

**APPLICATION OF SUBSTITUTION ALGORITHM AND
FREQUENCY ANALYSIS ON INFORMATION SECURITY
BASED CRYPTANALYSIS**

by

OLAMIJULO, Gbemisola Yetunde
PGD/MCS/2008/1260

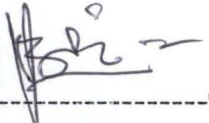
Submitted to Department of Mathematics / Computer Science
Federal University of Technology
Minna

In partial fulfillment of requirements leading to the award of
Postgraduate Diploma (PGD) in Computer Science
Federal University of Technology, Minna

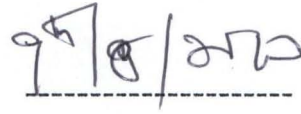
July 2010

CERTIFICATION

This project titled Application of substitution Algorithm and frequency analysis on information security by Olamijulo Gbemisola Y. (PGD/MCS/2008/1260), meets the regulations governing the award of Post-Graduate Diploma in computer science of Federal University of Technology, Minna, Niger State.



Dr. V.O. Waziri



Date

Prof. N.I. Akinwande
Head of Department

Date

ACKNOWLEDGEMENT

My sincere gratitude goes to the Almighty God who has called me specially and also given me the grace to go through this programme in line with his divine will and purpose for my life. My special thanks goes to my project supervisor, Dr V.O. Waziri for his concern and guidians. I express my warmest thanks to the Head of Department, Prof. N.I. Akinwande, the P G D Coordinator Mr Ndanusa A. for their support.

I also acknowledge the rest of the members of staff of Mathematics /computer whose importation of knowledge helped me in carrying out this project, and my classmates for their team spirit.

DEDICATION

This project is dedicated to Almighty God and personal saviour, to my family, for their love and support in all my endeavours.

TABLE OF CONTENT

Content	Page
TITLE PAGE	i
CERTIFICATION	ii
ACKNOWLEDGEMENT	iii
DEDICATION	iv
TABLE OF CONTENT	v
ABSTRCT	vi
CHAPTER ONE	
1.0 INTRODUCTION	1
1.1 OBJECTIVE OF THE STUDY	2
1.2 AIM OF THE STUDY	2
1.3 METHODOLOGY	3
1.4 SIGNIFICANCE OF THE STUDY	3
1.5 SCOPE OF THE STUDY	3
1.6 TERM DEFINTION	3
CHAPTER TWO	
2.0 LITERATURE REVIEW	5
2.1 INTRODUCTION	5
2.2 LETTER FREQUENCY IN SPANISH	6
2.3 FREQUENCY ANALYSIS IN FICTION	7
2.4 CRYPTOGRAPHY OVER VIEW	7
2.5 THE COMMUNICATION CHANNEL	7
2.6 THE SUBSTITUTION CIPHER	8
CHAPTER THREE	
3.0 ANALYSIS OF CASE STUDY	10
3.1 INTRODUCTION TO SUBSTITUTION CIPHER	10
3.2 HOW SUBSTITUTION ALGORITHM WORKS	11
3.2.1 SENDING A MESSAGE	11
3.2.2 RECEIVING THE MESSAGE	11
3.2.3 USING SUBSTITUTION CIPHER IN ENCRYPTING PLAINTTEXT	12
3.2.4 DECRYPTING OF SUBSTITUTION CIPHER	14
3.3 CRYPTANALYSIS	15
3.3.1 CRYPTANALYZING THE SUBSTITUTION CIPHER	16
3.3.2 SITUATION CONSIDERED	16
CHAPTER FOUR	
4.0 FREQUENCY ANALYSIS FOR SUBSTITUTION CIPHER	19
4.1 CRYPTANALYSIS OF THE SUBSTITUTION CIPHER	19
4.2 FREQUENCY OF OCCURRENCE OF 26 CIPHERTEXT LETTERS	19
4.3 THE DECRYPTED TEXT	25
CHAPTER FIVE	
5.0 SUMMARY, FINDINGS AND CONCLUSION	
5.1 SUMMARY	26
5.2 FINDINGS	26
5.3 CONCLUSION	26
5.4 REFERENCE	27

ABSTRACT

For centuries now, security of information transmission has been of importance such that security measure on information were used by an aged leader named Julius Caesar, who's name was given to a cipher (caesar cipher), who according to Suetonius, used a shift of three to protect message of military significance. With increase in technology, digital communication, electronic data exchange, the use of computers and the high rate of crime, information security has become an important issue in industry, business and administration. This project attempts to look into the issue of Substitution Algorithm as a cryptosystem providing essential technique for securing information and protecting data by the use of keys of alphabetic characters.

CHAPTER ONE

GENERAL INTRODUCTION

1.1 INTRODUCTION

Substitution Algorithm is a means of encrypting a plaintext into a ciphertext by associating each key word with an alphabetic characters. A substitution cipher involves the use of the twenty six (26) alphabets and integers to recreate words or sentence. These newly created words are decrypted by another person that's understands the key. Example is sending of email between two friends.

A substitution cipher is developed from cryptography. Cryptography is the science of writing in secrete codes. It is a means of permuting the twenty six letters for decrypting and encrypting messages.

The system enables two people, say (John and Blessing) to communicate over an insecure channel in such a way that an opponent, say Moses can not understand what has being said. This channel could be a telephone line or computer network. For example the information which John wants to send to Blessing is called a "plaintext" can be either English text or Numerical data. John encrypted (coded) the English text that is (plaintext), using a predetermined key and sends the resulting ciphertext (coded text) over the channel. The opponent, upon seeing the ciphertext in the channel or by eavesdropping, can not determine what the plaintext was. But Blessing, who

knows the encryption key, can decrypt the ciphertext and reconstruct the plaintext.

Frequency Analysis is the study of how often letters or group of letters occurs in ciphertext. Substitution algorithm and frequency analysis are necessary in cryptanalysing ciphertext into plaintext (that is from code to message) and from plaintext to ciphertext.

1.2 OBJECTIVE OF THE STUDY

- (1) To enable two people to communicate over an insecure channel in such away

An opponent can not understand what is being said.

- (2) To use mathematical notation to encrypt text. For example counting of ciphertext letters and then associating guess plaintext letter with them.

- (3) To be able to send a text with a full assurance that no one can understand it

Without your knowledge.

- (4) To find corresponding plaintext letter for each ciphertext.

1.3 AIM OF THE STUDY

It's enable you to have confidence that you have either sent a message or received message that is highly secured.

1.4 METHODOLOGY

The method I used in this project work are substitution Algorithm and Frequency Analysis Algorithm. The methods are used to encrypt plaintext (e_k) and the resulting ciphertext is subsequently decrypted using (d_k) for achieving the original plaintext.

1.5 SIGNIFICANCE OF THE STUDY

The study is based on encryption and decryption as permutations of Alphabetic characters

1.6 SCOPE OF THE STUDY

The study is based on securing of mail or text from a third party that wants to Intercept it.

1.7 TERMS DEFINITION

- (1) **Plaintext:** This is the information drafted to send to another person.
- (2) **Encryption:** I using mathematical notation and alphabet to code the drafted message to a new one that is meaningless.
- (3) **Ciphertext :** This is the newly created message that has no meaning.
- (4) **Decryption:** means of reconstructing the ciphertext back to a plaintext.
- (5) **Substitution cipher:** Is a means of encrypting or decrypting a text with the use of permutation.

- (6) **Cryptography:** Is a means of using secret code for securing information from eavesdroppers, hiding information.
- (7) **Cryptanalysis:** Is the breaking of coded text (ciphertext) to a meaningful message.
- (8) **Frequency Analysis:** It is the counting or knowing the number of Occurrence of the twenty six alphabetic characters in the Ciphertext.
- (9) **Security:** Activities involved in protecting against attack.
- (10) **Cryptographers:** they are people who do cryptography.
- (11) **Cryptanalyst:** Specialist in breaking coded message.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

In a Substitution Cipher, each letter of the plaintext is replaced with another, and any particular letter in the plaintext will always be transformed into the same letter in the ciphertext. For instance, if all occurrence of the letter e turn into the letter x, a ciphertext message containing numerous instances of the letter X would suggest to a cryptanalyst that X represents e.

The basic use of frequency analysis is to first count the frequency of ciphertext letters and then associate guessed plaintext letters with them. More X's in the ciphertext than anything else suggests that X corresponds to e in the plaintext, but this is not certain, t and a are also very common in English, so X might be either of them also. It is unlikely to be a plaintext z or g which are less common. Thus, the cryptanalyst may need to try several combinations of mappings between ciphertext and plaintext letters.

More complex use of statistics can be conceived, such as considering counts of pairs of letters (digrams), triplets (trigrams), and so on.

The first page of Al-kindī's 9th Century Manuscript on Deciphering Cryptographic Messages was the first known recorded explanation of frequency analysis on cryptanalysis has been suggested that close textual study of the Qur'an first brought to light that Arabic has a characteristic letter frequency. Its use spread, and similar system were widely used In European State by the time of the Renaissance. By 1474 Cicco Simonetta had written a manual on deciphering encryption of Latin and Italian text.

Several schemes were invented by cryptographers to defeat this weakness in simple substitution encryption. These include:

1. **Use of Homophones:** Several alternatives to the most common letters in otherwise monoalphabetic substitution ciphers (for example, for English, both X and Y ciphertext might mean plaintext E.)
- **Polyalphabetic Substitution:** the use of several alphabets chosen in assorted, more or less devious ways.
- **Polygraphic Substitution Scheme:-** Where pairs or triplets of plaintext letters are treated as unit for substitution, rather than single letters (for example, the playfair cipher invented by Charles Wheatstone in the mid 1800s).

A disadvantage of all these attempts to defeat frequency counting attacks is that it increases complication of both enciphering and deciphering, leading to mistakes. Famously, a British foreign secretary is said to have rejected the playfair cipher because, even if school boys could cope successfully as Wheatstone and Playfair had shown, our attache's could never learn it.

The rotor machines of the first half of the 20th century example, the (Enigma Machine) were essentially immune to straightforward frequency analysis. However, other kinds of analysis ("attack") successfully decode images from some of those machines.

2.2 **LETTER FREQUENCY IN SPANISH**

Frequency analysis requires only a basic understanding of the statistics of the plaintext language and some problem solving skill, and if performed by hand, some tolerance for extensive letter bookkeeping. During World War II (WW II), both the British and the Americans recruited code breakers by placing crossword puzzles in major newspapers and running contests for who could solve them the fastest. Several of the ciphers used by the axis powers were breakable using frequency analysis (for example, some of the consular ciphers used by the Japanese). Mechanical methods of letter counting and statistical analysis (generally IBM card type machinery) were first used in WWII, possibly by the US Army's SIS. Today, the hard work of letter

counting and analysis has been replaced by computer software, which can carry out such analysis in seconds with modern computing power, classical ciphers are unlikely to provide any real protection for confidential data.

2.3 FREQUENCY ANALYSIS IN FICTION: Is the description of people and events that are not real.

Frequency analysis has been described in fiction. Edgar Allan Poe's "The Gold Bug", and Sir Arthur Conan Doyle's Sherlock Holmes tale "The Adventure of the Dancing Men" are examples of stories which describe the use of frequency analysis to attack simple substitution ciphers. The ciphers in the Poe story is encrusted with several deception measures, but this is more a literary device than anything significant cryptographically.

2.4 CRYPTOGRAPHY OVER VIEW

The fundamental Objective of Cryptography is to enable two people, to communicate over an insecure channel in such a way that an opponent cannot understand what is being said. This channel could be a telephone line or computer network, for example. The information that Blessing wants to send to John is called plaintext, can be English text or numerics, its structure is arbitrary. Blessing encrypts the plaintext, using a predetermined key, and sends the resulting ciphertext over the channel. The opponent, upon seeing the ciphertext in the channel by eavesdropping, cannot determine what the plaintext is.

These ideas are described using the following mathematical notation.

- 1} P is a finite set of possible Plaintext
- 2} C is a finite set of possible Ciphertext
- 3} K, the keyspace, is a finite set of possible Keys.
- 4} For each $K \in K$, there is an encryption rule e_k and a corresponding decryption rule d_k

$d_k(e_k(x)) = X$ for every plaintext element $X \in P$

2.5 THE COMMUNICATION CHANNEL

Looking from the above mathematical notation for each $K \in K$, it says that if a plaintext x is encrypted using e_k , and the resulting ciphertext is subsequently decrypted using d_k , then the original plaintext x results.

Suppose Blessing wants to communicate a message to John over an in secure channel. We suppose that this message is a string.

$$X = x_1, x_2, x_3, \dots, x_n,$$

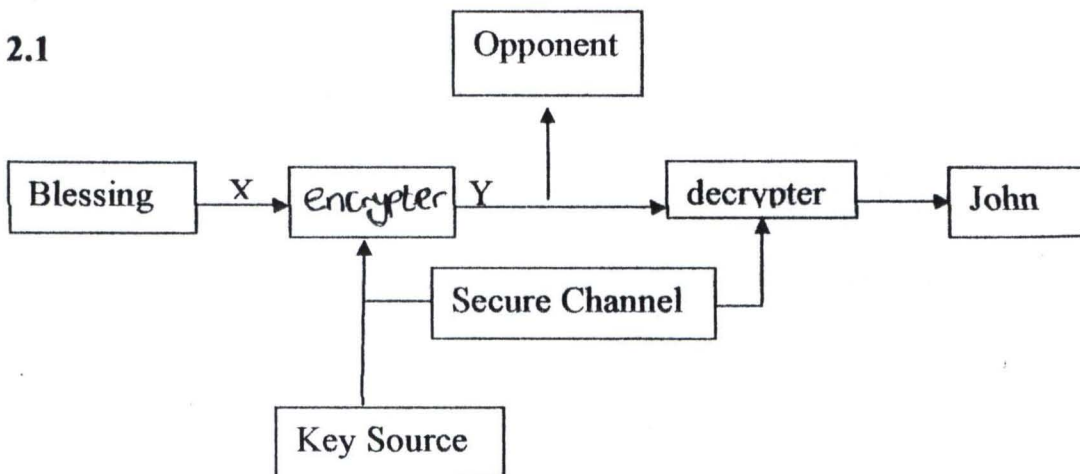
For some integer $n \geq 1$, where each plaintext symbol $X_i \in P, 1 \leq i \leq n$. each X_i is encrypted using the encryption rule e_k specified by the predetermined key K . Hence, Blessing computed $y_i = e_k(x_i), 1 \leq i \leq n$, and the resulting g ciphertext string.

$$Y = y_1, y_2, y_3, \dots, y_n$$

This is sent over the channel. When John receives $y_1, y_2, y_3, \dots, y_n$ he decrypts it using the decryption function d_k , obtaining the original plaintext string, $x_1, x_2, x_3, \dots, x_n$.

A picture for communication channel is found in figure 2.1 over leaf.

Fig. 2.1



Note that if $P = C$, it follows that each encryption function is a permutation. That is, if the set of plaintexts and ciphertexts are identical, then each encryption function just rearranges (or permutes) the elements of this set.

2.5.1 THE SUBSTITUTION CIPHER

The substitution cipher takes P and C both to be 26-letter English Alphabets. Substitution thinks of encryption and decryption as permutations of alphabetic characters. Example of a "random" permutation, π ,

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
X	N	Y	A	H	P	O	G	Z	Q	W	B	T	S	F

p	q	r	s	t	u	v	w	x	y	z
L	R	C	V	M	U	E	K	J	D	I

Thus, $e_{\pi}(a) = X$, $e_{\pi}(b) = N$ ----- $e_{\pi}(z) = I$

The decryption function is the inverse permutation. This is by writing the second line first, and then sorting in alphabetical order. The following is obtained.

2.5.2

A	B	C	D	E	F	G	H	I	J	K	L	M	N
d	l	r	y	v	o	h	e	z	x	w	p	t	b

O	P	Q	R	S	T	U	V	W	X	Y	Z
g	f	j	q	n	m	u	s	k	a	c	i

Hence, $d_{\pi}(A) = d$, $d_{\pi}(B) = L$, -----, $d_{\pi}(Z) = i$

Recall that, plaintext characters are written in lower case and ciphertext characters are written in upper case.

CHAPTER THREE
ANALYSIS OF CASE STUDY

3.1 INTRODUCTION TO SUBSTITUTION ALGORITHM

The substitution cryptosystem involves a 5 tuple (stages).

C : Finite set of ciphertext

P : Finite set of plaintext

K: Finite set of keys

E : Encrypting plaintext

D : Decrypting ciphertext

$d_k : P \rightarrow C$ (plaintext to ciphertext)

$e_k : C \rightarrow P$ (Ciphertext to plaintext)

$e_k(x)$ for every plaintext

let x be a string of message(plaintext)

$X = x_1, x_2, x_3, \dots, x_n$

$n \geq 1$

Plaintext symbol is x

$x : E_p, \leq i \leq n$

Each x_i is encrypted using encryption rule e_k with a predetermined key K .

$Y_i = e_k(x_i) = 1 \leq i \leq n$

$Y = y_1, y_2, y_3, \dots, y_n$ for every ciphertext

You decrypt y as it stands for ciphertext.

$Y = d_k(y_i) = d_k(y_2)$

Let $P = C = K = Z_{26}$

3.2 HOW SUBSTITUTION ALGORITHM WORKS

3.2.1 SENDING A MESSAGE

Sending message can be in different forms either through handset, internet or local post. All we are interested in is that we don't want the third person to understand our message.

John and Blessing want to communicate in such away that an opponent, Juliet, can not understand what is being said. John and Blessing will first choose a random key which they both agree on. After agreement on the key they choose, they can then go their different ways.

Blessing computes plaintext, she encrypt it using encryption rule e_k the text becomes:

HPHTWWDPPRLRITYLEXTOYTRSEXJMLMJ

Which is the ciphertext and sends it over the channel to John.

3.2.2 RECEIVING THE MESSAGE

The message sent to John was successfully received by him. Opening the message it reads:

HPHTWWDPPRLRITYLEXTOYTRSEXJMLMJ

Knowing the key, he sits down to decrypt the message with the key.

To decrypt the ciphertext, John first convert the ciphertext to a sequence of integers from 0 to 25 representing the 26 character of alphabet. Knowing the key between him and Blessing.

Let the key between them be 11. In decryption, you have $dk(y) = (y-k) \pmod{26}$

Subtract 11 from each value.

H=7, P=15, H=7, T=19, -----, J=9

H P H T W W D P P L R T Y L E X T O Y T R

7 15 7 19 22 22 3 15 15 11 17 19 24 11 4 23 19 14 24 19 17

11 11 11 11 - 11

22 4 22 8 11 11 8 4 4 0 6 8 13 0 19 12 8 3 13 8 6

S E X J M L M J

18 4 23 9 12 11 12 9

11 - - - - - - - - -

7 19 12 24 1 0 1 24

Using $dk(y) = (y-k) \pmod{26}$

We will see again at midnight my baby (plaintext without space)

Giving space and punctuation mark;

We will see again at mid night my baby.

3.2.3 USING SUBSTITUTION CIPHER IN ENCRYPTING PLAINTEXT TO CIPHERTEXT

The transformation can be represented by aligning two alphabets, the cipher alphabet is the plain alphabet rotated left or right by some number of positions.

For example; A left rotation of three places

The key here is 3. That is $k=3$

Plaintext: a b c d e f g h I j k l m n o p q r s t u v w x y z

Rotated by shift 3

The Ciphertext: D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

Plaintext is written in lower case and ciphertext is written in upper case

The encryption can also be represented using modular arithmetics by first transforming the letter into numbers, according to the scheme.

A=0,B=1,C=2,-----,Z=25

Plaintext: a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
3	3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<hr/>																
3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
<hr/>																
q	s	t	u	v	w	x	y	z								
17	18	19	20	21	22	23	24	25								
3	3	-	-	-	-	-	-	-								
<hr/>																
20	21	22	23	24	25	26	27	28								

There is no 27 and 28 in the English alphabet, and since am using 0 to 25, that means

$$26-26=0 \rightarrow A, 27-26=1 \rightarrow B, 28-26= 2 \rightarrow C.$$

Ciphertext: D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

$$D_k(x) = (x+k) \text{ mod } 26$$

3.2.4 DECRYPTING OF SUBSTITUTION CIPHER

Giving the ciphertext:

MGZVYZLGHCMHJMYXSSFMNHAHYCDLMHA

By rearranging the 26 alphabet

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
M	G	Z	V	Y	Z	L	G	H	C	M	H	J	M	Y	X	S

r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
S	F	M	N	H	A	H	Y	C	D	L	M	H	A			

The plaintext a=M, b=G, c=Z,-----,z=C,-----,e=A.

Ciphertext : WKH TXLFN EURZQ IRA MXPSV RYHU WKH ODCB GRJ

Using mathematical model where $W=22, K=10, -, -, -, -, -, -, -, -, -$. The chosen key is 3

22	10	7	19	23	11	5	13	4	20	17	25	16	8	17	0	12	23	15	18
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
19	7	4	16	20	8	2	10	1	17	14	21	13	5	14	23	9	20	12	15
21	17	24	7	20	22	10	7												
3	3	3	3	3	3	3	3												
18	14	21	4	17	19	7	4												

The plaintext is: thequickbrownfoxjumpoverthelazydog (before punctuation mark).

After punctuation mark: the quick brown fox jump over the lazy dog.

Decrypting a net work recharge card.

My dear Blessing, please recharge your phone with this token of #200.00 card.

The card is encrypted and sent over the net work.

Ciphertext : ABDDFGIIEGEICB

Knowing the code, she decrypt to have this, by first rearranging the ciphertext.

Rearranging the ciphertext to plaintext

Using 0-25, where a A =0,B=1

F=5, B=1, D=2, D=2, F=5, G=6,I=8, I=8, E=4,G=6, E=4,I =8, C=2, B=1

The recharge number is: 5122568846821.

3.3 CRYPTANALYSIS

Cryptanalysis is the method of breaking ciphertext to plaintext. The general assumption that is usually made is that the opponent, knows the cryptosystem being used. This is known as Kerckhoff's principle. But my goal in designing a cryptosystem is to obtain security while assuming that Kerckhoff's principle holds. There are different types of attack model on cryptosystem. The most common type of attack model are as follows;

- (1) **Ciphertext only attack:** Here the opponent possesses a string of ciphertext y .
- (2) **A known plaintext attack:** the opponent possesses a string of plaintext, x , and the Corresponding ciphertext, y .

(3) **The chosen plaintext attack:** The opponent obtains temporary access to the encryption machine. Hence he can choose a plaintext string, x , and construct the corresponding ciphertext string y .

(4) **The chosen ciphertext attack :** Here the opponent has obtained temporary access to the decryption machine. Hence he can chose a ciphertext string, y , and construct the corresponding plaintext, x .

In each case, the objective of the adversary (opponent) is to determine the key that is used. This will allow the opponent to decrypt a string, and further' to decrypt any additional ciphertext string that are encrypted using the same key.

3.3.1 CRYPTANALYZING THE SUBSTITUTION CIPHER

The substitution cipher can be cryptanalyzed in a ciphertext only scenario.

Cryptanalization is the method or means of breaking a ciphertext into a plaintext. Cryptanalysis is done for security purpose.

3.3.2 SITUATIONS CONSIDERED

Two situations are considered:

- (1) An attacker knows (or guesses) that some sort of substitution cipher has been used, but not specifically that, it is a substitution scheme.
- (2) An attacker knows that a Caesar cipher (substitution) is in use, but does not know the shift value.

In the first case, the cipher can be broken using general substitution cipher, such as frequency analysis or pattern words. The distribution of letters in English

language text can be used, the shift “rotates” this distribution, and it is possible to determine the shift by examining the frequency graph.

Another way of approach is to match up the frequency distribution of the letters.

By graphing the frequency of letters in the ciphertext, and by knowing the expected distribution of these letters in the original language of the plaintext, a person can easily spot the value of the shift by looking at the displacement of a particular features of the graph. This is known a frequency analysis.

Many techniques of cryptanalysis use statistical properties of the English language. Relative frequencies of the 26 letters in novels, magazines and newspaper are Partition and grouped into three. The estimates were obtained by Beker and Piper.

- (1) It is believed that in every novel the letter E always have the highest frequency, followed by t and so on.

Letters arranged in level of their occurrence are as follows:

E, T, A, O, I, N, S, H, R, D, L, C, U, M, F, G, Y, P, B, V, K, J, X, Q, Z

- (2) The 30 most common bigram or digram are:

TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU,
EA, NG,

AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF.

- (3) The twelve most common trigrams are:

THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH.

In cryptanalysis, frequency analysis is the study of the frequency of letters or group of letters in a plaintext. The method is used as an aid to breaking classical cipher (wide ciphers).

Frequency analysis is based on the fact that, in any given stretch of written Language, certain letters and combination of letters occurs with varying frequencies.

The characteristics distribution of letters that is roughly the same for almost all samples of that language, given for instance, a section of English language, e tends to be very common, while x is very rear as explained earlier.

CHAPTER FOUR

FREQUENCY ANALYSIS FOR SUBSTITUTION CIPHER

In a simple substitution cipher, each letter of the plaintext is replaced with another, and a particular letter in the plaintext will always be transformed into the same letter in the ciphertext.

(4.1) CRYPTANALYSIS OF SUBSTITUTION CIPHER

considering the cipher text obtained from substitution cipher.

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

NDIFEFMZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

NZUCDRJXYYSMRTMEYIFZWDYVZVYZVYFZUMRZCRWNZDZJJ

XZWGCHSMRNMDHNCMFQCHZJMXJWIEJYUCFWDJNZDIR

Suppose Juliet has intercepted the cryptogram above, and it is known to be ... encrypted using a simple substitution cipher.

4.2 FREQUENCY OF OCCURRENCE OF 26 ALPHABET CIPHER TEXT LETTER

Letter	Frequency	Letter	Frequency
<u>A</u>	<u>0</u>	<u>N</u>	<u>9</u>
<u>B</u>	<u>1</u>	<u>O</u>	<u>0</u>
<u>C</u>	<u>15</u>	<u>P</u>	<u>1</u>
<u>D</u>	<u>13</u>	<u>Q</u>	<u>4</u>
<u>E</u>	<u>7</u>	<u>R</u>	<u>10</u>
<u>F</u>	<u>11</u>	<u>S</u>	<u>3</u>
<u>G</u>	<u>1</u>	<u>T</u>	<u>2</u>
<u>H</u>	<u>4</u>	<u>U</u>	<u>5</u>
<u>I</u>	<u>5</u>	<u>V</u>	<u>5</u>
<u>J</u>	<u>11</u>	<u>W</u>	<u>8</u>
<u>K</u>	<u>1</u>	<u>X</u>	<u>16</u>
<u>L</u>	<u>0</u>	<u>Y</u>	<u>10</u>
<u>M</u>	<u>16</u>	<u>Z</u>	<u>20</u>

The most frequent letter is Z =20

Recall that uppercase letters are used to denote ciphertext, lower case letters are used to denote plaintext.

$Z \sim e$ is used to express a guess that cipher text letter Z represents plaintext e in the English language frequency properties. Having this, say, $dk(Z) = e$. The remaining ciphertext characteristics that occur at least ten(10) times each are M,C,D,F,J,R,Y. We might expect that these letters are subset of t,a,o,i,n,s,h,r, but the frequencies really do not vary enough to tell us what the corresponding might be.

At these stage, I will look at digrams having $-Z$ or $Z-$, since Z decrypt to e . $ZM=1, ZR=2, DZ=4, ZD=0, ZP=1, RZ=2, ZW=4, WZ=0, ZC=2, CZ=0, ZK=1, ZU=3, UZ=0, ZV=1, NZ=3, HZ=2, XZ=2, FZ=2, ZJ=2$.

The most common digram here is $DZ=4$ and $ZW=4$;

$ZU=3$ and $NZ=3$

$RZ, HZ, XZ, FZ, ZR, ZV, ZC, ZD, ZJ$ (all occurs twice each).

Taking $ZW=4$ but $WZ=0$ and W occurs less often than many other characters, i guess that $dk(W)=d$.

DZ occurs 4 times and ZD occurs twice. $dk(D)$ might decrypt either (r, s, t) but I cant guess the right (letter).

Now having $Z \sim e$, $W \sim d$, I check the ciphertext for trigram having Z and W .

$ZRW=1$

RZW=1

RW=2

The next is to check common digram for plaintext letter other than ed, cause there is d already, there is nd and R is a common letter in the ciphertext. That means $R \sim_n dk(R) = n$, as the most likely possibility.

Now i have;

YIFQFMZRWQFYVECFDZPCVMRZWNMDZVEJBTXCDDMJUNM

XZ

----- e n d ----- e ----- n e d ----- e -----

- e

NDIFEFMZCDMQZKCEYFCJMYRNCWJCSZREXCHZ

----- e ----- e ----- n - d ----- e

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

- e - - n - - - - - n - - - - - e d - - - e - - e - - n e - n d - e - e - -

XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

- e d - - - - - n - - - - - e - - - e d - - - - - d - - - e - - n

Notice that from the ciphertext the letter N is always before or after W,R,Z, so the next step is to try $dk(N) = ?$ N occurs 9 times from the common plaintext segment $N \sim h$, $dk(N) = h$. NZ is a common digram, occurs three times and ZN does not. With this guess the plaintext ne-nd-e

suggest ne-ndhe and C in ciphertext suggest that $dk(C)=a$. In cooperating these guess, it becomes;

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUJ

----- e n d ----- a - - e - a - - n e d h - - e - - - - a - - - -

NDIFEFMDZCMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

h - - - - - e a - - e - a - - - a - - - n h a d - a - e n - - a - e - h - - e

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

h e - a - n - - - - - n - - - - - e d - - - e - - - e - - n e a n d h e - e - -

XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

- e d - a - - - n h - - - h a - - - a - e - - - e d - - - - - a - d - - h e - - - n

Going back to frequency table, the second most common letter is M=16. Checking the plaintext and ciphertext substituted cipher RNM,NRZ,MZ,they all contain M, so I guess M in the ciphertext segment RNM,which I believe decrypt to nh-,suggest that h- begins a word, so M is probably representing a vowel.

Recall that M=16 and i have a, and e, i guess $M \sim I$, so $dk(M)=I$ or o, checking the ciphertext CM occurs, and ai is a much more likely digram than ao, so the cipher digram CM suggests that i use $dk(M)=I$ at first.

The guess now is

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

---- iend ---- a-i-e-a-ined hi-e-----a---i-

NDIFEFMDZCDMQZKCEYFCJMRNCWJCSZREXCHZUNMXZ

h-----i-ea-i-e-a---a-ihad-a-en--a-e-hi-e

NZUCDRJXYYSMRTMEYIFZWDYVZVFZUMRZCRWNZDZJJ

he-a-n-----in-i---ed---e--e-ineand he-e--

XZWGCHSMRNMDHNCFQCHZJMXJZWIEJYUCFWDJNZDIR

-ed-a--inhi--hai--a-e-i--ed-----ad--he--n

Taking a good look at the ciphertext and plaintext, we can suggest F=11 and it comes before

or after M,C,Z,W,it can decrypt say $dk(F)=r$.

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

--r-riend-r--ari-e-a-ined hi-e-----a---i-

NDIFEFMDZCMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

h--r-ri-ea-i-e-a--ra-i-nhad-a-en--a-e-hi-e

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

he-a-n-----in-i---red---e--re-ineand he-e--

XZWGCHSMRNMDHNCFQCHZJMXJZWIEJYUCFWDJNZDIR

-ed-a--inhi--hair-a-e-i--ed-----ard--he--n

Having the cipher character D, J, Y, the subset of the plaintext of t,o,s, but

i can't guess which is suitable for each. Trying the most common digram

and trigram for D,J,Y,

MDZ=3,JMR=1,NMD=2,MD=4,DM=1,JM=2

Now am able to suggest meaningful word .see ; Qriend can be guess friend, dk(Q)=f, dk(D)=s,

dk(Y)=o

Now D~s, Y~o, Q~f, the most common letter D,J,Y, with plaintext {s, t ,o} will now be

dk(D)=s, dk(Y)=o, dk(J)=t.

Now I have;

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

o-rfr iendfro-- arise-a-i ned hise--t--- ass - i t

NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

hs-r- riseasi fe-a-orati onhadta-en- -a -e- hi -e

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

he -asnt-oo-i n-i -o-red so -e-ore- i neandhesett

XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

-ed - a-- in his- hair fa-eti -ted--to- ardsthes-n

Following the ciphertext and plaintext ,i can easily decrypt the whole

sentence dk(X)=l,

dk(G)=b, dk(I)=u, dk(U)=w, dk(E)=p, dk(K)=v, dk(S)=k, dk(P)=x,
dk(H)=c, dk(V)=m, dk(B)=y,

dk(T)=g

Now i have;

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

our friend from paris examined his empty glass with

NDIFEFMZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

h surprise as if evaporation had taken place while

NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

he was not looking i poured some more wine and he sett

XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

led back in his chair face tilted up towards the sun

Now decryption is complete, the plaintext is

4.3 THE DECRYPTED TEXT

Our friend from paris examined his empty glass with surprise as if evaporation ha

D taken place while he was not looking i poured some more wine and settled back in

his chair face tilted up towards the sun (with out space and punctuation mark)

Our friend from paris examined his empty glass with surprise, as if

evaporation had taken place, while he wasn't looking, i poured some

more wine and settled back in his chair, face tilted up towards the sun.

As you can see, the guess where all correct for our Paris friend.

This would not always be the case, the variation in statistics for individual plaintexts can mean that initial guess are incorrect. It is also possible that the plaintext does not exhibit the expected distribution of letter frequency.

CHAPTER FIVE

SUMMARY, FINDINGS AND CONCLUSION

5.1 SUMARRY

The Substitution Algorithm Cipher which is a cryptosystem is a means of permuting letters of the twenty six characters use in securing message from being intercepted. Even when intercepted it can not be easily deciphered.

5.2 FINDINGS

Substitution cipher keys are very difficult to calculate and compute. Substitution cipher is difficult because the encryption and decryption involves permutation of alphabetic characters along with frequency analysis.

5.3 CONCLUSION

Substitution Algorithm is a means by which each letter in plaintext is replaced by a letter with some fixed number of positions down the alphabet. Substitution cipher is a confidential way of securing message for security purpose. Say, writing plaintext to ciphertext, this is by changing the order of the letters of the alphabets, that, not a word could be made out. If any one wishes to decipher these and get at meaning, he must go through permuting of letter of the alphabet or knowing the keys.

At first, substitution cipher looks as if writing in unknown foreign language. Substitution cipher is in use today in children's toys such a secret decoder rings, novels have been written that omit the letter "e" altogether – a form of literature known as a Lipogram puzzle, "cryptograms" in newspapers are examples of substitution cipher.

REFERENCE

- 1 H. BEKER AND F. PIPER. Cipher system, The protection of communications. John Wiley and sons, 1982.
- 2 A. BEIML AND B. CHOR. Interaction in key distribution scheme. Lecture note in computer science, 1994.
- 3 M. BELLARE, J.KILIAN AND P. ROGAWAY. The security of the cipher Block chaining message authentication code. Journal of computer and system science, 2000.
- 4 N. KOBLITZ. Elliptic curve cryptosystem. Mathematics of computer, 1987.
- 5 A.SHAMIR. How to share a secrete. Communications of the ACM,1979.
- 6 C.E. SHAWNON. A mathematics theory of communication. Bell system Technology journal,1948.
- 7 S. S. WAGSTAFF, JR. Cryptanalysis of Numbers Theoretic cipher. Chapman & Hall/ CRC, 2003.
- 8 F.PIPER AND S. MURRHY CRYPTOGRAPHY. A very short introduction.
Oxford. 2002
- 9 K. NYBERG. Differentially uniform mapping for cryptography. Lecture note
in computer science. 1994.
- 10 R. A. RUEPPEL AND P. C. VAN OORSCHOT. Modern key agreement technic. Computer communication, 1994.

- 11 R. SAKAI, K. OHGISHI AND M. KASAHARA. Cryptosystem based on pairing. Present at the symposium on cryptography and information security, Okinawa, Japan.
- 12 Dr SALOMON. Data privacy and security, 2003.
- 13 S.Y. YAN. Number theory for computing springer, 2000.
- 14 D. KAHN. The codebreaks. Scriber, 1996.
- 15 H.M. HEYS AND S.E TAVARES. Substitution permutation networks resistant to differential and linear cryptanalysis. Journal of cryptology, 1996.