Real-Time Malware Uniform Resource Locator Detection System Based on Multi-Layer Perceptron Neural Networks

Mohd Taufik Abdullah[1, 2], Morufu Olalere[2], Ramlan Mahmod[2] and Azizol Abdullah[2]
[1]Cyber Security Science Department Federal University of Technology Minna, Nigeria
[2]Information Security research Group Faculty of Computer Science and information Technology, Universiti Putra Malaysia, Malaysia
taufik@upm.edu.my

Abstract: With the rapid proliferation of Internet technologies, mobile devices, and web applications, attackers now use the Web as a vector for introducing malware into enterprise networks. This change in attack vector has forced many organizations to subscribe to blacklisting services of malware Uniform Resource Locators (URLs) which are provided by a range of techniques including manual submission of suspected malware URLs and honeypots. However, the blacklist approach to detect malware URLs is no longer sufficient as many new malware URLs are not blacklisted immediately they are launched on the Internet. To address this problem, there is a need for real-time malware URL detection system that will be able to detect malware URL on the fly. The few previous studies that addressed this problem was unable to achieved high level of accuracy and none of the study has proposed Artificial Neural Network (ANN) based on detection system. Consequently, this study proposed a real-time detection system that is based on Multi-Layer Perceptron (MLP) Neural Networks. With the application of WEKA data mining tool, the performance of the proposed detection system was tested using an existing dataset comprises of malware URLs and benign URLs. The proposed detection system in this paper outperformed previous studies with accuracy of 97.33%.

Keywords: malware URL, real-time malware URL detection, artificial neural network, multi-layer perceptron, blacklisting

1. Introduction

With the rapid proliferation of Internet technologies, mobile devices, and web applications, attackers now use the Web as a vector for introducing malware into enterprise networks through employee's mobile devices in an environment such as Bring Your Own Device (BYOD). The personal mobile device is used to access a web application through the Internet either by typing a URL in the web browser or by clicking a URL link to the web application. In any case, URLs serve as a means of obtaining access to web applications, thus making it an exploitable tool for attackers to infect malware into the device of their victim (Olalere et al., 2016).

However, this change in attack vector has forced many organisations to subscribe to blacklisting services of malware URLs which are provided by a range of techniques including manual submission of suspected malware URLs and honeypots. With 571 new websites available on the Internet per minute (CoNet, 2016), the blacklist approach to detect malware URLs is no longer sufficient as many new malware URLs are not blacklisted immediately they are launched on the Internet. More so, since the blacklist is created by volunteer experts, human error in classification is unavoidable. Exact matching in blacklisting also renders it easy to be evaded (Choi, Zhu and Lee, 2014)

To address blacklisting challenges, a real-time malware URLs detection system is necessary. Consequently, previous studies have proposed detection systems that rely on machine learning algorithms. These systems detect malware URLs as soon as they are encountered, without having to visit the blacklist server. To the best of the knowledge of the authors, little work has been done in the area of real-time malware URL detection system. A recent survey (Patil and Patil, 2015), concerning malicious web page detection techniques reported works by (Sayamber and Dixit, 2014) and (Choi et al., 2011) as the only studies on malware URL detection systems. However, none of these studies has proposed ANN based malware URL detection system. Perhaps, this explains why the performance accuracy of the previous studies are relatively low. In this paper, a real-time malware URL detection system which is based on an MLP Neural Network is proposed.

2. Our proposed malware URL detection system

As shown in figure 1, our proposed detection system comprises of three units namely, input unit, detection unit and output unit. A URL to be tested is inputted into the classification unit through input unit. The detection unit which contains a well-trained MLP NN then perform classification of the inputted URL. The output (weather benign or Malware URL) of the classification is outputted through output unit. If the inputted URL is classified as malware URL then malware URL is detected otherwise, the inputted URL is benign URL.