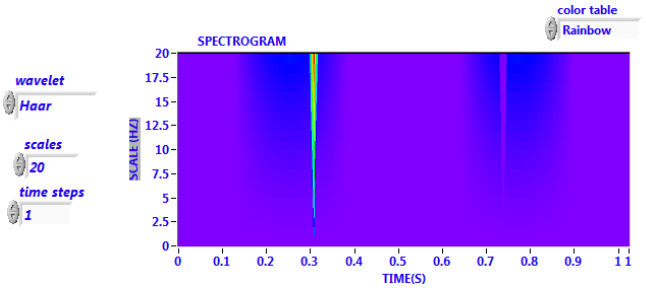


Fig. 11b. Scalogram of Signal.

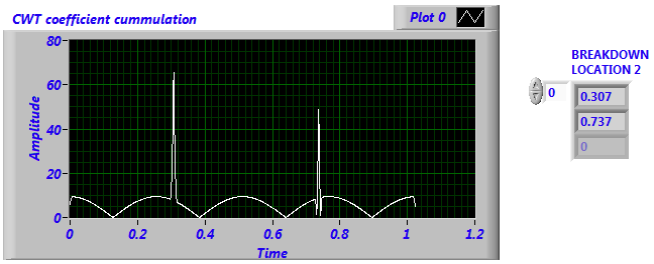


Fig. 11c. Breaking Point Detection

B. Data Analysis Module

For the purpose of computation, the designed oscilloscope has data analysis and data manipulation modules.

Data analysis functions includes curve fitting and statistics function (mean, median, range, and histogram) on XY data. Data analysis functions carried out on the oscilloscope on a set of XY data is shown in Fig. 12a to Fig. 12d.

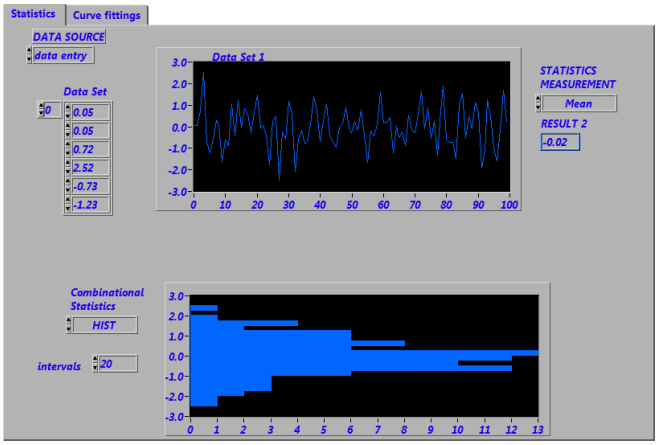


Fig. 12a. Descriptive Statistics and Histogram plot.

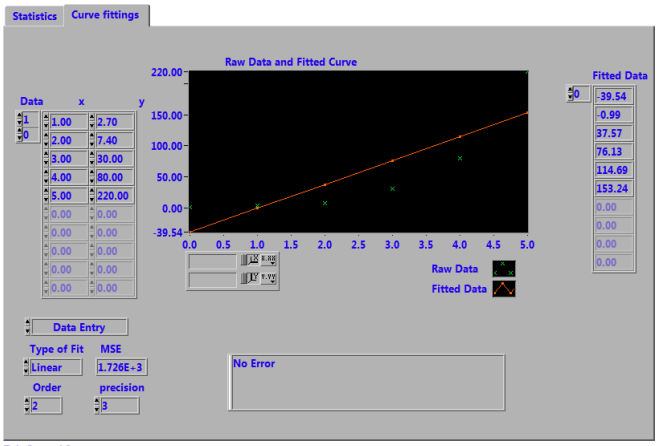


Fig. 12b. Linear Curve Fitting.

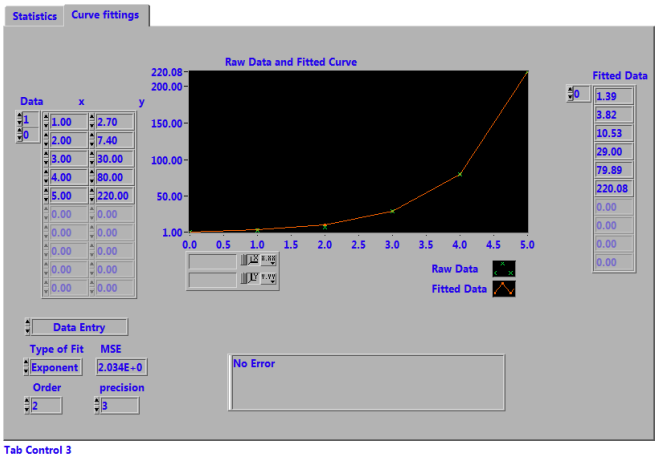


Fig. 12c. Exponential Curve Fitting.

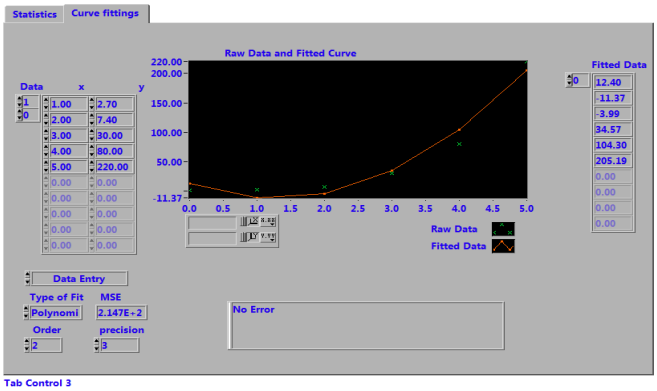


Fig. 12d. Polynomial curve fitting.

C. *Performance Comparison*

The performance of the proposed oscilloscope has been compared with existing oscilloscopes reported in recent literature. A valid observation across the reviewed virtual oscilloscopes is that most of them are limited in functionality in relation to our proposed scope.

For example, unlike our oscilloscope, the virtual oscilloscope presented in [2] could only analyze signals in time-domain just like the system presented in [12]. Although the system presented in [3] is also designed on the platform of NI Labview it is limited to the analysis of stationary signals and it does not analyze nonstationary signals. This is exactly the problem with the scopes presented in [5] and [6].

Joint time and frequency analysis cannot be carried out by the systems presented in [4] and [8 – 10] even though they can analysis signals in time and frequency domain including FFT computation in some cases.

IV. CONCLUSION

A multifunctional virtual oscilloscope based on the graphical programming capabilities of National Instruments Labview with signal acquisition and signal analysis function (stationary and non stationary signals), data analysis and data manipulation functions have been developed. The major features of the oscilloscope include Signal acquisition functions from both NI DAQ card and PC sound card, Time domain measurements (Amplitude, frequency Peak to peak), Spectrum analyzer (plot of magnitude vs. Frequency) and Short-time Fourier transforms spectrogram with the additional choice of specifying the window length for the STFT analysis.

Other features are Wavelet based Time-Frequency analysis with the option of selecting time-frequency sampling parameter like time steps, scales and the choice of the wavelet function and Data analyses function on XY data such as descriptive statistics (mean median and mode), histogram plots and curve fitting functions.

REFERENCES

[1] [1] D. Kwock. "Virtual instrument" in Electronic Instrument Handbook. New York: Mc Graw –hill, 2004, pp. 1128-1151. [Online]. Available: ww.digitalengineeringlibrary.com.

[2] [2] C. Bhunia ; S. Giri ; S. Kar ; S. Haldar ; P. Purkait ,

[3] "A low-cost PC-based virtual oscilloscope," IEEE Transaction. Edu. vol. 47, no. 2, pp. 295-299, 2004.

[4] [3] J. Wei and Y. Fang. "Design of oscilloscope based on virtual instrument technique". In proc. Second International Conference on Power Electronics and Intelligent Transportation System, 2009, pp 284-287.

[5] [4] G. Ping and Z. Wei. "Design and implementation of multifunctional virtual oscilloscope using USB data - acquisition card", in International Workshop on Information and Electronics Engineering, Procedia Engineering 29 (2012) 3245-3249, 2012.

[6] [5] R.T Payal, and N. Chauhan. "A Virtual Instrument Oscilloscope for Signal Measurements". International Journal Of Engineering Research and General Science Volume 3, Issue 3, pp. 1014-1018, May-June, 2015.

[7] [6] T. Gozde and C. Cuneyt. 'Applying Virtual Oscilloscope to Signal Measurement in Scintillation Detectors'. Radiation Science and Technology, Vol. 1 No. 1, pp. 1-5.2015

[8] [7] L Guili, K Quancum, ' Design of Virtual Oscilloscope Based on GPIB Interface and SCPI', in The 11th IEEE International Conference on Electronic Measurement and Instruments, 2013 .

[9] [8] D. Yi, "Research and Design of Virtual Oscilloscope Based on Sound Card", in International Conference on Electrical and Control Engineering, 2010 © IEEE. doi:10.1109/iCECE.2010.386

[10] [9] W Jiang, F.Yuan, "Design of Oscilloscope Based on Virtual Instrument Technique", in 2 nd International Conference on Power Electronics and Intelligent Transportation System , 2009

[11] [10]Z. Xian-Ling, "The Virtual Instrument Based on Labview and Sound Card", in International Conference on Computational Aspects of Social Networks, 2010 © IEEE. doi: 10.1109/CASoN.2010.170

[12] [11]W Lijun, et al., " Development of Multi-functional Virtual Oscilloscope Based on Soundcard", in Third International Conference on Measuring Technology and Mechatronics Automation, 2011. doi: 10.1109/ICMTMA.2011.264

[12] A. Chen and J. Liu, "A Kind of Virtual Oscilloscope Used in Experiment Teaching Based on Sound Card", in Second International Conference on Education Technology and Training, 2009. doi: 10.1109/ETT. 2009. 34

Machine Learning for Strategic Urban Planning

S. N. Odaudu , I. J. Umoh , M. B. Mu'azu, E. A. Adedokun
Department of Computer Engineering
Ahmadu Bello University
Zaria, Kaduna, Nigeria

Abstract - Data mining is an important part of strategic planning for the development of modern urban settlement with capacities to accommodate population explosion. Developing countries are fast becoming urbanized giving the developments and opportunities that are lacking in rural areas. Data regarding urban development such as satellite image need to be analysed to ascertain the possibilities for further development or opening up of new settlements. This work presents a binary sub-pixel and feature based method of classification to detect water bodies and vegetation in earth observatory images. In this work, the images were subjected data pre-processing, feature extraction, and analysed the data using machine learning classification method to detection regions that support urban expansion or development of new settlement. The proposed method achieved 88.93% accuracy and 0.06% RMSE.

***Index Terms* - Satellite image, object detection, machine learning, Haar- cascade, feature engineering.**

I. INTRODUCTION

Data management has become a crucial ingredient for national development and human capacity development as it account for virtually every part of our daily life. Over 50billion device are expected to communicate and generate data that will exceed petabytes in 2020. Data in some quarters has been referred as the black oil of the millennium far richer in content than crude and other resources. This informs the concept of data mining and analytics for national development. This potential holds the future of modern governance, security, transportation, Agriculture and urban development. According to the United Nation assessment on urban growth in developing country, the percentage of urbanization rose to about 10.5 % between the period of 1980 -1996 [1, 2]. This movement of people from the rural areas into cities places a strain on basic amenities and governmental presence. [2, 3]. Due to this increase there is need for new urban development and expansion of existing ones to accommodate the increase in both human and goods influx from rural to urban settlement.

To address this challenge of continuous population explosion in urban settlement, there is need to develop a strategic and good data system that can account for these population in terms of planning and decision making.

Data analytic is a scientific and mathematical approach that involves data cleaning, manipulation, analysing for the purpose of decision making [4]. Data available for development includes population index, national boundary data, and satellite image information.

Satellite information for national development can be obtained in the form of a digital image [5]. A digital image is one that has its pixels represented in X, Y coordinates. Information capture digitally upon analysis can be translated into meaningful decision tool for urban development [5, 6]. In every urban settlement, water is essential for agriculture, manufacturing, human consumption and other activities. However, it's a major problem in most urban settlement in developing countries. The inability of governments to provide quality drinkable water, for irrigation farming and industry has led to scarcity of water. Sometime, lack of information on locating presence of water bodies has contributed to poor urban planning and location of new settlement. However, it has become imperative to employ the use of data for future urban planning [7]. Thus, this work employs the use of machine learning classification models and feature selection algorithms to detect water contents in satellite images for urban development.

II. RELATED WORKS

This section discuss related literatures, study area, dataset, tools, and algorithms. Pasquale *et al.,* **(2017)** [8] Used satellite imagery and crowd sourcing to assess population covering of a health demographic surveillance. Their work reported results from satellite image and information obtained from recruited volunteers[1] from crowdcrafting.org to ascertain the demographic surveillance system of Malawi. Xian, G., & Crane, M. (2005) [9] evaluated the influence of urbanization on urban climate with remote sensing and climate observation by utilizing remote sensing data to access land cover and its thermal characteristic by mapping sub-pixels values. An inverse relationship between imperviousness (sum of roads, parking lots, rooftop and other form of modern landscape) and Normalized difference of vegetation index NDVI was calculated. Authors in [10-12] Proposed a quantitative approach to land use change with focus on the changes in urban agglomeration and demography, land use and cover; the work seek to establish the correlation between population increase in an urban setting and the problem of water supply and other

environmental degradation. Satellite image were analysed based on features like water, urban, agriculture and forest. Kumar *et al.,* **(2017)** [13] Characterized land cover features into four class through the use of random forest a machine learning algorithm to explore sub-pixels for the purpose of estimating Global land cover fraction from satellite data using high performance computing. The work was reported to have 6% improvement in unmixing pixels based classification when compared to pre-pixel classification.

I. A. Study Area

Nigeria, a West African state that has urbanization index of about 49.4% as at 2017 and urban growth of 4.3% annual growth [14] is a choice study area for this study because of her rapid urban growth. Nigerian north central geopolitical zone comprise of the following state Benue (30732.69sqkm), Kogi (29044.79sqkm), Kwara (36066.31sqkm), Nasarawa (26385.04sqkm), Niger (72200.9sqkm) and Plateau whose dwellers are predominantly farmers and Civil servants [15]. Due to population increase and rapid urbanization, state around Abuja are bound to experience population explosion as such there is need for adequate planning into need settlements [15].

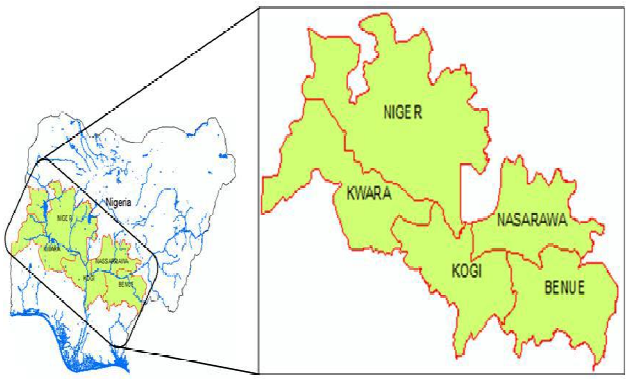


Figure1: Map of Study Area [15]

SEX RATIO OF DEMOGRAPHY OF NORTH CENTRAL, NIGERIA 2006 CENSOR [16]

State	Male (%)	Female (%)	Ratio of Males to females
Abuja	55	45	122.2
Benue	58.3	41.7	139.9
Kaduna	57.7	42.3	136.5
Kwara	54.2	45.8	118.4
Kogi	61.1	38.9	157
Nasarawa	62.2	37.8	164.8
Niger	62.4	37.6	165.7
Plateau	57.7	45.7	119

TABLE2:
LANDSAT8 IMAGE RASTER DATA EXTRACTED

Band	Wavelength	Colour
Band1	0.45-0.52	Blue
Band1	0.52-0.62	Green
Band3	0.63-0.69	Red
Band4	0.76-0.90	Near infrared
Band5	1.55-1.75	Mid – infrared
Band6	10.40-12-50	Thermal infrared
Band7	2.08-2.35	Mid infrared

B. Dataset

The dataset considered for this study are landsat8 satellite data covering the north central region of Nigeria with emphases on Kaduna, Niger state and Abuja FCT. The Figure 2 represent a landsat8 satellite imagery captured in 2016 the image contains features such as water, settlements, Vegetation and other features. Data were collected from Nigerian geological survey Agency, Abuja.

Figure 2: Landsat8 Image obtained from USGS Database

Landsat image contains seven bands represent different color spectrum details are shown in table 2.

C. Tools

R language is a data analytic tool developed for processing and analyzing data.in this study, R as a tool was used to visualize the dataset as to understand the distribution of the pixel properties and to perform feature classification using the Raster Library.

D. K- MEANS

K-Mean algorithm is an algorithm that group or cluster unlabeled dataset. It has two iterations the first is the cluster assignment step and the second is the move centroid step. The first step assigns data into a group with similarities or nearness of features and the second on the other hand classify the dataset based on statistical parameters such as mean and average [17].

To perform K-means clustering, first, the number of clusters must be defined then, the algorithm groups the dataset into the specified numbers of clusters. Equation (1) represents Kmeans model [18]:

$$U = \sum_{j=1}^{k}\sum_{j=1}^{h}(g_i - g_j)(1)$$

Having a dataset of h , applied equation (1) will partition the data into k number of clusters having a distance function d the mean feature and i_{th} features that are j_{th} related is represented as g_i and g_j respectively. Distance d between two points in an image is given as

$$d = \sqrt{\sum_{i=1}^{h}(P_{1i} - P_{2i})} \qquad (2)$$

Where $(P_{1i} - P_{2i})$ represents the distance between point 1 and 2 on an image is is, h represents number of observation in the set.

E. HAAR-CASCADE

The Haar-like feature is a machine learning approach developed by Viola and Jones in 2001 for visual object detection which has the capacity of processing image very fast and achieving high detection rate [19]. The principle is such that all human faces have a certain feature that is similar in nature. And these regularities can be used for matching. Below are some of the features that are common to all human face [20]:

1. The eyes region is darker than the nose and upper cheek
2. The nose bridge is brighter than the eyes.

Trainable features used in matching faces:

1. The location and size of human eyes, mouth and nose bridge
2. Value – the intensity of pixels forming a location. The algorithm relies on the above mention four features to detect and match the human face. Training the algorithm requires that each region is represented in the form of a black and white rectangle which is dark and bright respectively [20]. The algorithm takes into account the sum difference between pixel value taken from the dark region and compared to the sum value of the pixel of bright region. The algorithm is an ensemble learning that requires a combination of both weak and strong classifiers in other to produce a better result [20].

The algorithm is of four sequences which are [20]:

1. Haar feature selection
2. Creating an integral image

$$vv(x,y) = \sum_{\substack{x' \le x \\ y' \le y}} (x', y') \qquad (3)$$

$$\sum_{(x,y)\in PQRS} v(x,y) = vv(S) + vv(P) - vv(Q) - vv(R)\,(4)$$

Where $h^{(i)} i = 1, \ldots\ldots k$ in equation (5) represents weak classifiers selected by boosting in training process and $\propto^i i = 1, \ldots\ldots, k$ are the corresponding weight. Figure 3

3. Adaboost Training**:** Adaboost is a member of the boosting Meta algorithm family of classifiers that gives stronger classifiers by combining a set of weak classifiers. The predictions from numerous weak classifiers are consolidated through weighted majority voting to create the prediction of the strong classifier. For parallel characterization issues, the strong classifier has the structure

$$H(x) = \begin{cases} 1 \sum_{i=1}^{k} \propto^{(i)} h^{(i)}(x) \geq 1 \\ 0\ Otherwise \end{cases} \qquad (5)$$

4. Cascading of classifiers.

Cascading or combining the classifiers, the training process involves [20]:

F= the minimum accepted false positive rate

D = the minimum accepted detection rate

P = set of positive

N = set of negative

In this method of learning, each classifier depends on the behaviour of the classifiers before it, the false positive rate can be calculated as[20]:

$$f = \prod_{i=1}^{k} f_i \qquad (6)$$

And detection rate is:

$$D = \prod_{i=1}^{k} d_i \quad (7)$$

III. METHODOLOGY

A. Feature Engineering

Feature extract form the basic and fundamental process in every machine learning Data Mining process for image processing. The essence of this process is to extract information that will guild us into understanding hidden pattern in dataset. In this work, our feature selection techniques is based on calculated distance between features that are near to each other in an image L.that is if feature L_i is randomly selected from its class, and further search for Q_i

of its nearest neighbours of same class B and K nearest neighbours from a different class M at distance d , it update the Accuracy measures u_i for feature i based on the value of L_i , Q_i and K . If the features in L_i and features in B has conflicting estimating value on feature i , then the estimation of L_i will decrease in value otherwise, Q_i feature and M feature will experience increase on estimate value of i feature. When the process is performed iteratively for n

number of time, the update function is given as

$$u_1 = u_1 - \frac{\sum_{k=1}^{k} dQ_1(k)}{n.k} + \sum_{c=1}^{c} P_c \frac{\sum_{k=1}^{k} d_m(k)}{n.k} \quad (8)$$

From equation (2), the distance between two point and coordinate in an image using k-mean algorithm is given as d . Therefore, the sum of distance between selected features and the K_{th} nearest neighbours B and (M) in equation (8) $dQ_1(k)$or $d_m(k)$, P_c Represent the probability of class$c$.

B. Haar Training

Haar cascade is a feature base classification model that take digital image as features for training. Feature detector is a classifier that classifies selected feature within an image dataset.

To train a classifier, it required that the set

$$(D_2L_2), \ldots\ldots\ldots (D_mL_m) \quad (8)$$

Where $D_i \in R^n, l_i \in \{-1,+1$ and M represent the sum of training samples. In Haar algorithm, selected feature is represent by using +1 for positive region and -1 label for negative region. Extracted feature vector is given as $D_{i\,i}$ which in turn has a n number of observation. Features extracted from positive data and negative data sum the total information of the dataset.

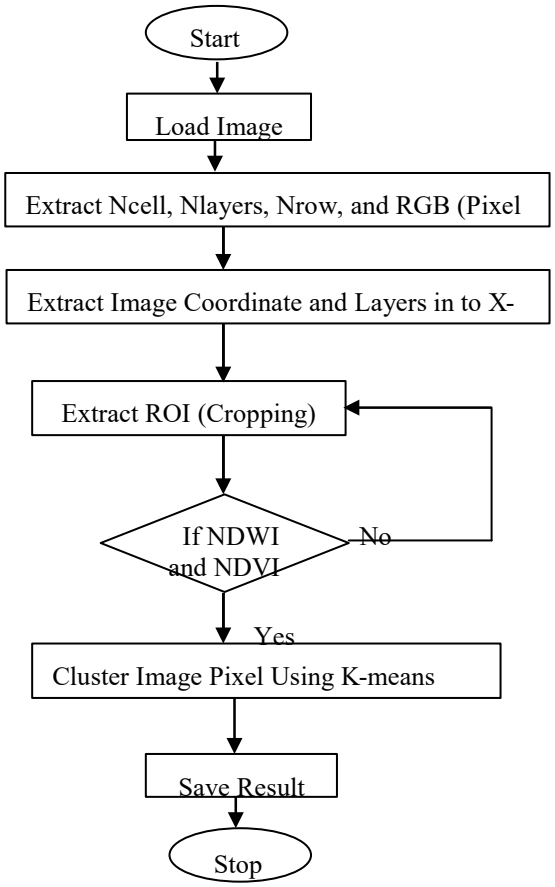


Figure 3: Flowchart of feature Extraction.

C. Training for Detection

When a feature is input to a classifier, it is classified by the cluster of classifiers as [0, 1] using the weighted average to make decision that will tell if the object is positive or negative and this forms their final decision. Results from weak classifiers are further combined and process by stronger classifier. To recombine the object registered by weak classifiers the model is represented as equation (10)

$$V_{(x)} = \begin{cases} 1 \prod_{i=1}^{k} (v)^{i(x)-1} \\ 0 \end{cases} \quad (10)$$

Where $V^i(x)$ represent boosted set of weak classifiers that wrongly detect features and k is the numbers of cluster of classifiers

D. Normalized Difference Vegetation Index (NDVI) And Normalized Difference Water Index (NDWI)

This study also considered the implementation of Normalized difference vegetation index and normalized difference water index. The process is binary pixel classification technique to investigate the percentage both water and vegetation content in the set of data.

The two approach are fundamental to investigating vegetation and monitoring biological content in a location through satellite images. These principles rest on approach of Feature extraction from satellite image for analysis which requires the use experts and high definition tools such as computer vision application and data analytic tools to unveil information contained in an image that is not visible to the human eyes. Water bodies and vegetation region can be extracted through the use of NDWI and NDVI. NDVI represent the index of vegetation density (-1,-0.1) while NDWI is the index of water body within a giving location duly represented in a digital satellite image. NDWI can be calculated by the using equation

(11)

$$NDWI = \frac{NIR - GREEN}{NIR + GREEN} \quad (11)$$

Vegetation can be detected by comparing reflection from the visible and the infrared reflection of light from vegetation. Equation (12) represents the Mathematical model for calculating NDVI.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (12)$$

IV. Result and Discussion

This work implement an object detection in satellite image, using Haar – like algorithm, data was collected, pre-processed, feature engineering and finally analysed the features with Haar. The work shows an improved performance in accuracy, precision and Recall as the size of data increase from 40% to 100%. Table 3 represents result of this work:

Figure 4 represents the results of feature visualization using Kmeans algorithm. The algorithm clustered the features into groups of related pixel colour bands in such that each feature is given a cluster class. Green patch in the image represents vegetation, the yellow patch is interpreted as land or rock while the third feature represent water bodies.

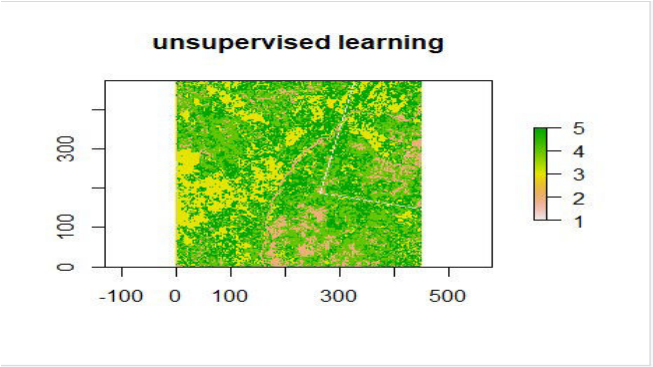


Figure 4: Unsupervised clustering

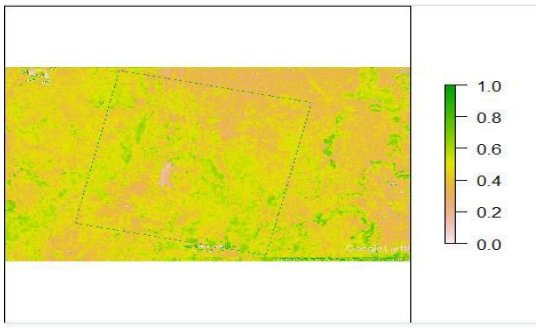


Figure 5: NDVI implementation

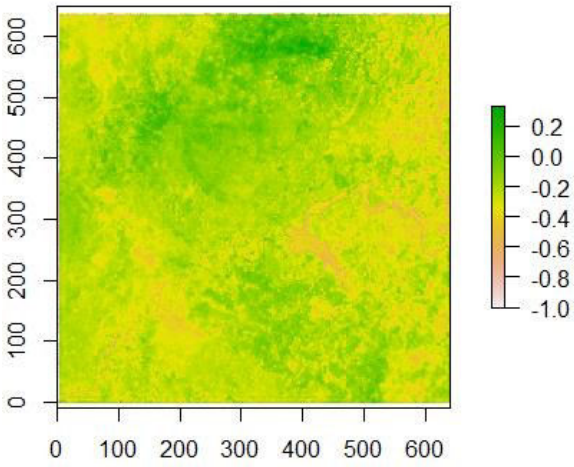


Figure 6: NDWI

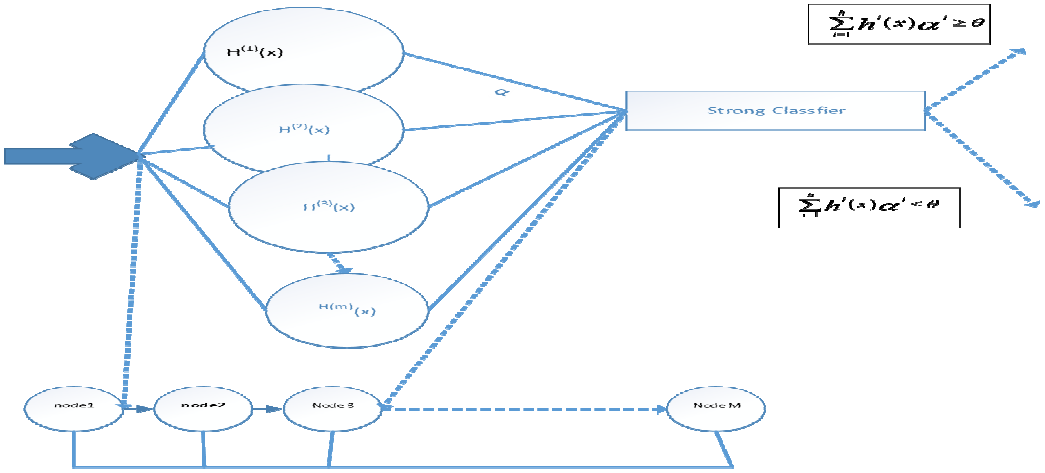


Figure 3: combined structure of ensemble classifiers arranged decision tree fashion.

Figure 5 and Figure 6 are the pictorial result of NDVI and NDWI respectively. These approach were adopted to investigate through object visual method to extract features for training of our model. The result shown affirm the presence of the desired features.

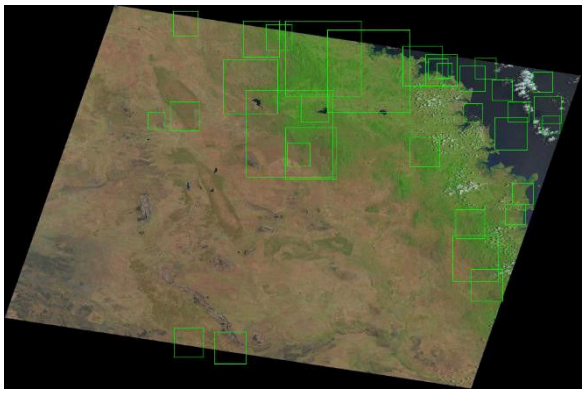


Figure 7: Water detection and identification

A. **Table 3: Confusion matrix Table**

Data size (%)	Accuracy	Sensitivity	Specificity
100	88.2	89.7	82.2
80	82.55	82.6	72.2
60	65.1	75.3	63.7
40	63.81	77.9	82.2
20	63.76	77.9	82.2

Figure 7 and Table 3 has shown that our proposed feature based algorithm have high accuracy results in detecting and identifying water body within satellite image. The results also shows that the error ratio was very low while specificity and sensitivity remains high.

Conclusion

In this paper, we propose the application of Machine Learning technique on satellite images for enhance urban development and planning of new settlement. The technique adopted used dataset obtained from Landsat 7 and 8 from the U.S. Geological survey and the Nigerian Geological survey Agency, Abuja and Haar Algorithm to detect features that are key to urban development. Feature extraction was carried out using NDWI and NDVI for the purpose of feature visualization. The propose technique has an Accuracy of 88.2%, 89.7% of Sensitivity and 82.2% for Specificity.

Acknowledgement

The authors would like to appreciate members of Computer and control research group of the Department of Computer Engineering, Ahmadu Bello University, Zaria for their immerse contribution in the course of this work. Also the U.S. Geological Survey (USGS) and the Nigeria Geological Survey, Abuja for making available the dataset for this work

REFERENCES

[1] C. B. d'Amour, F. Reitsma, G. Baiocchi, S. Barthel, B. Güneralp, K.-H. Erb, H. Haberl, F. Creutzig, and K. C. Seto, "Future urban land expansion and implications for global croplands," *Proceedings of the National Academy of Sciences,* vol. 114, no. 34, pp. 8939-8944, 2017.

[2] L. Jiang, and B. C. O'Neill, "Global urbanization projections for the Shared Socioeconomic Pathways," *Global Environmental Change,* vol. 42, pp. 193-199, 2017.

[3] D. Dodman, "Environment and Urbanization," *The International Encyclopedia of Geography*, 2017.

[4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM,* vol. 39, no. 11, pp. 27-34, 1996.

[5] Y. Zhang, *Image analysis*: Walter de Gruyter GmbH & Co KG, 2017.

[6] Z. Zhou, Y. Wang, Q. J. Wu, C.-N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," *IEEE Transactions on Information Forensics and Security,* vol. 12, no. 1, pp. 48-63, 2017.

[7] L. Woltersdorf, M. Zimmermann, J. Deffner, M. Gerlach, and S. Liehr, "Benefits of an integrated water and nutrient reuse system for urban areas in semi-arid developing countries," *Resources, Conservation and Recycling,* vol. 128, pp. 382-393, 2018.

[8] A. Di Pasquale, R. S. McCann, and N. Maire, "Assessing the population coverage of a health demographic surveillance system using satellite imagery and crowd-sourcing," *PloS one,* vol. 12, no. 8, pp. e0183661, 2017.

[9] Xian, G., & Crane, M. (2005,). Evaluation of urbanization influences on urban climate with remote sensing and climate observations. In *Proceedings of the ISPRS joint conference 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas (URBAN 2005) 5th International Symposium Remote Sensing of Urban Areas (URS 2005) Tempe, AZ* (pp. 14-16).

[10] C. Wei, H. Taubenböck, and T. Blaschke, "Measuring urban agglomeration using a city-scale dasymetric population map: A study in the Pearl River Delta, China," *Habitat International,* vol. 59, pp. 32-43, 2017.

[11] A. K. Misra, M. Masoodi, and R. P. Poyil, "Water demand and waste management with respect to projected urban growth of Gurugram city in Haryana," *Beni-Suef University Journal of Basic and Applied Sciences*, 2018.

[12] I. Buchori, A. Sugiri, M. Maryono, A. Pramitasari, and I. T. Pamungkas, "Theorizing spatial dynamics of metropolitan regions: A preliminary study in Java and Madura Islands, Indonesia," *Sustainable Cities and Society,* vol. 35, pp. 468-482, 2017.

[13] U. Kumar, S. Ganguly, R. R. Nemani, K. S. Raja, C. Milesi, R. Sinha, A. Michaelis, P. Votava, H. Hashimoto, and S. Li, "Exploring Subpixel learning Algorithms for Estimating Global Land Cover Fractions from Satellite Data Using High Performance Computing," *Remote Sensing,* vol. 9, no. 11, pp. 1105, 2017.

[14] M. E. H. Arouri, A. B. Youssef, C. Nguyen-Viet, and A. Soucat, "Effects of urbanization on economic growth and human capital formation in Africa," 2014.

[15] M. Ojigi, F. Abdulkadir, and M. Aderoju, "Geospatial mapping and analysis of the 2012 flood disaster in central parts of Nigeria." pp. 1067-1077.

[16] O. Okafor Samuel, "Scoring the Census Priority Table for Age Heaping and Shifting: A Study of 2006 Nigeria Population Census Result."

[17] A. K. Sangaiah, A. E. Fakhry, M. Abdel-Basset, and I. El-henawy, "Arabic text clustering using improved clustering algorithms with dimensionality reduction," *Cluster Computing*, pp. 1-15, 2018.

[18] A. K. Jain, "Data clustering: 50 years beyond Kmeans," *Pattern recognition letters,* vol. 31, no. 8, pp. 651-666, 2010.

[19] R. Mustafa, Y. Min, and D. Zhu, "Obscenity detection using Haar-like features and gentle adaboost classifier," *The Scientific World Journal,* vol. 2014, 2014.

[20] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features." pp. I-I.

[21] A. Kaehler, and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*: " O'Reilly Media, Inc.", 2016.

Buffer Occupancy Threshold Selection Strategy for Optimal Traffic Distribution Across a Hierarchical Network Onchip

Umar I. Mohammed, Umoh J. Imeh, Abdulrazaq M. Bashir
Department of Computer Engineering
Ahmadu Bello University, Zaria

***Abstract*— This research work uses a frequency sweep to determine the buffer occupancy that gives the best throughput for uniform random traffic pattern. Load imbalance in Network on Chip occurs when certain nodes experiences heavy traffic situation than they can carry while other nodes experiences less traffic. This situation causes congestion into the network and subsequently, performance deterioration. Among the several solutions offered to address this problem is to set a bridge node that can be used to jump across nodes experiencing heavy traffic. This is done by setting a buffer occupancy threshold value on each bridge link, which decides when to route packets across any node experiencing congestion. These solution, despite the performance increase recorded, has serious short-comings. There is a trade-off in selecting a suitable threshold value. The trade-off is between low threshold which triggers more frequents jumps and high threshold which injects more packets and imbalances the network. In this research, a sweep of frequencies, at step interval of 0.1 is applied across the range 0.0 to 1.0 to determine the buffer occupancy that gives the best throughput. It was observed that the best throughput occurred around a buffer occupancy of 0.5. This optimizes the network for the uniform random traffic pattern and gives the best trade-off.**

***Index Terms*—Congestion, Network on Chip, Router, Threshold, Topology, Traffic Load Distribution.**

I. INTRODUCTION

Background to the Research

High radix Network on Chip (NoC) exhibit low performance due to large end-to-end packet hop counts which raises latency [1]. This is an undesirable situation. Among the solutions given by researchers is the use of hierarchical topologies, which divides the network into hierarchical layers. Upper hierarchies are mostly used as communication short-cuts to distant destination nodes. To effectively use the hierarchies, packets are usually divided into local and global traffics based on their calculated Manhattan distances to destinations [2]. Local traffics only communicate with their neighbours, which are not more than a few hops distances a way, while global traffics traverses longer distances usually through the upper hierarchies [3]. The upper hierarchy sizes are usually a fraction of the base layer and therefore have fewer communicating nodes. This means reduced path diversity and bandwidth, which pose no problem as far as the local traffic dominate the network data. However, if the global traffic dominates, more packets traverse the upper layers than they can take, which causes imbalance and subsequently, severe congestion can easily form [4], [5]. Mitigating congestion by balancing the network load among the hierarchies was looked into by Bai *et al.,* (2016) who proposes the use of fixed threshold value across the link connecting any two layers. As good as this solution is, it is limited by trade-off and application specific requirements: usually a time-consuming, repetitive trial and error process.

Significance of the Research

The ever increasing requirements by computing applications for large processing power under low energy and power constraints led to the continued integration of onchip components. This is depicted in Multiprocessor System on Chip (MPSoC) and Chip Multiprocessors (CMP) designs that have tens to hundreds of Intellectual Properties (IP) consisting of processing elements, memory units and controllers [6]. As the density of chip integration continuous to increase, NoC becomes the de-facto means of onchip communication. The large number of IPs in these onchip systems necessitate the requirements for low radix NoC communication infrastructure, so that end-to-end latency is kept as low as possible for improved performance [7]. Two dimensional hierarchical mesh NoC topology is one of such low radix NoC that can be implemented in MPSoC and CMP.

The significance of this research is that, it provides a practical means of determining the buffer occupancy threshold that gives the best throughput for a given traffic pattern, across various traffic injection rates. The sampled injection rates may cover any desired level of saturation. The determined buffer occupancy threshold can then be used to make routing decisions through the bridge links which is hped will give better performance in a given hierarchical NoC topology.

Statement of the Problem

The problem definition can be stated in the following form. Given a hierarchical NoC topology that has two or more hierarchy layers, how can the buffer occupancy threshold be determine such that it meets good performance requirement for all traffic conditions and applications? A trial and error process is time-wasting and involves a trade-off. It is expected that a low value setting triggers off the link frequently and deplete the upper layers of packets, while a high value setting triggers the link less frequently and injects more packets into the upper layers, thus

disturbing the balance. A sweep of buffer occupancies at 0.25 step interval over the range 0.0 to1.0 is plotted against the throughput for the uniform random traffic pattern. The buffer occupancy that gives the best throughput is selected as the best threshold value.

Aim

To improve hierarchical NoC performance, by determining the buffer occupancy threshold that give the best network throughput.

Objectives

The following are the intended achievable objectives of this research work.

i. Develop a hierarchical, three layer NoC topology called Prism_Mesh

ii. Modify the XY routing algorithm to use a bridge node routing that link the layers.

iii. Run a sweep of buffer occupancy to the get the value that gives the best throughput for uniform random traffic pattern.

II. THEORETICAL BACKGROUND

This section reviews the literatures that are fundamental to this research. The first part, a review of fundamental concepts, introduces NoC by discussing the necessity of its emergence, the performance requirements, components and its associated algorithms. The second part is a discussion of the review of similar works in relation to the context of this research.

Review of Fundamental Concepts

Before the emergence of Network on Chip (NoC), data-buses were the means of onchip communication. They provided communication busses that convey data packets across Processing Element (PE), but as the number of PE blocks increased, so also does the complexity of the data-bus connections [8]. This led to the problems of scalability: Data-bus is poorly scalable. Another inherent problem of data-bus is predictability [9]. All the PE blocks share the same data-bus for communication, which led to the possibility of data collision along the bus when two or more IP emit to the bus simultaneously. Even with collision detection techniques added, collision is still possible as the number of PE increases. Secondly, the electrical characteristics of data-bus such as parasitic capacitance and cross-talks to adjacent wires are difficult to predict in the early design stages and may subsequently lead to data corruption in the final run [10],[11]. The combine effect of poor scalability and predictability make the data-bus unreliable for communication-intensive applications.

In an effort to address the above problems,[11],[8], proposed a complete replacement of data-bus with NoC. NoC provides the advantages of structure, performance and modularity. Structural wise, NoC is scalable and provides larger bandwidth, high propagation velocity and significantly lower power dissipation than the data-bus. In terms performance, a simplified link layout gives a more controlled electrical parameters leading to predictive behaviours. Therefore, packets corruptions due to unpredictability are eliminated. In effect, NoC provides replacement with better scalability, predictability and reliability.

1) NoC performance requirements

NoC performance is measured against four metrics: Latency, Throughput, Packet Loss and Power consumption, which are now explained briefly [7].

Latency: This is the required time needed to transfer specified amount of data from source to destination. The design consideration, in NoC as in other communication systems is to make this value as low as possible [10].

Throughput: This is the actual number of received packets by the destinations, coming from a source, per unit time interval. Higher throughput is the ultimate design requirement because it translate to more number of packets reaching their destination.

Packet Loss: This refers to a situation when some packets are discarded before they could reach their destination, usually due to error, or due to abnormal delay in the route. This is an undesirable property because, for some systems, retransmission is required, which implies double effort; and a bad retransmission policy can upset the network balance and create congestion.

Power Consumption: NoC are usually design to meet stringent power requirement because of their on-chip manufacturing characteristics and the lower the power consumption the better the performance.

These performance requirements are key elements in the choice of other NoC design components. They inform the decisions on the router design, the choice of flow control, routing algorithm and topology.

2) NoC Router

The most important component used to transfer data between communication nodes is the Router. It is a device that receives incoming data packets, checks their destinations and decides the best paths to forward to, towards their destination. It is the backbone of communication in NoC and the efficiency with which it is designed determines the performance of the Networking system [10],[12]. It has the following components: Input/output Ports, Virtual Channels, Arbitration Unit, Address Decoder and Cross-Bar Switch.

I/O Ports: A 2D Mesh topology router consists of five input and five output ports along five directions, namely: East, West, North, South and the local port connecting to a processing core. This local port takes care of the injection and reception of packets from and to the network[13].

Memory Buffers: To each input port of a router is attached a set of memory buffers for temporary storage of data packets during period of traffic congestion. The number and size of the buffers per port is arbitrary decided during the design stage. It may range from one buffer to be shared by all input ports, to four buffers per each input port. When the number of buffers per port exceed one, the buffer is then referred to as Virtual Channel and can be used for deadlock avoidance. The former requires a carefully design sharing scheme and is suitable when congestion estimate is

expected to be minimal, while the later, even though has no sharing overhead, can handle heavier congestion estimates [14].

Arbitration Unit: Flow control policies are executed by an arbitration unit known as the Arbiter, which defines the approach used in moving data packets through the NoC [11]. The arbitration unit resolves conflict when multiple input sources are requesting for a single output port.

Cross-Bar: The cross-bar (X-bar) is a multiple input, multiple output switch that connects specified input port to a specified output. When a new packet is received at the input port, an address decoder decodes the destination address and the output port to pass the packet to, towards the next routing node. The information is pass to the port controller, which communicate with the arbiter for contention, once there is no contention, the port controller signals the x-bar to link the input port to the output port for onward routing of the packet [11].

Other control signals establish handshake with neighbouring routers for communication control purposes. Such as the acknowledgement of a received packet, initiation and termination of transmission or sending credit signal to upstream routers.

3) Network flow control

Flow control is a scheme that controls when packets move and how they move through the network. It decides when and the amount of data a sender can transmit without overloading the receiver [15]. It is analogous to automobile traffic control light that allow or block passage at junction points. If the sending node produces data faster than the receiving node can process, the receiver will be overloaded with data. To regulate the flow of data, the receiver needs to send some feedback to the source to inform the latter that it is overloaded. Network flow control is also referred to as routing or switching mode. Several methods of flow control are used in NoC as shown below [14].

Store-and-Forward routing mode: Data Packets move through the network in a store and forward manner. At each router, the router's buffer stores the entire packet, before it is delivered to the succeeding router. In effect, the buffer's capacity must be able to contain the largest packet in the network. Latency is calculated as the summation of time to receive a packet and time to forward it to the next hop through the hop length of the packet path.

Virtual Cut-Through routing mode: The difference between this mode and the store-and-forward mode is in the condition to start forwarding the received packet. This mode doesn't require the whole packet to be received before transmission will commence; it only requires the receiving router's permission to receive data and forwarding will commence. Therefore, it has low latency than the store-and-forward mode.

Wormhole routing mode: In this mode, data packets are broken-down into smaller equally-sized flow control units called flits. Virtual cut-through routing is performed for only the first flow control unit. Thereafter, the path is reserved for the other flits to follow directly without passing through the router's buffer at all. It is clear that Wormhole uses less buffer space because only the first flit is stored and therefore gives lower latency. It major drawback is a high possibility of deadlock occurrence and head on line blocking [16].

4) Routing

Routing is the logic that directs the movement of packets within a network. It selects one output from a set of output ports, in a node, to forward a received packet. The selected output is based on the routing information available in the packet header. Overtime, several routing algorithms were used in NoC, each one leading to different trade-off between performance and cost. What follows is a brief discussion of the major types of routing algorithms [17].

Oblivious routing algorithms: This is ignorant about information that describes network conditions, such as traffic amount and congestion situation. A router only routes packets based on the grounds of some algorithm. The three mojor variants are the dimension order (DOR), turn-model and deterministic routing algorithms. DOR is based on the use of a pre-set packets direction of flow, an example is the XY routing, which routes through the x-direction before turning to y-direction. Turn-model is based on using a selective set of turns while prohibiting other turns. This is to prevent forming a routing circle, which is a requirement for deadlock avoidance [18]. Example is the West First routing algorithm. Deterministic routing uses circuit switching techniques to routes packets through a pre-determined routing path, usually based on minimal hop distance to destination. Example is the distance vector routing algorithm.

Adaptive routing: These algorithms make routing decisions dynamically, always considering the present state of the network. It routes packets by making use of the dynamic congestion estimates of various paths in the network. It dynamically changes path, to avoid region experiencing congestion as it journeyed towards its destination based on the pre-decided selection strategy in it design. It effectiveness is determined by the metric that is used to capture congestion indicators [19]. The congestion indicators are: The number of free virtual channels, buffer occupancy, crossbar requests and link utilization.

In adaptive routing, the ability to select one route from multiple available routes is a critical factor in its performance determination. Output selection is the process of using a suitably designed output selection function to select one output port from multiple available ports based on the congestion metrics. Similar to this, there can also be a scenario in a router when multiple flits compete for the same input port concurrently, in which case input selection is done based on parameters like age of the flit or upstream node congestion [19]. Moreover, based on the depth of congestion information, adaptive routing schemes can be either local or global to which global adaptive schemes show better performance [20].

5) Congestion control

Congestion in computer networking refers to the decrease in performance that occurs due to delay, or temporal blockage, to the flow of packets in some region of a network. This results if a node has more packets inflow than it can processed, or when the bandwidth of a link is

below the volume of data injected into it. The reduction in performance may be observed as a rise in queuing delay, rise in amount of discarded packets and blocking of incoming traffic into the affected region. Furthermore, a small increase in load usually decreases performance such as; reduced network throughput, increase latency or the overall failure of a network system [21].

6) Topology

NoC topology is the medium through which the routing unit, routing algorithm and flow control mechanism are implemented and has a significant effect on the NoC performance. Topology is the structural layout of the components of NoC that can be described by a graph consisting of nodes and links. Essentially, the topology consists of five components: The PE core, the Network Interface (NI), the routing node and the links (internal and external links).

PEs are the points of processing tasks. They also form the source and sink of packets that traverse the NoC. PE may be any of the followings: Processing cores, DSP, GPU, RAM memories and peripheral controllers.

Network Interface (NI), as the name implies interfaces the PE block with the routing node. It is the point in which the message to be communicated is broken into packets sizes with each packet having a header, a payload and a tail. Error checking mechanism is implemented, source and destination addresses are added and finally, flow control scheme is incorporated.

Routing Nodes: Essentially, the routing node consists of the router through which suitable routing algorithm is implemented.

Links are of two types: Externals links are bi-directional data channels that connects the PE to routing nodes and internal links are uni-directional data channels that connect any two routing nodes.

7) 2D Mesh topology

The most widely used NoC topology is the 2D mesh due to its simplicity of design, scalability and easier chip fabrication. Nonetheless, it suffers from low path diversity compared to 3D mesh [13]. In addition, the network hop count increases linearly with increased network size. This limitation necessitates the need for low radix NoC. Figure 1 and 2 below give graphical representation of the 2d mesh concepts.

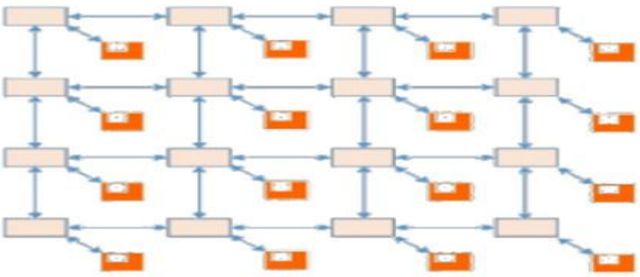


Figure 2: A 4x4 2D Mesh topology.

Low radix networks may come as ring, torus and hierarchical topologies. Ring topology is a linear arrangement to which the last node is looped back to the first. Torus is a 2d mesh to which each row forms a ring and each column also forms a ring.

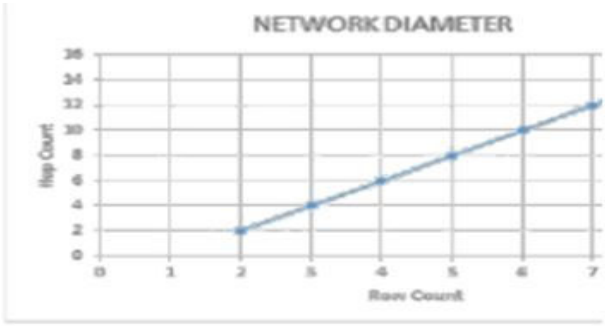


Figure 3: Size/distance relationship of an 8x8 mesh topology.

8) Hierarchical mesh topology

Hierarchical mesh topology consists of two or more mesh topologies that are connected in hierarchies. The mesh structures used, may be of the same or of different sizes. For a given number of routing nodes, differently sized hierarchical topologies may be constructed using layer counts and clusters size as structural parameters.

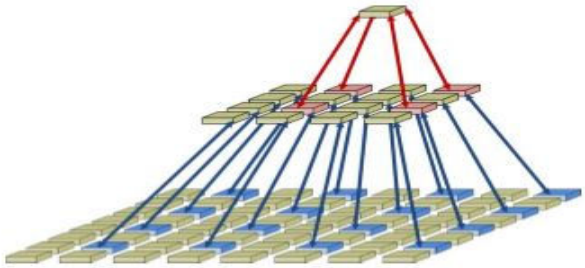


Figure 4: A three layer hierarchical topology: Prism_Mesh [22].

Hierarchical topologies offer the benefits of reduced end-to-end hop counts and large path diversity. More so, packets are divided into local and global traffics during routing decisions. The major limitations of this topology is the extra overhead in the number of routing nodes that serve only as routing switches. In addition, it suffers from traffic load imbalance across the layers. The attached figures above show sample of hierarchical NoC topologies. Figure 3: above is a prism_mesh topology; the focus of this research work.

III. METHODOLOGY

The steps used to achieve the objectives of the research are; building the topology, development of a modified XY routing algorithm to handle the bridge link route and deciding on the traffic that will use the bridge link. These steps are elaborated below.

A. Building the topology

1) Build the layers

Structurally, the prism_mesh topology consists of a 2d mesh base layer followed by one or more 2d hierarchical layers of lower sizes that decrement in size as one ascends the hierarchy structure. This gives it a pyramidal resemblance. The layers are connected through bridge nodes and links. The prism_mesh topology can be constructed with the given structural parameters in table 1.

Table 1: Prism_mesh Structural Information.

LEVEL	NAME	SIZE	NO OF NODES	NO OF PE
0	Base_layer	8x8	64	64
1	Layer_1	4x4	16	Nil
2	Layer_2	2x2	4	Nil

The base layer is calculated to have 64 routers and therefore, 64 processing elements. Using the cluster size of 2, the first upper layer is calculated to have 16 nodes (4x4) and the second layer is calculated to have 4 nodes (2x2). Note that the first and second upper layers have no processing elements and serve only as routing switches.

2) Links the layers.

Links between the layers are established as follows. Starting with row(0) and column(0), up-links are set for each odd row (row(odd)) and odd column (column(odd)). Starting at row(0) and column(0) in the layer above, a down-link are set for each row(i) and column(i). The first up link in the bottom layer is connected to the first down link in the upper layer, repeatedly to the last node. The above procedure is repeated for the next hierarchy (first upper layer and second upper layer), until the topmost layer is linked, if more than three.

B. Develop a routing algorithm

A routing algorithm is developed, for the prism_mesh topology as follows. By considering that all the layers are planar structures that can use the normal XY routing, an extension to the third dimension, that include vertical links is added, which represent the bridge links. Global traffic at the bridge link are routed up through the link, else they routed using XY routing. Non global traffic at layer 1 and layer 2 are routed down through the links if they are within the destination cluster otherwise they are routed using XY routing. The modified XY routing algorithm will henceforth be referred to as TBR algorithm.

C. Determine the routing parameters

Calculate Manhattan distance: Using the source and destination address contained in the packets header, calculate the Manhattan distance M_D. Manhattan distance is the number of hop counts from source node to destination node, going through each node along the chosen path. When a packet is emitted into it source node, the source node calculates M_D with the following expression.

$$M_D = |X_d - X_s| + |Y_d - Y_s| \quad (1)$$

where X_d and Y_d are the horizontal and vertical coordinates of the the destination node, while X_s and Y_s are the horizontal and vertical coordinates of the source node.

3) Decides the traffic type

Use relation (2) below to decides if the packet is a global or local traffic. Once T_R is global at a bridge link use TBR algorithm, otherwise XY routing is used at the local level.

$$T_R = \begin{cases} global\ if\ M_D > log N_n + 2 \\ local\ otherwise \end{cases} \quad (2)$$

4) Obtain the buffer occupancy B_U

Buffer occupancy (B_U) is one of the metrics used to determine the congestion situation at a node. It is the the ratio of the occupied buffer space to the total buffer space. The TBR algorithm will read this value at the bridge node. Equation 5.4 below is the mathematical expression. It ranges from 0.0 to 1.0.

$$B_U = \frac{Occupied\ buffer\ space}{Total\ buffer\ space} \quad (3)$$

5) Set a threshold for the bridge link

Run a sweep of threshold from 0.0 to 0.25 at interval of 0.1. A plot of injection rate and throughput is done and the threshold that gives the highest throughput is selected, to be used in subsequent simulation.

6) Apply the Parameters into the TBR algorithm.

The parameters obtained above are necessary towards making a routing decision as is shown in equation (4) below. The equation will be coded into the TBR routing algorithm so that both local and global traffics are appropriately routed.

$$R_D = \begin{cases} Route\ through\ bridge\ link\ if\ B_U \leq B_T \leq 1 \\ Use\ alternate\ route\ otherwise \end{cases} \quad (4)$$

IV. IMPLEMENTATION

The methodology described above is implemented using the NoC simulator Garnet2.0 in Gem5 [23]. To test the constructed Prism_Mesh topology, a comparative plot with the inbuilt Mesh_XY topology of the simulator is done using injection rate and throughput as parameters. The plot is in the figure 4 below.

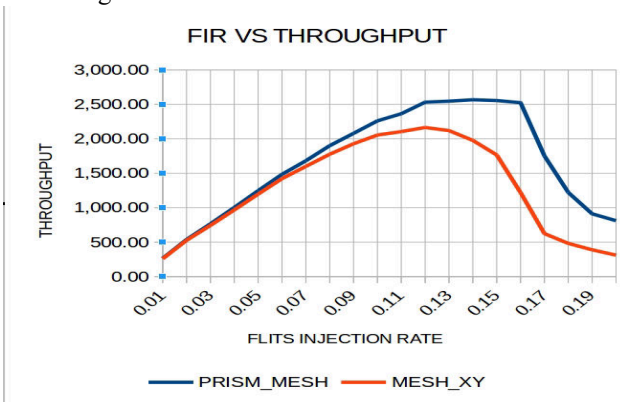


Figure 5: A comparative plot of injection rate against throughput for Prism_Mesh and MeshXY topologies.

Two things are observed from the plot:

1) The Prism_Mesh topology, which has a peak through of about 2,566.72 at 0.14 fir, outperforms the Mesh_XY topology, which has a peak throughput of 2,164.28 at 0.12 fir.

2) Saturation for Prism_Mesh topology starts at about 0.12 injection rate from where a sharp decrease in throughput rise is noticed, while saturation for Mesh_XY starts at about 0.10 where a sharp decrease in through rise is also noticed.

The Prism_Mesh topology can now be subjected to further investigation. To evaluate the Prism_mesh with injection rate far below saturation will yields insignificant results because it is running below optimal capacity. Similarly, evaluating the Prism_Mesh far beyond saturation,

can lead to unpredictable results because the network is being pushed far past its capacity. Instead, it is best to evaluate the topology at a point just below the saturation level and to gradually move deep within the saturation region upto about 0.20 fir, so that we can see the pattern of behaviour for each curve. It is also important to note that, the buffer depth is 4 flits; that is, the buffer capacity can only have one of five states: empty, ¼, ½, ¾, and full capacity. Using this range as the threshold values, a plot against throughput for various injection rates is given below.

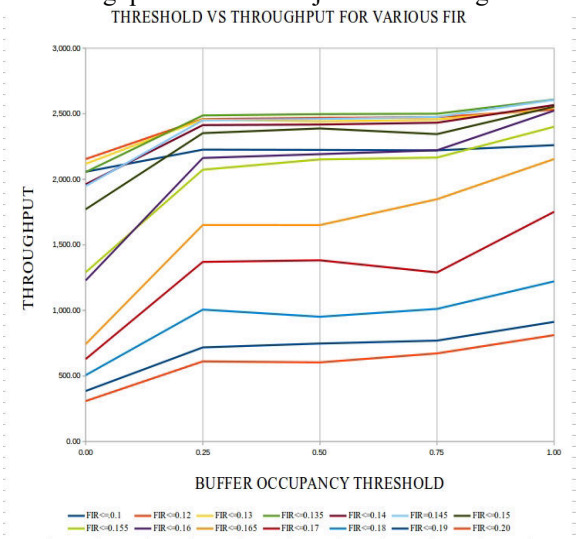


Figure 6: A plot of Threshold against throughput at various injection rates.

The plot above shows some interesting patterns that need further investigation. Threshold value of 0.00 always has the minimum throughput value and threshold value of 1.00 always has the highest throughput value. The threshold values of 0.25 and 0.50 have little variation from each other across the chosen fir. Threshold value of 0.75 have the most remarkable variation for fir of 0.165 and 0.170, in which it value falls below those of 0.25 and 0.50 for 0.170 fir and rose above 0.25 and 0.50 for 0.165 fir.

Injection rate values from 0.16 to 0.20 are deep within the saturation region as they carry more than 60 percent of the zero-load latency values and we practically don't want to operate in that region. We will therefore concentrate on injection rate values in the range 0.10 to 0.15, a region of saturation that is of more practical importance.

A rescaling of the above plot to highlights this range is given below.

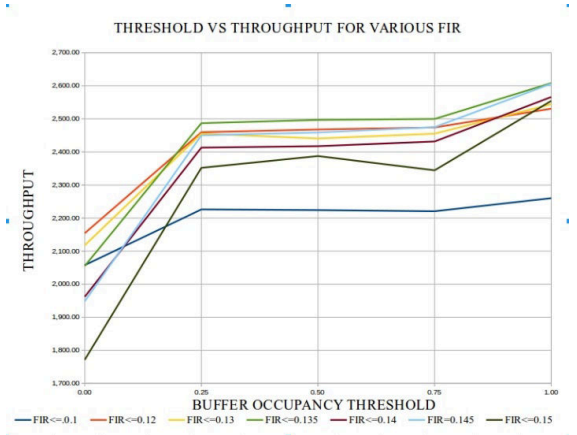


Figure 7: A rescaled threshold vs throughput plot to highlight fir range of 0.1 to 0.15

The rescaled figure shows that below saturation at 0.10 fir, the threshold with the maximum throughput is 1.00 and that with the minimum throughput is 0.00. Threshold settings of 0.25 to 0.75 have insignificant variations. At light saturation region (<40% zero-load latency) the best threshold is 1.00 followed 0.50, except for fir 0.13 where 0.25 and 0.75 give better reading that 0.50. The worst threshold reading is always 0.00.

The best curve is the green plot which corresponds to injection rate of 0.135. It falls behind red and yellow curves only at 0.0 threshold setting. The worst curve is the dark-green curve with fir 0.15, which reading is only better than the yellow and red curve at threshold setting of 1.00.

At threshold setting of 1.00, the throughput reading are in the following order: green (0135 fir), sky-blue (0.145 fir), maroon (0.140 fir), dark-green (0.15 fir), yellow (0.13 fir) curves and then the red (0.12 fir) curve

At threshold setting of 0.75, the throughput reading are as follows: green is best, sky-blue and red coincides, then yellow, followed by marron, and dark-green is least. At threshold setting of 0.5, the following throughput is observed in descending order: green, red, sky-blue, yellow marron and dark-green. At threshold setting of 0.25, the following throughput is also observed decrementing order: green, red and yellow sky-blue, maroon, and dark-green. At threshold setting 0.0, which gives worst throughput output for each fir, the following order is observed decrementally: red, yellow, green, marron sky-blue, and dark-green.

Why is threshold value 1.00 given the highest throughput and threshold value 0.00 given the lowest throughput reading? The possible explanation is that: Congestion starts developing at the base layer when saturation is reached, therefore, any routing through the bridge link relieves that particular node of congestion. The threshold setting of 1.00 means that at any buffer occupancy level, routing at the bridge link towards the next layer always mitigates the congestion level in the base layer thus improving the throughput.

This will always hold true if congestion at the upper is less than that at the lower layer, which can be determine by considering two factors. One, routing nodes at the upper layers are not connected to PE and therefore, packets only come from the lower layers. Two, the percentage of global traffic in the lower layers does not dominate that of the local traffic.

For the threshold setting of 0.0, this means that routing to the upper layer is done only when the buffer occupancy is zero. The number of routers with empty buffers can only be a small fraction of the total number of routers, hence impacting negligibly on the congestion level which leads to very low throughput output.

Another situation is the anomalous decrease in throughput at 0.50 and 0.75 threshold setting for the dark-green and yellow curves. It is expected that, the throughput at 0.50 threshold would be better than those at 0.25 threshold setting and those at 0.75 threshold would be better than those at 0.50 threshold setting. The green, red and sky-blue curves shows that compliance, but the yellow (0.25 has better throughput than 0.5and 0.75 threshold

setting), and dark-green (0.50 has better throughput than 0.75 threshold setting) clearly shows a contradiction.

The explanation to the observation above is not obvious and outside the set objectives of this research, but will be a subject of further investigation.

V. CONCLUSION

This research work looked at ways of improving NoC performance by using a traffic load distribution scheme across hierarchical layers in network on Chip. The scheme used buffer occupancy threshold to decides when to route traffic between the network layers. A sweep of buffer occupancy set as threshold at step interval of 0.25 across the range 0.0 to 1.0 was ran and the value that gave the best throughput was identified at 1.00 and chosen as the best value. The result indicates that routing packets through the bridge link always mitigates the congestion at the base layer.

ACKNOWLEDGMENT

I extend my thanks and appreciation to the following, all of the Department of Electrical and Computer Engineering, ABU, Zaria, Nigeria: All members of the Academic Staff for their mentoring and instruction in the course of my studies. PG Library staff and other Non-Academic Staff for helping in my access to study materials. My Course Mates for their cooperation and encouragement during difficult moments in the course of this work and the programme of studies. Thank you all.

REFERENCE

[1] [1] A. Ben Ahmed and A. Ben Abdallah, “Architecture and design of high-throughput, low-latency, and fault-tolerant routing algorithm for 3D-network-on-chip (3D-NoC),” *J Supercomput*, vol. 66, no. 3, pp. 1507–1532, Dec. 2013.

[2] [2] M. Bai, D. Zhao, and H. Wu, “CATBR-Congestion Aware Traffic Bridging Routing among hierarchical networks-on-chip,” in *2016 29th IEEE International System-on-Chip Conference (SOCC)*, Seattle, WA, USA, 2016, pp. 52–57.

[3] [3] R. Manevich, I. Cidon, and A. Kolodny, “Handling global traffic in future CMP NoCs,” in *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, 2012, pp. 1–5.

[4] [4] R. Manevich, I. Cidon, and A. Kolodny, “Design and dynamic management of hierarchical NoCs,” *Microprocessors and Microsystems*, vol. 40, pp. 154–166, Feb. 2016.

[5] [5] J. Kemmerer and B. Taskin, “Range-based Dynamic Routing of Hierarchical On Chip Network Traffic,” in *Proceedings of SLIP (System Level Interconnect Prediction) on System Level Interconnect Prediction Workshop - SLIP ’14*, San Francisco, CA, USA, 2014, pp. 1–9.

[6] [6] J. Balfour and W. J. Dally, “Design tradeoffs for tiled CMP on-chip networks,” in *25th Anniversary International Conference on Supercomputing Anniversary Volume* -, Munich, Germany, 2014, pp. 390–401.

[7] [7] A. More, “Network-on-Chip (NoC) Architectures for Exa-scale Chip-Multi-Processors (CMPs),” Drexel, Philadelphia, 2013.

[8] [8] W. James. Dally and B. Towles, “Route packets, not wires: on-chip interconnection networks,” in *Proceedings of the 38th Design Automation Conference (IEEE Cat. No.01CH37232)*, Las Vegas, NV, USA, 2001, pp. 684–689.

[9] [9] H. J. Mahanta, A. Biswas, and Md. A. Hussain, “Networks on Chip: The New Trend of On-Chip Interconnection,” in *2014 Fourth International Conference on Communication Systems and Network Technologies*, Bhopal, India, 2014, pp. 1050–1053.

[10] [10] L. Benini and G. De Micheli, *Networks on chips: technology and tools*. Amsterdam ; Boston: Elsevier Morgan Kaufmann Publishers, 2006.

[11] [11] W. James. Dally and B. Towles, *Principles and practices of interconnection networks*. Amsterdam ; San Francisco: Morgan Kaufmann Publishers, 2004.

[12] [12] N. E. Jerger and L.-S. Peh, “On-Chip Networks,” *Synthesis Lectures on Computer Architecture*, vol. 4, no. 1, pp. 1–141, Jan. 2009.

[13] [13] N. E. Jerger, T. Krishna, and L.-S. Peh, “On-Chip Networks, Second Edition,” *Synthesis Lectures on Computer Architecture*, vol. 12, no. 3, pp. 1–210, Jun. 2017.

[14] [14] A. Aldammas, A. Soudani, and A. Al-Dhelaan, “The efficiency of buffer and buffer-less data-flow control schemes for congestion avoidance in Networks on Chip,” *Journal of King Saud University - Computer and Information Sciences*, vol. 28, no. 2, pp. 184–198, Apr. 2016.

[15] [15] U. Y. Ogras and R. Marculescu, *Modeling, Analysis and Optimization of Network-on-Chip Communication Architectures*, vol. 184. Dordrecht: Springer Netherlands, 2013.

[16] [16] E. Rijpkema, K. Goossens, and P. Wielage, “A Router Architecture for Networks on Silicon,” in *In Proceedings of Progress 2001, 2nd Workshop on Embedded Systems*, 2001, pp. 181–188.

[17] [17] M. S. Obaidat, P. Nicopolitidis, and F. Zarai, Eds., *Modeling and simulation of computer networks and systems: methodologies and applications*. Waltham, MA: Morgan Kaufmann, 2015.

[18] [18] E. Fragkakis, “Deadlock avoidance with virtual channels,” Universitat Autònoma de Barcelona., Spain, 2009.

[19] [19] R. Oommen and J. Jose, “Congestion management in adaptive NoC routers using cost-effective selection strategies,” *HiPC-SRS: IEEE*, p. 5, 2013.

[20] [20] M. Debnath, D. Konstantinou, C. Nicopoulos, G. Dimitrakopoulos, W.-M. Lin, and J. Lee, “Low-cost congestion management in networks-on-chip using edge and in-network traffic throttling,” in *Proceedings of the 2nd International Workshop on Advanced Interconnect Solutions and Technologies for Emerging Computing Systems - AISTECS ’17*, Stockholm, Sweden, 2017, pp. 8–11.

[21] [21] T. Kogel, R. Leupers, and H. Meyr, *Integrated system-level modeling of network-on-chip enabled multi-processor platforms*. Dordrecht: Springer, 2006.

[22] [22] J. Kemmerer and B. Taskin, “Dynamic Routing of Hierarchical On Chip Network Traffic,” p. 30, 2014.

[23] [23] N. Agarwal, T. Krishna, L.-S. Peh, and N. K. Jha, “GARNET: A detailed on-chip network model inside a full-system simulator,” in *2009 IEEE International Symposium on Performance Analysis of Systems and Software*, Boston, MA, USA, 2009, pp. 33–42.

PMU Optimal Placement in Smart Grid using WLAV Power System Estimator

Mohammad Shoaib Shahriar[1*], Yusuf Abubakar Sha'aban[1,2], Ibrahim Omar Habiballah[3], Farhan Ammar Ahmad[4],
[1]Department of Electrical Engineering, University of Hafr Al Batin, Saudi Arabia
[2]Department of Computer Engineering, Ahmadu Bello University, Zaria, Nigeria
[3]Department of Electrical Engineering, King Fahd University of Petroleum & Minerals Dhahran, Saudi Arabia
[4]Department of Electrical Engineering, University of Management and Technology Lahore, Sialkot Campus, Pakistan

Abstract - **The monitoring, observation and security of a modern-day smart grid system is highly dependent on the accuracy of the selected state estimator. State estimation provides accurate states of a power system in a real-time frame. The conventional meters, which are mostly installed in current smart grid systems are prone to provide wrong measurements to the estimator, which reduces the efficiency of the state estimation results. Over the last two decades, new and reliable meters have been introduced to electric utilities. For example, phasor measurement units (PMU), which are devices that can provide voltage and current phasors with a very high accuracy are readily available. However, they are expensive and need to be optimally placed on the smart grids. This paper presents a technique for optimizing the placement of PMUs, with the objective of improving the estimation performance using Genetic Algorithm (GA). The weighted least absolute value (WLAV) estimator, which is one of the well-established robust estimators, is being selected for simulation. The IEEE-14 and 30-bus power systems are chosen to demonstrate the effectiveness of proposed optimal PMU placement.**

Keywords — Genetic Algorithm, Optimal PMU Placement, Phasor Measurement Unit, State Estimation, Weighted Least Absolute Value.

I. INTRODUCTION

Traditionally, static state estimation based upon measurements provided by supervisory control and data acquisition (SCADA) system are used for planning and control features in the utilities. A major drawback of such an arrangement is that the system is not capable of capturing measurements at a very fast rate for dynamic operation. Another disadvantage is the inability of the SCADA measurements to provide the phase angle of bus voltages or line currents. However, phasor measurement units (PMUs) can provide instantaneous voltage and current phasors at very high refresh rates and are becoming very popular in modern electric power systems. Typically, a PMU measurement achieves a magnitude accuracy of 0.1% and synchronization precision of one microsecond when compared with traditional SCADA measurement [1], [2].

In the future, more PMUs will be required to enhance the monitoring, control, operation, and protection of a power system. These PMUs are generally more expensive when compared with traditional SCADA systems. This makes it impractical to place PMUs at all or most of the bus locations in a power system. A practical solution is to place a few PMUs at appropriate bus locations such that the entire power system is completely observable. This approach is known as the optimal PMUs placement and has been considered by several researchers [3]–[6] where the authors presented comprehensive studies on the optimal placement of PMUs along with the objective functions and constraints for the respective optimization problems.

In the electric power industry, the cost associated with state estimation with only PMUs, placed on every bus is prohibitive. Therefore, using the measurements from both SCADA and PMU measurements simultaneously is a more practical approach in enhancing the accuracy of state estimation. In [7]–[10], some approaches, that combine both SCADA and PMUs measurements with the theory, mathematical formulation and computer simulations, are presented. In [11], an incremental PMU placement technique was proposed and tested to mitigate state estimation uncertainties associated with conventional SCADA measurements. In [12], a study was carried out to compare different approaches for employing PMU in state estimation. However, the work was done only based on the voltage phasors of PMU. In [13], a heuristic PMU placement approach for enhancing the estimation accuracy was proposed for limited number of PMUs. However, the technique is prone to be trapped in a local minima during PMU location selection. In [14], genetic algorithm (GA) was used to get the optimal number of PMUs for the Weighted Least Square (WLS) installed power system only. However, despite the above efforts, the power systems with WLAV installed estimators need an approach of optimizing the PMU locations.

This paper presents an optimization technique to select the optimal locations to place PMUs in a smart grid, which will ensure the least estimation error. The proposed technique could be an effective solution for the sequential replacement of conventional measuring meters by the highly efficient PMU meters. It can provide the optimal set of any desired number of PMUs to be placed in the power system. One of the mostly used heuristic optimization techniques, genetic algorithm (GA) is used to solve the optimization problem. One of the widely used robust estimators, Weighted Least Absolute Value (WLAV) is also used to ascertain the accuracy of estimation. The proposed PMU placement was implemented on two IEEE power systems with 14-bus and 30-bus.

II. Modified Weighted Least Absolute Value (WLAV) with Phasor Quantities

Linear programming approaches such as the simplex and interior point methods can be used to solve the WLAV estimation problem. Studies have shown that while WLAV fails in the presence of leaverage points, it gives improved estimation accuracy in the precense of Bad-data [15].

Consider a power system network that has n buses. If there are m meter locations, then the measurement vector has a size $(m \times 1)$ while the state vector has a dimension $(n \times 1)$. Let the measurement error be $\in_i$ also with dimension $(m \times 1)$ and $h_i(x)$ with dimension $(m \times 1)$ be the nonlinear state equation. Then,

$$z_i = h_i(x) + \in_i, \qquad i = 1 \cdots m \tag{1}$$

The objective function minimized by WLAV is given as [16], [17]:

$$f(x) = \sum_{i=1}^{m} W_i \left| h_i(x) - z_i \right| = \sum_{i=1}^{m} W_i \left| \in_i \right| \tag{2}$$

Where, $k = 1,...n$, W_i represents i^{th} measurement error variance reciprocal.

$$W_i = \frac{1}{\sigma_i}, \qquad W = R^{-1} \tag{3}$$

The absolute error $\left(\left|\in_i\right|\right)$ should be less than a threshold value, ε_i such that $\left|\in_i\right| \le \varepsilon_i$.

A PMU can accurately measure voltage and current phasors of all buses and branches connected to it. After a PMU is installed at a bus, the phasor measurements need to be added in the measurement vector and hence, the measurement Jacobian matrix should be modified accordingly. The mathematical formulation of WLAV estimator has been modified in this paper to execute the state estimation accurately with both SCADA and PMU phasor measurements. The modified Jacobian matrix, H is given in (4) [14].

The Jacobian matrix in (4) contains the elements with partial derivatives of power flows (P_{flow}, Q_{flow}) in branches, power injections (P_{inj}, Q_{inj}) at buses, current phasors ($IF_{,Real}$, $IF_{,imag}$) and voltage magnitudes (V) with respect to system's state variables: voltage magnitude (V) and phase angle (θ).

The following assumptions are made during the phasor measurement incorporation in the state estimation [14]:

- To avoid time skew, the SCADA and PMUs phasors where measured concurrently.
- A PMU must be located at a slack bus and it should always be available to ensure same reference angle for SCADA and PMU measurements.
- After a PMU has been placed on a specified bus, it can accurately measure the voltage phasor and current phasors of the bus and all connected branches respectively.

$$H = \begin{bmatrix} \delta P_{inj}/\delta\theta & \delta P_{inj}/\delta V \\ \delta P_{flow}/\delta\theta & \delta P_{flow}/\delta V \\ \delta Q_{inj}/\delta\theta & \delta Q_{inj}/\delta V \\ \delta Q_{flow}/\delta\theta & \delta Q_{flow}/\delta V \\ \delta V/\delta\theta & \delta V/\delta V \\ \frac{\partial I_{F,Real[ij]}}{\partial\theta} & \frac{\partial I_{F,Real[ij]}}{\partial Vm} \\ \frac{\partial I_{F,Imag[ij]}}{\partial\theta} & \frac{\partial I_{F,Imag[ij]}}{\partial Vm} \end{bmatrix} \tag{4}$$

III. Optimization with Genetic Algorithm

The goal of the developed optimization problem is to find the optimal locations for any defined number of PMUs to ensure optimum estimation. Therefore, the problem formulation could be stated as follows:

Let the set for available bus locations is denoted as X. Therefore, $(x_1, x_2, ...x_n) \in X$ and n is the largest bus number. Where, X_{PMU} is the set of bus locations installed with PMUs in the SCADA-based power system. X_{opt} is the set of optimal PMU locations. The installation of PMUs in X_{opt} locations will ensure the maximum estimation efficiency. x_{opt} is an individual bus location in the optimal solution set, X_{opt}. The proposed optimization problem uses GA to find out X_{opt}. A_i and E_i denote the true and estimated state variables respectively. The absolute difference between them reflects the error of the estimation process. r is the reference/slack bus where s represents interesting buses i.e. those associated with important measurements. Z represents the measurement vector used by the state estimator. Z_{PMU} contains the measurements which are coming from PMUs.

The objective function and associated constraints are as follows:

$$Min \ \left(\frac{1}{2 \times n} \sum_{i=1}^{n} \sum_{X_{opt} = X_{PMU}}^{X_{opt} \subset X} \sum_{Z_{PMU} \subset Z} \left(\left|A_i - E_i\right|\right)\right)$$

$$s.t. \ r \in X_{opt}$$

$$x_1 \le x_{opt} \le x_n$$

$$s \in X_{opt}$$

The optimization problem searches for the X_{opt} set by minimizing the value of $\left|A_i - E_i\right|$.

The optimal solution X_{opt} is a subset of X and contains the same elements as X_{PMU} to ensure that the buses to be installed with PMUs are the optimal solutions. The constraints ensure that the slack bus (s) and the buses (r)

which covers the critical zones will be the element of optimal set, X_{opt}. The PMU installed in the slack bus will ensure the accuracy of the phase angle of the reference, which will be compared with the remaining bus-angles. Covering the critical locations by PMUs will ensure the least effect of bad-data on the performance of the state estimation.

The optimization process ensures that the PMU measurements (Z_{PMU}) are included in the overall measurement series (Z) during estimation process. One of the significances of the proposed optimization method is that, the error indicator is applicable for measuring the accuracy of any estimator even though different estimators have different objective functions and problem formulations. Therefore, the optimization problem could be implemented and tested with any estimation algorithm without modifying the problem formulation and solution strategy.

Figure 3 is a flow chart which shows how Binary Genetic Algorithm (BGA) works to reach the optimal solution. It starts with the selection of initial values appropriately and setting up the proper elitism. Then the set of initial population is generated and used for operations like selection, crossover and mutation in each iteration. This is repeated until convergence to a solution, having minimum NCE, which is considered the best.

IV. TEST CASE PREPARATION

The developed approach was applied to the IEEE 14 and IEEE 30 bus system, in a simulated environment, in the presence of bad-data. The parameters for the considered scenarios are taken from the reference [17].

There are total 55 measurements in IEEE 14 bus system and 126 measurements in IEEE 30 bus system. The IEEE-14 and IEEE-30 bus system has global measurement redundancy level equal to 2.03 and 2.13 respectively. To execute any state estimator algorithm, a minimum of (2n-1) measurements need to be provided, which fulfills the observability condition. The requirement of observability is achieved by selecting the proper meter locations for the respective test cases. The accuracy of the traditional SCADA measurements is affected by random white noise during the metering process. To incorporate random noise in the measurements, Gaussian Noise was added. The respective parameters for the applied white noise were obtained from the true values of the load flow solution.

For the test case of the IEEE-30 bus, there are two critical measurements: PF_12-13 and QF_12-13. In applying the GA, bus number 12 served as the elite genes for the considered IEEE-30 bus system. Here, the critical zones were covered by PMUs. Optimal PMU locations are found in the SCADA-measurement based power system. An installed PMU, gives measurements such as the magnitude and phase of the bus and the resulting current through the bus [4-5]. Other additional phasor measurements for certain PMUs are presented in Table II.

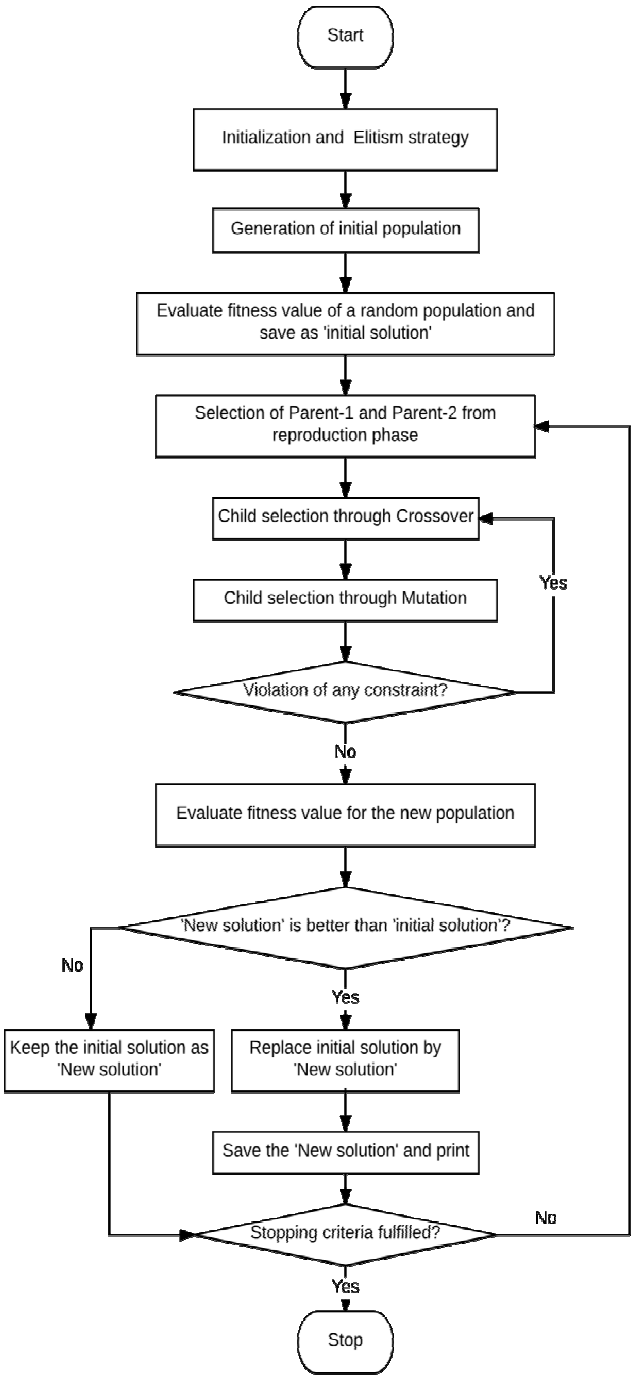


Figure 3. Flow diagram of BGA steps

TABLE II. ADDITIONAL PMU METERS LOCATION FOR OPTIMAL PMU SET 1,5,19 (THE IEEE-30 BUS SYSTEM)

Measurement Type	Measurement locations
Voltage-magnitude	1, 5, 19
Voltage-angle	1, 5, 19
Current flows	1-2, 1-3, 5-7, 19-20

The standard deviation of each measurement has to be chosen carefully to increase the accuracy of the measurement. The voltage magnitude measurement should have maximum weight as compared with other types of measurements during the estimation process and the components of the reactive power injection and flows should have the lowest weights. The

standard deviation of the real power flows and injections may have a value that is between the magnitude of the voltage and the reactive power [18]. The applied weight is computed by finding the inverse of the associated sigma value. This explains the reason why the selection criteria of sigma values, which has been kept same for all the test scenarios.

TABLE III. THE VALUES AND TOLERANCES OF SIGMA USED IN SIMULATION

Measurement Types	Sigma for WLAV
Vm	0.01
PG	0.02
QG	0.04
PF	0.02
QF	0.04

GA needs to select the appropriate initial values before going into operation. The size of the initial population, maximum iteration number, criteria for stopping the iterations etc. should be decided properly. Besides, the appropriate probabilities for the operations (selection, mutation, and crossover) need to be specified. Elitism is selected according to the constraints of the optimization problem in order to cover the slack bus and the critical zones. To stop the iteration of GA, it is required to either reach the maximum iteration threshold of 250, or the fitness value kept unchanged for consecutive 100 iterations.

V. SIMULATION RESULTS

The simulation results for both the 14-bus and 30-bus test cases were obtained. The result for the 14-bus system is presented in Table IV. The slack bus of the 14-bus system is bus-1 and the test case does not have any critical measurements. Therefore, bus-1 is present in all the solution sets. It is noticed that the error indicator, NCE kept decreasing with the increase in the number of the PMU. For example, NCE value dropped to 0.0305 from 0.0492 with the installation of 4 PMUs in the SCADA-based measurement system.

TABLE IV. OPP RESULTS FOR THE IEEE 14-BUS SYSTEM

Number of PMUs	Optimal PMU locations	NCE Value
Only SCADA meters	NA	4.92E-02
2	1, 12	4.12E-02
3	1, 6, 12	3.65E-02
4	1, 6, 9, 12	3.05E-02

The optimal locations of PMUs for the 30-bus IEEE power system is presented in Table V where bus 1 and bus 12 are present in all the solution sets because of the elitism strategy of the optimization. Again, the NCE indicator kept decreasing with the PMU installation. With the presence of 5 PMUs in the test case, NCE indicator becomes 0.0304 which was initially 0.0479 without the PMU meter. The high accuracy of PMU measurements clearly improved the estimation performance.

TABLE V. OPP RESULTS FOR THE IEEE 30-BUS SYSTEM

Number of PMUs	Optimal locations of PMUs	NCE Value
Only SCADA meters	NA	4.79E-02
2	1, 12	3.92E-02
3	1, 12, 18	3.56E-02
4	1, 6, 12, 25	3.23E-02
5	1, 12, 25, 28	3.04E-02

The covering of critical locations is considered as one of the constraint in this optimization problem. The purpose of this is to counter the effect of bad-data on the results of the estimation process. However, the presence of bad-data, if it occurs from or around a critical location, has a severe impact. Therefore, the PMU placement should cover such sensitive zones and ensure the robustness against such scenario. Table VI shows the effect of bad-data on estimation performance around the critical zone. It also shows that the covering of critical zone by highly accurate phasor measurements will help to cope with the presence of such bad-data.

TABLE VI. NCE INDICATOR FOR 30-BUS SYSTEM WHEN BAD-DATA ARE AROUND THE CRITICAL ZONE

Cases	Location of Bad-data	NCC (Not Covering Critical)	CC (Covering Critical)
White noises only	N/A	3.25E-02	3.56E-02
Bad-data locations	Vm_12	4.18E-02	3.66E-02
	QT_15-12	4.48E-02	3.56E-02
	PF 12-14	3.88E-02	3.52E-02
	PF 12-16	3.96E-02	3.68E-02
	QF 12-14	4.23E-02	3.50E-02
	QF 12-16	4.06E-02	3.60E-02

In the 30-bus test case, a PMU was installed at bus-12 to cater for the critical measurements, which are PF_12-13 and QF_12-13. Therefore, the optimal solution set by BGA is 1, 12 and 18. If bad-data comes from the critical zone in the presence of these three PMUs, it is noticed from the table that the NCE indicator remains almost unchanged. Initially, with no bad-data presence, error indicator was 0.0356 with the optimal PMU placement. Later, with the presence of bad-data around the critical zone, this indicator remained almost constant thereby showing significant robustness. The reason behind this is the installation of a PMU in bus-12 which covers the critical zone with high accurate phasor measurements.

On the other hand, if the constraint of covering the critical zone is not there in the optimization (NCC case), the solution set for 3 PMU installation is 1, 10, 28. When the system is considered without bad-data, the NCE value is 0.0315 which is lower than the CC case as it results the global solution. However, the presence of bad-data around the critical location has changed the complete scenario and shows how the estimation performance falls with a single measurement. The table shows that the NCE values are very high for NCC case when the system was considered to deal with bad-data. Thus, the significance of covering the critical zone by the highly accurate PMU meter is proved in this paper.

VI. CONCLUSIONS

This paper presents an effective approach for finding the optimal location of PMUs, for a specified and limited number of PMUs in a SCADA-based power network. The BGA has been used to solve the optimization problem. The objective of the optimization technique is to ensure the optimum estimation performance. The problem formulation also ensures the effect of bad-data presence in the estimation performance was minimized by covering the critical zones. A major problem with the covering of the critical locations is that it could lead to a non-global solution. However, our results show that the solution is robust to the effect of the presence of bad-data. The obtained results for the optimal solution sets and the corresponding NCE indicators for both the IEEE-14 and IEEE-30 bus power systems have been presented in this paper, which can be further applied to any size of power system. In today's modern era, there is a general inclination towards the installation of PMUs over existing SCADA-system for smart grid applications. Therefore, the proposed method of finding the optimum locations for the placement of a predetermined number of PMUs would be a valuable and effective tool for both the planning and operations of modern electricity power systems.

REFERENCES

[1] M. Hurtgen and J.-C. Maun, "Advantages of power system state estimation using Phasor Measurement Units."

[2] A. Gómez-Expósito, A. Abur, P. Rousseaux, A. De La, V. Jaén, and C. Gómez-Quiles, "On the Use of PMUs in Power System State Estimation."

[3] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, "Optimal placement of phasor measurement units: A literature review," in *2011 16th International Conference on Intelligent System Applications to Power Systems*, 2011, pp. 1–6.

[4] K. K. More and H. T. Jadhav, "A literature review on optimal placement of phasor measurement units," in *2013 International Conference on Power, Energy and Control (ICPEC)*, 2013, pp. 220–224.

[5] "Application of heuristic algorithms to optimal PMU placement in electric power systems: An updated review," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 214–228, Oct. 2015.

[6] K. Negash, B. Khan, and E. Yohannes, "Artificial Intelligence Versus Conventional Mathematical Techniques: A Review for Optimal Placement of Phasor Measurement Units," *Technol. Econ. Smart Grids Sustain. Energy*, vol. 1, no. 1, p. 10, Dec. 2016.

[7] R. Zivanovic and C. Cairns, "Implementation of PMU technology in state estimation: an overview," in *Proceedings of IEEE. AFRICON '96*, vol. 2, pp. 1006–1011.

[8] M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, "An Alternative for Including Phasor Measurements in State Estimators," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930–1937, Nov. 2006.

[9] R. F. Nuqui and A. G. Phadke, "Hybrid Linear State Estimation Utilizing Synchronized Phasor Measurements," in *2007 IEEE Lausanne Power Tech*, 2007, pp. 1665–1669.

[10] N. H. Abbasy and H. M. Ismail, "A Unified Approach for the Optimal PMU Location for Power System State Estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 806–813, May 2009.

[11] K. Zhu, L. Nordstrom, and L. Ekstam, "Application and analysis of optimum PMU placement methods with application to state estimation accuracy," in *2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1–7.

[12] M. Yucra, F. Schmidt, and M. C. de Almeida, "A comparative analysis of inclusion of PMUs in the power system state estimator," in *2015 IEEE 6th Latin American Symposium on Circuits & Systems (LASCAS)*, 2015, pp. 1–4.

[13] J. Chen, Y. L.-J. E. E. E. T. 1, and undefined 2012, "Optimal placement of phasor measurement units for improving power system state estimation accuracy: A heuristic approach," *pdfs.semanticscholar.org*.

[14] M. Shahriar, I. Habiballah, and H. Hussein, "Optimization of Phasor Measurement Unit (PMU) Placement in Supervisory Control and Data Acquisition (SCADA)-Based Power System for Better State-Estimation Performance," *Energies 2018, Vol. 11, Page 570*, vol. 11, no. 3, p. 570, Mar. 2018.

[15] A. Abur, "A bad data identification method for linear programming state estimation," *IEEE Trans. Power Syst.*, vol. 5, no. 3, pp. 894–901, 1990.

[16] A. Abur and M. K. Celik, "Least absolute value state estimation with equality and inequality constraints," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 680–686, May 1993.

[17] M. S. Shahriar and I. O. Habiballah, "Least measurement rejected algorithm for robust power system state estimation," in *2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, 2017, pp. 1–6

[18] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

A Direct Optimal Control of the Jebba Hydropower Station

[1]Olalekan Ogunbiyi, [2]Cornelius T. Thomas, [3]Benjamin J. Olufeagba,[4] Ibrahim S. Madugu, [5]Busayo H. Adebiyi, [1]Lambe M. Adesina
[1]Electrical and Computer Engineering, Kwara State University, Malete, Malete, Nigeria.
[2]Electrical and Information Engineering, Achievers University, Owo, Nigeria
[3]Electrical and Electronics Engineering, University of Ilorin, Ilorin, Nigeria
[4]Electrical Engineering, Kano University of Science and Technology, Wudil, Nigeria
[5]Computer Engineering, Ahmadu Bello University, Zaria, Nigeria

***Abstract* — The optimal power generation from the Jebba hydroelectric power station is subject to the reservoir operating head, weather-related factors, units' availability and system dynamics. In this paper, a computer control system is designed to ensure safe operation and maximize power generation. The controller is an optimal controller, which determines the amount of inflow required to regulate the reservoir operating head. The control law is an optimal control procedure developed around the steepest descent and conjugate gradient algorithm. The algorithms determine the control signal and state trajectories for minimization of a performance index defined for the regulation of the reservoir operating head. The results show that the two techniques are feasible in estimating an optimal inflow needed to move the reservoir operating head from any level to the nominal head. The two techniques were compared under different operating conditions of the hydropower system, the conjugate gradient algorithm performs better in terms of computational time. The control algorithm is recommended for use in the realization of a computer control system for the station.**

Keywords *— Conjugate gradient, differential equation, operating head, optimal control, performance index.*

I. INTRODUCTION

The hydroelectric power system consists of various dynamical responses such that research in this field seems unending. These include the reservoir dynamics, turbine dynamics, generator dynamics, cooling system[1]. The Jebba Hydroelectric Power Station (JHEPS) in Nigeria is not an exception. It is located at relatively on $09^0 08' 08'' N,\ 04^0 47' 16'' E$, 103 km downstream of the Kainji Hydroelectric Power Station (KHEPS) [2], [3].

The JHEPS was installed with six units of fixed blade propeller type turbo-alternator and each is rated at 96.4 MW. The nominal operating head is specified at 27.6 m with a 380 m^3/s maximum flow per unit. The turbo-alternators are connected to a transformer, each rated 119 MVA to step the voltage up to $330\ kV$ [3], [4]. While the KHEPS is designed such that the operating head can varies within 24m to 24m, JHEPS was designed in such a manner that the operating head remains relatively constant for system stability and station safety. The nominal operating head is 27.1m while the min net head is 24m and the maximum of 29.3m. Electric power generation at JHEPS is highly affected by the seasonal inflow on the River Niger with its white and black flood characteristics, the operation on the upstream reservoir at Kainji, the number of available units at a time and environmental factors that are weather related[5]. In order to ensure an optimal power generation in the presence of varying and uncertain inflow, an exponentially smoothed time series model of the inflow was suggested by Ale et al.[6].

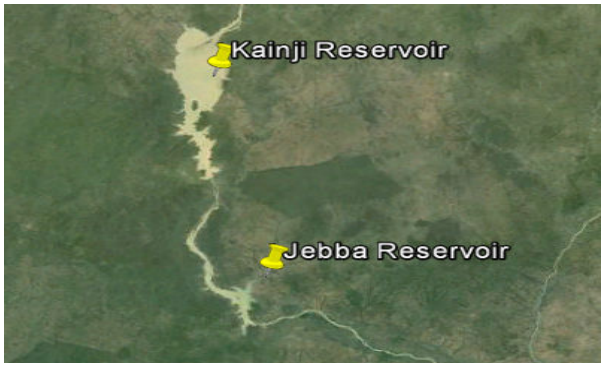


Fig. 1. Plan View of the KHEPS and JHEPS

The model could forecast the probable inflow on the River Niger over the year but it cannot handle the variation in reservoir head due to the turbo-alternators (TAs) availability. The failure of a unit at KHEPS reduces the inflow into JHEPS except on free fun, this can lead to a reduction in the number of operating TAs at JHEPS, thus reducing the power generated [3]. In Nwobi-Okoye and Igboanugo [7], a time series analysis of the inflow into KHEPS and the generated power output for a 10 years period resulted in a transfer function. This provided a means of estimating the efficiency of the power generation. Unfortunately, the model formulated using this method is non-causal, hence cannot be used in control design.

To obtain optimal power generation, a statistical model of inflow observation and the use of dynamic programming for optimal policy formulation have been widely presented[8], [9]. Similarly, an artificial neural network model was also presented in [10]. A more suitable technique is needed to handle a nonlinear dynamical system. The determination of the control input that is needed to force a dynamical system to follow a predefined trajectory and at the same time minimising a defined performance index within the system constraints is one of the problems solved in modern control engineering. Hence, an optimal control problem can be formulated such as to optimise the power generation from JHEPS and ensure a relatively constant operating head. The optimal control, in this case, is the amount of release from the upstream reservoir that will restore the head to the nominal value, whenever there is a deviation. Hence the station can be better managed and generation potential maximised.

The system dynamical model can be linear or nonlinear, there are usually, constraints on the boundary value and sometimes there can be constraints on the system state. The performance index can also be very complicated such that the solution requires sophisticated techniques. The solutions to an optimal control problem are grouped into two: indirect and direct methods [11]. The indirect methods solve the optimal control problem using the Pontryagin maximum principle. It is based on the calculus of variations to construct the formulation of the Hamiltonian function. The Euler-Lagrange equation is then used to obtain the optimal control canonical equations. Solutions to these equations ensure the determination of the optimal control and state trajectories. Usually, these leads to solving a multi-point boundary value problem by advanced numerical techniques involving the computation of partial derivatives. Also, powerful techniques may also be required where there are path inequalities. In the case of the direct methods, the optimal solution is determined by converting the infinite-dimensional optimization problem to a finite-dimensional optimization problem[12], [13]. The optimal control problems are formulated as a nonlinear programming problem by discretization of the state and the control variables. The solution is then obtained using an appropriate optimization algorithm[14]. In most cases, the solution is numerical such that the algorithm iterates until a certain condition or stopping criteria are met[15]. According to Rao (2009), the numerical methods of solving nonlinear programming problem is categorised into the gradient-based methods and the heuristic-based methods [13], [16]. The steepest descent and conjugate gradient methods are efficient gradient-based optimization algorithm for solving nonlinear programming problem. In this paper, the two methods used as a direct method of solving the optimal control of the JHEPS operating head.

II. The Optimal Control Problem at JHEPS

A typical close loop control system is as shown in Fig. 2, consisting of the desired operating head, the controller, the actuator and the system.An essential element of solving optimal control is the dynamic equation describing the system JHEPS [17]. The structural model of the JHEPS system is presented in Fig. 3, showing the reservoir, the turbine and the TAs. The electrical power (P) generated is related to the operating head by Equation (1), where n is the number of operating units, η represents the conversion efficiency of the turbo-alternators, ρ stands for the density of water in kg/m^3, A is the effective correctional area at the scroll casing and g is the acceleration due to gravity (m/s^2). Let $x(t)$ represents the operating head and the system state while $u(t)$ is the net inflow.Equation (1) implied that the power generation depends greatly on the operating head dynamics, hence the dynamical model for the operating head of JHEPS is represented by Equation (2), where α and μ are constants relating to the reservoir [5].

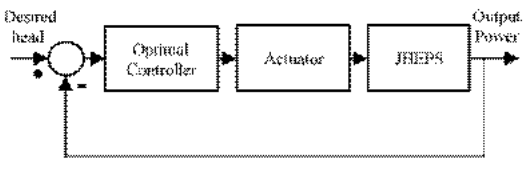


Fig. 2. The structural model of the JHEPS system

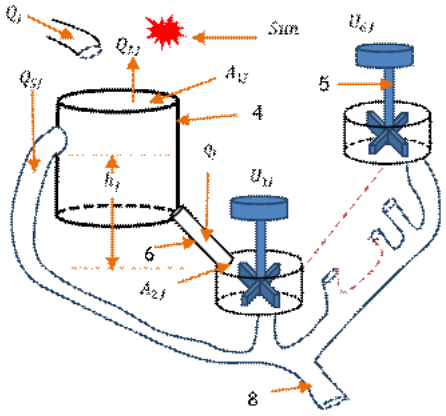


Fig. 3. The structural model of the JHEPS system

$$P = \sqrt{2}\, n\eta\rho A\, g^{3/2}\, x^{3/2}(t) \quad (1)$$

$$\frac{dx(t)}{dt} = -n\alpha\, x^{1/2}(t) + \mu u(t) \quad (2)$$

Equation (2) can be written in the standard form for the nonlinear system as shown in equation (3), where f represents a nonlinear function.

$$\dot{x}(t) = \boldsymbol{f}(x(t), u(t), t)\ ;\ t_0 \leq t \leq t_f \quad (3)$$

There are different kinds of optimal control problem associated with equation (2). The focus here is on the estimation of the control vector $u(t)$ that forces the state trajectories to move from a point $x(t_0)$ to another point $h(t_f)$ within a time frame $(t_0 \rightarrow t_f)$ and at the same time minimizing the performance index $I(x,u,t)$. $I(x,u,t)$ is defined in equation (4), where K_x is a positive scalar constant.

$$I(x,u,t) = min \int_{t_0}^{t_f} \left\{K_x\big(x(t) - x(T)\big)^2\right\} dt \quad (4)$$

The performance index is subject to the following constraints:

$$\dot{x}(t) = \boldsymbol{f}(x(t), u(t), t); \quad t_0 \leq t \leq t_f$$

$$x(t_0) = x_0 \quad (5)$$

$$x(t_f) = x(T) \quad (6)$$

$$[u_{min}(t) \leq u(t) \leq u_{max}(t)]$$

Since a direct method of solution is to be used for solving the optimal control problem, the time interval $t_0 \rightarrow t_f$ is divided into $N > 1$ intervals, resulting in equation (7). Within these sub-intervals, the infinite-dimensional control signal is approximated to a finite control vector of the form in equation (8). Similarly, the performance index is partitioned to create a vector of the form in equation (9).

$$\pi[t_0, t_f] = t_0 < t_1 < t_2 < \cdots < t_N = t_f \quad (7)$$

$$u(t) = [u_1, u_2, u_3, \ldots, u_N] \quad (8)$$

$$I(x,u,t) = [I_1, I_2, I_3, \ldots, I_N] \quad (9)$$

Hence the problem has been converted to a nonlinear optimization which is solved using the steepest descent and conjugate gradient algorithms.

III. Optimal Control using a Steepest Descent Algorithm

The steepest descent algorithm is a gradient-based technique used to solve a problem involving the determination of the minimum of a function that is differentiable. It is important that the performance index $I(x,u,t)$ must be differentiable such that the descent direction equals the path opposite $\nabla I(x,u,t)$.

Given an initial control input $u^{k=0}(t)$, the algorithm modifies the control as the iteration k increases until the minimum point is reached where $u^k(t) = u^*(t)$. If the initial guess is far from the feasible region, the algorithm may diverge [16], [18], [19].

At the minimum point, it is expected that;

$$\nabla I^* = \frac{\partial I^*(x,u,t)}{\partial u^*(t)} = \left\| \frac{\partial I_1^*}{\partial u_1^*}, \frac{\partial I_2^*}{\partial u_2^*}, \frac{\partial I_3^*}{\partial u_3^*}, \dots, \frac{\partial I_N^*}{\partial u_N^*} \right\| \approx 0 \qquad (10)$$

Let $\boldsymbol{\tau}(u^k)$ be a unit vector along the increasing gradient, the search direction in the opposite. The unit vector is defined as,

$$\boldsymbol{\tau}(u^k) = \frac{\left[\frac{\partial I^k}{\partial u^k}\right]^T}{\left\|\frac{\partial I^k}{\partial u^k}\right\|} = \frac{\left[\frac{\partial I_1^k}{\partial u_1^k}, \frac{\partial I_2^k}{\partial u_2^k}, \frac{\partial I_3^k}{\partial u_3^k}, \dots, \frac{\partial I_N^k}{\partial u_N^k}\right]^T}{\left\|\frac{\partial I_1^k}{\partial u_1^k}, \frac{\partial I_2^k}{\partial u_2^k}, \frac{\partial I_3^k}{\partial u_3^k}, \dots, \frac{\partial I_N^k}{\partial u_N^k}\right\|} \qquad (11)$$

Then the next control vector can be computed by equation (12), where ψ^k is step size along the search direction. A line search is usually performed to determine the optimum value for ψ.

$$\begin{bmatrix} u_1^{k+1} \\ u_2^{k+1} \\ u_3^{k+1} \\ \vdots \\ u_N^{k+1} \end{bmatrix} = \begin{bmatrix} u_1^k \\ u_2^k \\ u_3^k \\ \vdots \\ u_N^k \end{bmatrix} - \psi^k \boldsymbol{\tau}(u^k) \qquad (12)$$

IV. Optimal Control using a Conjugate Gradient Algorithm

Another gradient-based algorithm for the minimization of a differentiable function is the conjugate gradient algorithm. Lasdon (1967) used this algorithm to solve an unconstrained optimal control problem applied it before it was employed in solving a constrained optimal control problem where a penalty function imposed on the state and control [11], [20]–[22].

The algorithm incorporates the fast characteristics of steepest descent method whereby it determines a feasible solution after a few iterations but uses additional modification in ensuring that it converges to an optimum solution with few iterations.

The steps start by guessing and initial control vector $u^k(t) = [u_1^k, u_2^k, u_3^k, \dots, u_N^k]$, and determine the performance index $I(u_1^k, u_2^k, u_3^k, \dots, u_N^k)$ with the first iteration where $k = 0$.

The next step is to determine the gradient vector g^k.

$$g^k = \nabla I(u^k) = [g_1^k, g_2^k, g_3^k, \dots, g_N^k]$$

Determine the conjugate gradient parameter β^k, where

$$\beta^k = \frac{\int_{t_0}^{t_f} (g^k)^2 \, dt}{\int_{t_0}^{t_f} (g^{k-1})^2 \, dt} \qquad (13)$$

When $\mathrm{k} = 0$, $g^{k-1} = 1$

Proceed by calculating the vector s^k (direction of search).

$$\begin{bmatrix} s_1^k \\ s_2^k \\ s_3^k \\ \vdots \\ s_N^k \end{bmatrix} = \begin{bmatrix} g_1^k \\ g_2^k \\ g_3^k \\ \vdots \\ g_N^k \end{bmatrix} + \beta^k \begin{bmatrix} s_1^{k-1} \\ s_2^{k-1} \\ s_3^{k-1} \\ \vdots \\ s_N^{k-1} \end{bmatrix} \qquad (14)$$

$when\ k = 0\ , s^{(k-1)} = 0,$

Obtain the next control vector $u^{(k+1)}(t)$, where

$$\begin{bmatrix} u_1^{k+1} \\ u_2^{k+1} \\ u_3^{k+1} \\ \vdots \\ u_N^{k+1} \end{bmatrix} = \begin{bmatrix} u_1^k \\ u_2^k \\ u_3^k \\ \vdots \\ u_N^k \end{bmatrix} - \psi_{min} \begin{bmatrix} s_1^k \\ s_2^k \\ s_3^k \\ \vdots \\ s_N^k \end{bmatrix} \qquad (15)$$

The optimum step size ψ_{min} was determined via a minimization search on the function $I(\,u^{(k)} + \psi\, s^k)$.

Check if $|I(u^{k+1}) - I(u^k)| \leq 10^{-n}$ to terminate the iterative procedure, given that n is a scalar constant. Otherwise, increment k and repeat the steps using a new control input $u^{(k+1)}$.

The control procedure by the steepest descent and conjugate gradient algorithms were realised in a Microsoft EXCEL VBA® environment. Since the model, performance index and the constraints are nonlinear; the Adams-Moulton based numerical integrator was employed in the computations. The results in the computations are presented in the next section.

Since the aim of this work is to present the control law, the control procedure is thus presented in Fig. 4, Fig. 5 and Fig. 6. The procedure starts in a similar manner as presented in Fig. 4. For the steepest descent technique, then algorithm proceeds by following Fig. 5 while Fig. 6 is for the conjugate gradient.

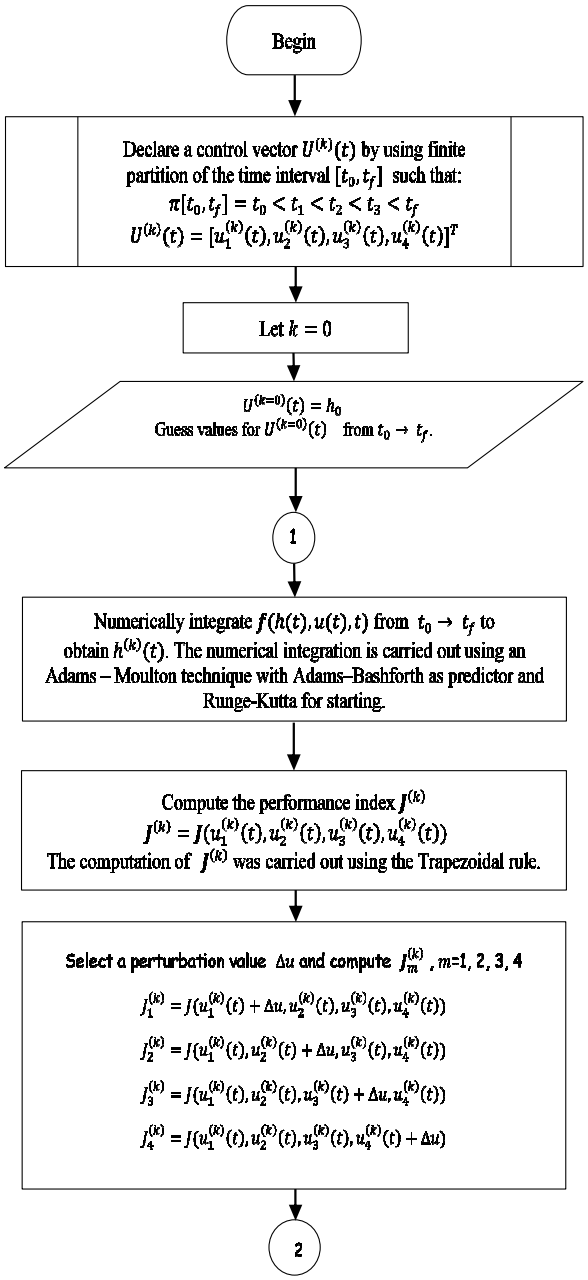


Fig. 4. Flow chart for the direct optimal control of the JHEPS

V. Results and Discussion

Given the operating conditions of the turbo-alternators, the reservoir characteristics, initial head and the desired head. The optimal controls required to achieve the set conditions are as presented in this section. The cases show the realization performed to study the potential of the two algorithms and comparison of their results. The notation for specifying operating conditions can be defined as follows: (*Total turbo-alternators that are running, the initial operating head (m), time intervals (Days), penalty*).

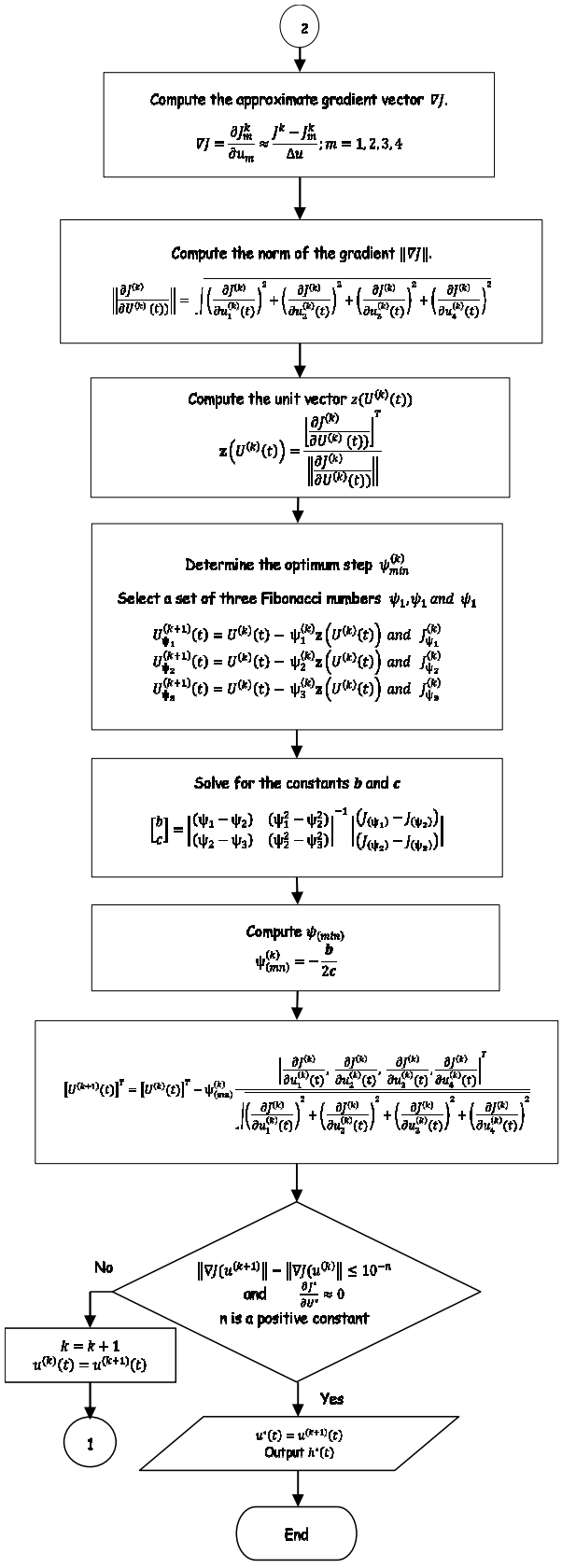
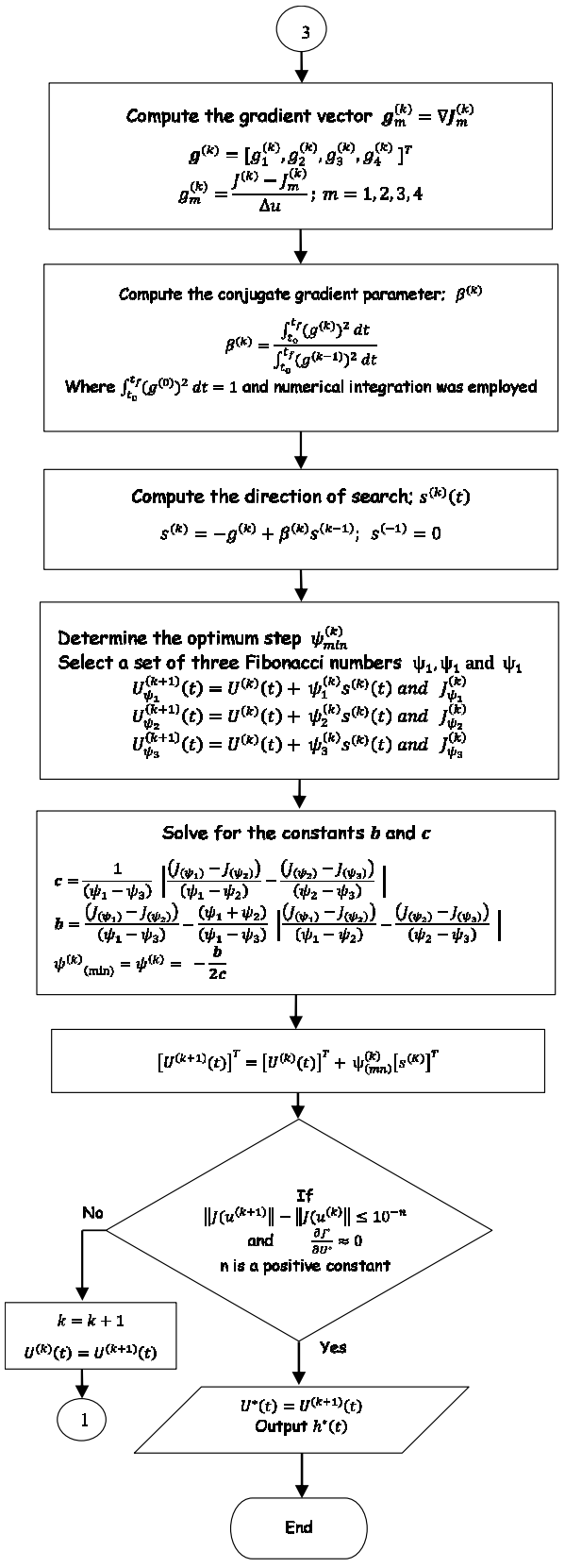


Fig. 5. Flow chart for the Optimal Control By Steepest Descent Algorithm



A flow chart for the Optimal Control by Conjugate Gradient Algorithm

A. *Case 1:* $(5, 25.8, 1, u_{max}\ unpenalized)$

Consider a case where the initial operating head $h(0) = 25.8\,m$ and the desired operating head is $h(T) = 26.1\,m$. The initial guesses for the controls are $u_1(0) = u_2(0) = u_3(0) = u_4(0) = 1000\ m^3/s$. Given that there are five (5) units in operation and the optimal control is required to move the operating head form $h(0)$ to $h(T)$ in $24\ hrs$ .Fig. 7 presents the trajectory of the performance index as the algorithm iterates until it converges. The Performance index must be minimum at the optimum point. A similar situation was experienced for both the steepest descent and conjugate gradient algorithm. This is a sufficient condition that the two techniques are feasible in the determination of optimal control for JHEPS.

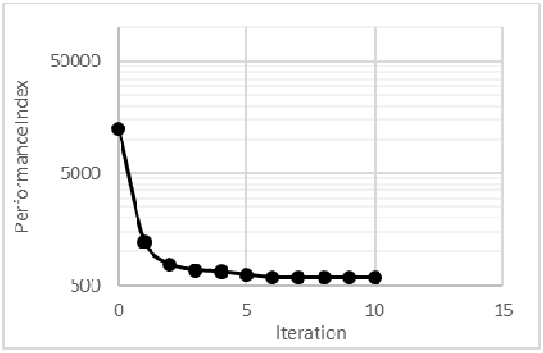


Fig. 6. Performance Index

It should also be noted that the optimal control can be bounded or unbounded, in this case, the maximum control is not penalized. When the input conditions were passed into the optimal control procedure, the operating head trajectory using the steepest descent and conjugate gradient algorithms presented in Fig. 8. Both methods can solve the direct optimal control problem with a unique characteristic of producing a slight overshoot in the first quarter of the time. From Fig. 9, the two techniques produced an optimal control around:

$$u_1^* = 6287.13\ m^3/s, u_2^* = 368.77\ m^3/s,$$
$$u_3^* = 1692.96\ m^3/s, u_4^* = 1725.85\, m^3/s$$

The performance of the two techniques differs in the computational time, the steepest descent algorithm converges after 13 iterations while the conjugate gradient algorithm converges after 10 iterations, this is evident in Fig. 10.Hence if the maximum control is not penalized, the computation based conjugate gradient algorithm is faster than that of the steepest descent. Meanwhile, the control of $6287.13\ m^3/s$ seems too large that is may not be feasible to release such from JHEPS.

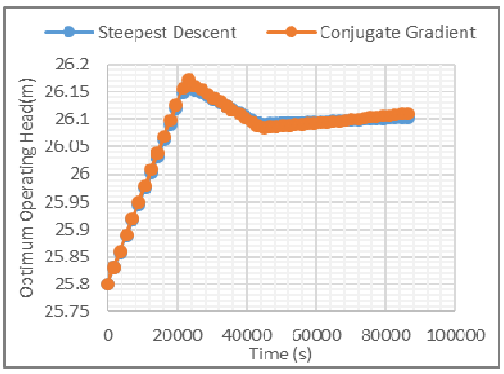


Fig. 7. Comparison of the operating Head by Steepest Descent and Conjugate Gradient under $(5, 25.8, 1, u_{max}\ unpenalized)$

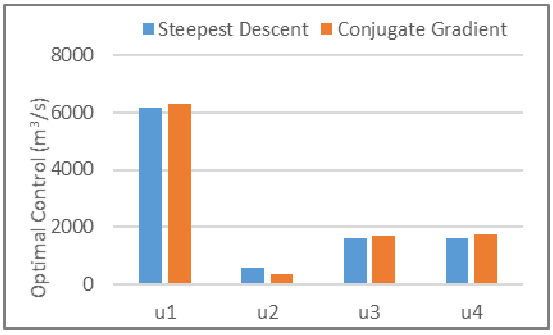


Fig. 8. Comparison of the optimal control by Steepest Descent and Conjugate Gradient under $(5, 25.8, 1, u_{max}\ unpenalized)$

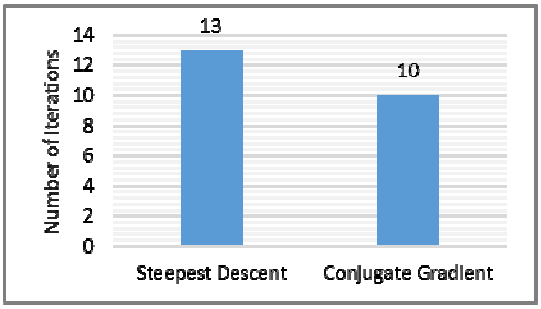


Fig. 9. Number of iterations for $(5, 25.8, 1, u_{max}\ unpenalized)$

B. *Case 2:* $(5, 25.5, 1, u_{max} = 4000\ m^3/s)$

In the operating conditions is slightly changed such that $h(0) = 25.5\ m$ and the maximum control is bounded to be $u_{max} = 4000\ m^3/s$ while all other operating conditions remain the same. In Fig. 11, the operating head trajectory for $h(0)$ to $h(T)$ do not produce an overshoot like in the unbounded case. Optimal control remained at $4000\ m^3/s$ until the last quarter of the time, it decreases to $2236.83\ m^3/s$, this can be seen from Fig. 12.

This is a better response as compared to the previous case, showing that it is better to compute the direct optimal control with a penalized maximum control. Fig. 13 gives a clear difference in the performance of the techniques, showing that the conjugate gradient-based algorithm that converges after 14 iterations is better than the steepest descent based technique that converges after 115 iterations.

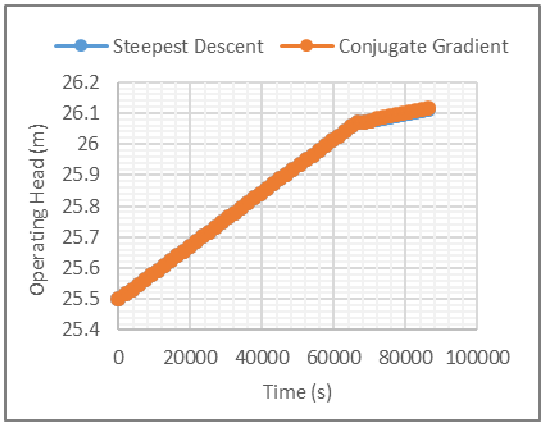


Fig. 10. Comparison of the operating Head by Steepest Descent and Conjugate Gradient under $(5, 25.5, 1, u_{max} = 4000\ m^3/s)$

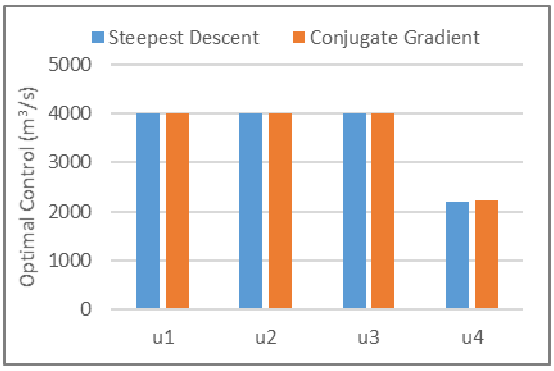


Fig. 12. Comparison of the optimal control by Steepest Descent and Conjugate Gradient under

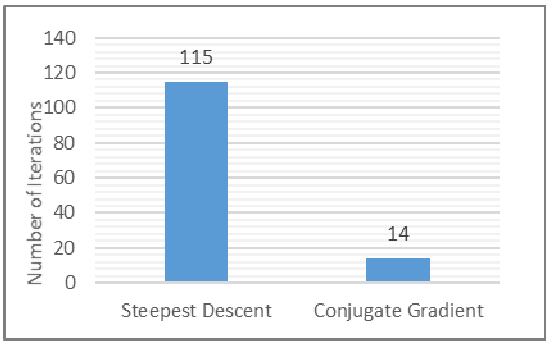


Fig.13. Number of iterations $(5, 25.5, 1, u_{max} = 4000\ m^3/s)$

C. *Case 3:* $(4, 25.8, 1, u_{max} = 4000\ m^3/s)$

Consider the case where $h(0) = 28.5\ m$, the maximum control is bounded at $4000\ m^3/s$, the number of operating machines is four (4) while other operating conditions do not change. Fig. 14 presents the operating head trajectories for the two techniques, the Conjugate gradient technique overshoot slightly beyond $26.1\ m$ at $h(T)$, while the response from the steepest descent technique seems slightly better. Meanwhile the optimal control plot of Fig. 15 shows that the inflow required will be less as compared to that of the steepest descent techniques. In Fig. 16, the result is similar to the previous case considered; the conjugate gradient technique converges after two iterations while steepest descent technique converges after 113 iterations.

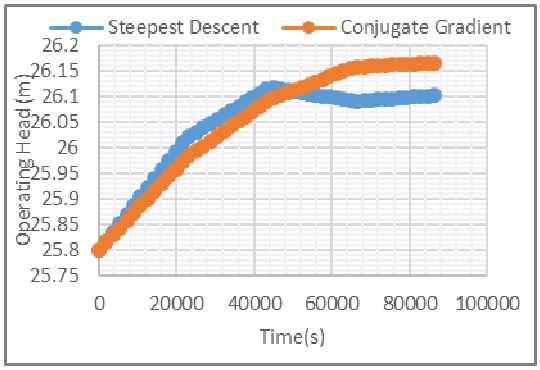


Fig 14.. Comparison of the operating Head by Steepest Descent and Conjugate Gradient Technique for (4, 25.8, 1, u_{max} = 4000 m^3/s)

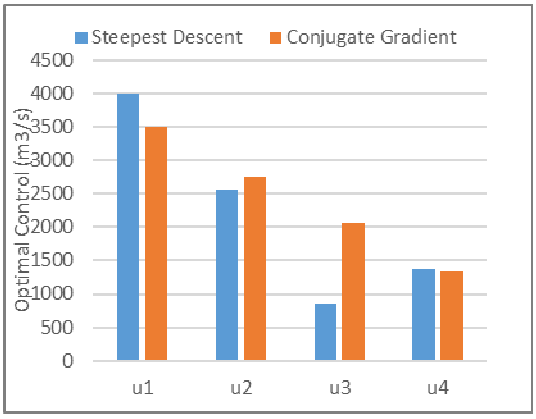


Fig. 15. Comparison of the optimal control by Steepest Descent and Conjugate Gradient Technique for (4, 25.8, 1, u_{max} = 4000 m^3/s)

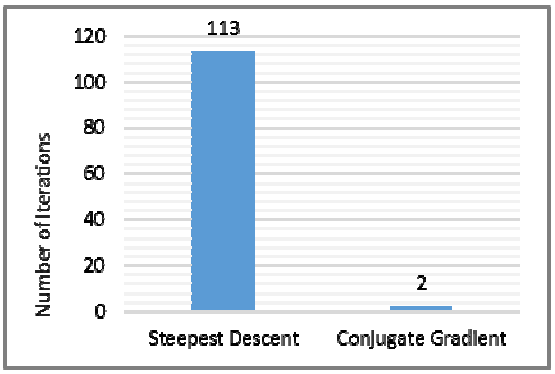


Fig 16. Number of iterations (4, 25.8, 1, u_{max} = 4000 m^3/s)

VI. Conclusion

This paper presents the determination of optimal control for the JHEPS operating head based on the steepest descent and the conjugate gradient technique. It was discovered that the two techniques have the potential of saving the direct optimal control method, but the conjugate gradient method converges faster than that of the steepest descent technique. In addition, when the two algorithms were compared in terms of the number of iterations before convergence, the conjugate gradient algorithm converges faster.

Therefore, the two control procedures are recommended for the development of control laws for the operation of the JHEPS and management of the operating head but the technique based on the conjugate gradient will be faster.

References

[1] N. Algorithm, "Optimal Operation of Hydropower Station Unit Based on," *Sci. Technol.*, pp. 2216–2219, 2010.

[2] C. T. Thomas, M. F. Akorede, O. Ogunbiyi, B. J. Olufeagba, and J. S. Samuel, "A Study of Energy Conversion at the Jebba Hydroelectric Power Station," in *2017 IEEE 3rd International Conference on Electro-Technology for National Development, NIGERCON 2017*, 2018, pp. 828–833.

[3] O. Ogunbiyi, C. T. Thomas, I. A. O. Omeiza, J. Akanni, and B. J. Olufeagba, "Dynamical Control Model of the Cascaded Kainji-Jebba Hydropower Operating Head," *FUOYE J. Eng. Technol.*, vol. 4, no. 1, 2019.

[4] T. S. Abdulkadir, A. W. Salami, A. R. Anwar, and A. G. Kareem, "Modelling of Hydropower Reservoir Variables for Energy Generation: Neural Network Approach," *Ethiop. J. Environ. Stud. Manag.*, vol. 6, no. 3, pp. 310–316, 2013.

[5] O. Ogunbiyi, C. T. Thomas, I. O. A. Omeiza, J. Akanni, and B. J. Olufeagba, "A Nonlinear Control Model and Operational Support System for the Kainji Hydroelectric Power System," *Niger. J. Technol.*, vol. 38, no. 2, pp. 449–455, 2019.

[6] T. O. Ale, K. E. Alowolodu, J. O. Babatola, and B. J. Olufeagba, "Inflow Forecasting for Kainji Dam Using Time Series Model," *Int. J. Math. Arch.*, vol. 2, no. 12, pp. 2844–2851, 2011.

[7] C. C. Nwobi-Okoye and A. C. Igboanugo, "Performance evaluation of hydropower generation system using transfer function modelling," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 245–254, 2012.

[8] J. O. Aribisala, "Water Use Forecast for Hydropower Generation," *J. Eng. Appl. Sci.*, vol. 2, no. 1, pp. 222–225, 2007.

[9] A. W. Salami, "Operational Performance of Water management Models for Hydropower System under Reservoir Inflow Forecast.," University of Ilorin, 2007.

[10] T. S. Abdulkadir, F. Sule, and A. W. Salami, "Application of Artificial Neural Network Model To the Management of Hydropower Reservoirs Along," *Ann. Fac. Eng. HUNEDOARA - International J. Eng.*, vol. 10, no. 3, pp. 419–424, 2012.

[11] R. W. H. Sargent, "Optimal Control," *J. Comput. Appl. Math.*, vol. 124, pp. 361–371, 2000.

[12] K. Ratkovic, "Limitations in direct and indirect methods for solving optimal control problems in growth theory," *Industrija*, vol. 44, no. 4, pp. 19–46, 2016.

[13] J. T. Betts, "A Survey of Numerical Methods for Trajectory Optimization," *AIAA J. Guid. Control. Dyn.*, vol. 21, pp. 193–207, 1998.

[14] J. T. Betts, *A Direct Approach to Solving Optimal Control Problem*. 1999, pp. 73–75.

[15] G. R. Rose, "Numerical Methods for Solving Optimal Control Problems," University of Tennessee, 2015.

[16] A. V Rao, "A survey of numerical methods for optimal control," *Adv. Astronaut. Sci.*, vol. 135, no. 1, pp. 497–528, 2009.

[17] D. E. Kirk, "Optimal Control Theory: An Introduction." Dover Publication, New York, 1998.

[18] H. Schättler and T. J. Tarn, "Descent Algorithms for Optimal Periodic Output Feedback Control," *IEEE Trans. Automat. Contr.*, vol. 37, no. 12, pp. 1893–1904, 1992.

[19] K. S. Srinivas and K. Vaisakh, "Application of Steepest Descent And Differential Evolution Methods For Optimal Control Of Power Systems," *Imp. J. Interdiscip. Res.*, vol. 3, no. 1, pp. 2257–2263, 2017.

[20] J. K. Willoughby, "Adaptations of the conjugate gradient method to optimal control problems with terminal state constraints," 1969.

[21] L. S. Lasdon, "The Conjugate Gradient Method for Optimal Control Problem," *IEEE Trans. Automat. Contr.*, vol. 12, no. 2, pp. 132–138, 1967.

[22] M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems 1," *J. Res. BNational Bur. Stand.*, vol. 49, no. 6, pp. 409–436, 1952.