OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

# DETECTION OF PHISHING WEB SITES BASED ON FEATURE CLASSIFICATION AND EXTREME LEARNING MACHINE

**Amol D Wakhare[1],V S Karvande[2]**

*P.G. Student, Computer Science & Engineering, Everest Educational Society's Group of Institutions, Maharashtra, India[1]*

*HOD, Computer Science & Engineering, Everest Educational Society's Group of Institutions, Maharashtra, India[2]*

---

***Abstract:*** ***Phishing website extracts the victims confidential data by averting them to surf a fake website page that resembles an honest to goodness one a type of criminal acts through the internet. Phishing is considered to be the most common yet the most hazardous attacks among all the cybercrimes. Phishing websites are unpredictable and its aim is to steal sensitive data of individuals or the organizations in order to conduct transactions. The purpose of conducting the study is detecting the fake web sites. Web pages differ with the feature set and thus, we use it as our prime weapon to prevent the phishing attacks. We have suggested some new rules to have efficient feature classification. The incitement of the proposed work is to perform Extreme Learning Machine (ELM) based classification for various features of the website. We thus come up with the model which uses machine learning techniques for detecting phishing web sites. ELM classification algorithm has a high rate of accuracy of detecting phishing websites as compared to other machine learning algorithms.***

*Keywords: Phishing, Extreme Learning Machine, Feature Classification*

---

## I INTRODUCTION

Technology is growing rapidly day-by-day and with this rapid growing technology internet has become an essential part of humans daily activities. Use of internethas grown due to the rapid growth of technology and intensive use of digital systems and thus data security has gained great importance. The primary objective of maintaining security in information technologies is to ensure that necessary precautions are taken against threats and dangers likely to be faced by users during the use of these technologies. Phishing is the fraudulent attempt to obtain sensitive information such as user-names, passwords and credit card details by disguising as a trustworthy entity in an electronic communication. Typically carried out by email spoofing or instant messaging, it often directs users to enter personal information at a fake website, the look and feel of which is identical to the legitimate site. Information security threats have been seen and developed through time along development in the internet and information systems. The impact is the intrusion of information security through the compromise of private data, and the victims may lose money or other kinds of assets at the end. Internet users can be affected from different types of cyber threats such as private information loss, identity theft, and financial damages. Hence, using of the internet may suspect for home and official environments. Identify and defend against privacy leakage efficient analytical tools are required for users to reduce security threats. Effective systems that can improve self-intervention must be formed using artificial intelligence-based information security management system at the time of an attack. Phishing is an Internet-based attack that seduces end users to visit fake websites and give away personal information such as user id and password. Phishing web pages are formed by fraudulent people to copy a web page from an original one. These phishing web pages are very similar to the original ones. Technical tricks and social engineering are extensively joined together for beginning a phishing attack. An important view of online security is to protect users from phishing attacks and fake websites. Intelligent methods can be used to develop fake web pages. For this reason internet users whether have enough experience in information security or not might be cheated. Phishing attacks can be launched via sending an e-mail that seems to be sent from a trusted public or private organization to users by attackers. Attackers get the users to update or verification their information by clicking a link within the e-mail. Other methods such as file sharing, blogs, and forums can be used by attackers for phishing. There are many ways to fight phishing including legal solutions, education, and technical solution. A significant number of studies on the phishing have been done.

Motivation

Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data, for example, passwords, account points of interest, or MasterCard numbers. In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavors in messages and identifying phishing substance on sites, phishers think of new and half breed strategies to go around the accessible programming and systems. Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble delicate and individual data, for example, passwords and charge card subtle elements by taking on the appearance of a dependable individual or business in an electronic correspondence. Phishing makes utilization of spoof messages that are made to look valid and implied to be originating
from honest to goodness sources like money related foundations, e-commerce
destinations and so forth, to draw clients to visit fake sites through joins gave in the
phishing email.

## II LITERATURE SURVEY

As a crime of employing technical means to steal sensitive information of users, phishing is currently a critical threat facing the Internet, and losses due to phishing are growing steadily. Feature engineering is important in phishing website detection solutions, but the accuracy of detection critically depends on prior knowledge of features. Moreover, although features extracted from different dimensions are more comprehensive, a drawback is that extracting these features requires a large amount of time. To address these limitations, we propose a multidimensional feature phishing detection approach based on a fast detection method by using deep
learning (MFPD). In the first step, character sequence features of the given URL are xtracted and used for quick classification by deep learning, and this step does not require third-party assistance or any prior knowledge about phishing. In the second step, we combine URL statistical features, webpage code features, webpage text features and the quick classification result of deep learning into multidimensional features. The approach can reduce the detection time for setting a threshold. Testingon a dataset containing millions of phishing URLs and legitimate URLs, the accuracy reaches 98.99%, and the false positive rate is only 0.59%. By reasonably adjusting the threshold, the experimental results show that the detection efficiency can be improved. [1]

Phishing detection is recognized as a criminal issue of Internet security. By deploying a gateway anti-phishing in the networks, these current hardware-based approaches provide an additional layer of defense against phishing attacks. However, such hardware devices are expensive and inefficient in operation due to the diversityof phishing attacks. With promising technologies of virtualization in fog networks, an anti-phishing gateway can be implemented as software at the edge of the network and embedded robust machine learning techniques for phishing detection. In this paper, we use uniform resource locator (URL) features and web traffic features to detect phishing websites based on a designed neuro-fuzzy framework (dubbed Fi-NFN). Based on the new approach, fog computing as encouraged by Cisco, we design an anti-phishing model to transparently monitor and protect fog users from phishing attacks. The experiment results of our proposed approach, based on a largescale dataset collected from real phishing cases, have shown that our system can effectively prevent phishing attacks and improve the security of the network. [2]

Phishing is one of the most harmful social engineering techniques to subdue end users where threat actors find a chance to gain access to critical information systems. A common approach in phishing is through the use of e-mail communication with an embedded hyperlink. The detection and mitigation of phishing attacks a grand challenge due to the complexity of current phishing attacks. Existing techniques are often too time-consuming to be used in the real world in terms of detection and mitigation time. Likewise, they employ static detection rules that are not effective in the real world due to the dynamics of phishing attacks. In this paper, we present PhishLimiter, a new detection and mitigation approach, where we first propose a new technique for Deep Packet Inspection (DPI) and then leverage with Software-Defined Networking (SDN) to identify phishing activities through e-mail and webbased communication. The proposed DPI approach consists of two components, phishing signature classification and real-time DPI. Based on the programmability of SDN, we develop Store and Forward (SF) mode and the Forward and Inspect (FI) mode to direct network traffic by using an Artificial Neural Network (ANN) model to classify phishing attack signatures and design the real-time DPI so that PhishLimiter can flexibly address the dynamics of phishing attacks in the real world. PhishLimiter also provides better network traffic management for containing phishing attacks since it has the global view of a network through SDN. Furthermore, we evaluate Phish Limiter using a real-world testbed environment and datasets consisting of real-world email with embedded links. Our extensive experimental study shows that Phish Limiter provides an effective and efficient solution to deter malicious activities. [3]

Phishing attacks continue to pose a major threat for computer system defenders, often forming the first step in a multi-stage attack. There have been great strides made in phishing detection; however, some phishing emails appear to

pass through filters by making simple structural and semantic changes to the messages. We tackle this problem through the use of a machine learning classifier operating on a large corpus of phishing and legitimate emails. We design SAFEPC (Semi-Automated Feature generation for Phish Classification), a system to extract features, elevating some to higher level features, that are meant to defeat common phishing email detection strategies. To evaluate SAFE-PC, we collect a large corpus of phishing emails from the central IT organization at a tier-1 university. The execution of SAFE-PC on the dataset exposes hitherto unknown insights on phishing campaigns directed at university users. SAFE-PC detects more than 70% of the emails that had eluded our production deployment of Sophos, a state-of-the-art email filtering tool. It also outperforms SpamAssassin, a commonly used email filtering tool. We also developed an online version of SAFE-PC, that can be incrementally retrained with new samples. Its detection performance improves with time as new samples are collected, while the time to retrain the classifier stays constant. [4]

Recent years have witnessed the increasing threat of phishing attacks on mobile computing platforms. In fact, mobile phishing is particularly dangerous due to the hardware limitations of mobile devices and mobile user habits. In this paper, we did a comprehensive study on the security vulnerabilities caused by mobile phishing attacks, including the web page phishing attacks, the application phishing attacks, and the account registry phishing attacks. Existing schemes designed for web phishing attacks on PCs cannot effectively address the various phishing attacks on mobile devices. Hence, we propose MobiFish, a novel automated lightweight anti-phishing scheme for mobile platforms. MobiFish verifies the validity of web pages, applications, and persistent accounts by comparing the actual identity to the claimed identity. MobiFish has been implemented on a Nexus 4 smartphone running the Android 4.2 operating system. We experimentally evaluate the performance of MobiFish with 100 phishing URLs and corresponding legitimate URLs, as well as phishing apps. The results show that MobiFish is very effective in detecting phishing attacks on mobile phones. [5]

## III PROBLEM STATEMENT

Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. On account of the last and in addition ambiguities in arranging sites because of the intelligent procedures programmers are utilizing, some keen proactive strategies can be helpful and powerful tools can be utilized. Several conventional techniques for detecting phishing website have been suggested to cope with this problem. However, detecting phishing websites is a challenging task, as most of these techniques are not able to make anaccurate decision dynamically as to whether the new website is phishing or legitimate. Several conventional techniques for detecting phishing website have been suggested to cope with this problem. To propose an intelligent model for detecting phishing web pages based on Extreme Learning Machine.

## IV SYSTEM DESIGN

The proposed methodology which imports dataset of phishing and legitimate URLs from the database and the imported data is preprocessed. Detecting phishing website is performed based on four categories of URL features: domain based, address based, abnormal based and HTML, JavaScript features. These URL features are extracted with processed data and values for each URL attribute are generated. The analysis of URL is performed by machine learning technique which computes range value and the threshold value for URL attributes. Then it is classified into phishing and legitimate URL. The attribute values are computed using feature extraction of phishing websites and it is used to identify the range value and threshold value. The value for each phishing attribute is ranging from f-1, 0, 1g these values are defined as low, medium and high according to phishing website feature. The classification of phishing and legitimate website is based on the values of attributes extracted using four types of phishing categories and a machine learning approach.

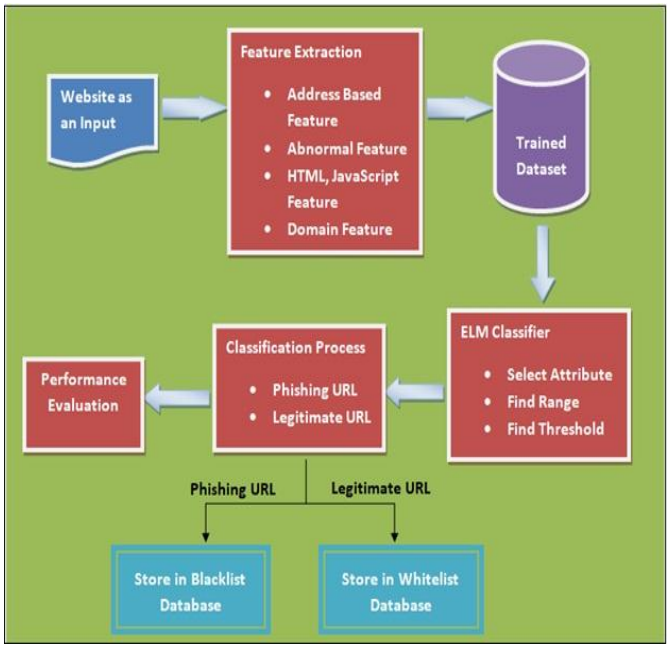


***Figure 4.1: Architecture Diagram***

4.1 DATA FLOW DIAGRAMS


***Figure 4.2: Level 0 Data Flow Diagram***

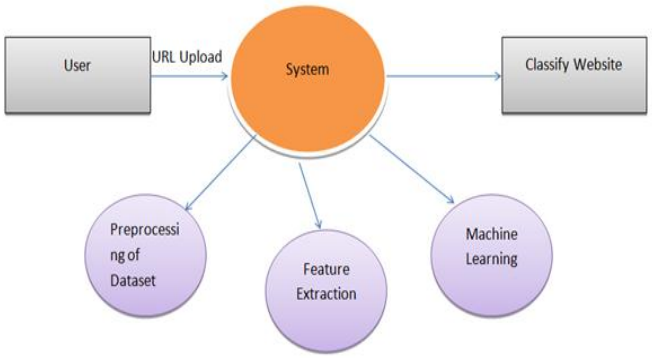


*Figure 4.3: Level 1 Data Flow Diagram*

## ALGORITHM DETAILS

### Extreme Learning Machine (ELM)

Extreme Learning Machine (ELM) is proposed as a single hidden layer feed-forward artificial neural network (ANN) model which ensure a high-performing learning and parameters such as threshold value, weight and activation the function must have appropriate values for the data system which is to be modeled. In ELM learning, the parameters are gradient-based, where the input weights are randomly selected while the output weights are analytically calculated. For the sake of activating the cells in the hidden layer of ELM, a linear function as well as non-linear (sinus, sigmoid, Gaussian), and the non-derivable or discrete activation functions canbe used.

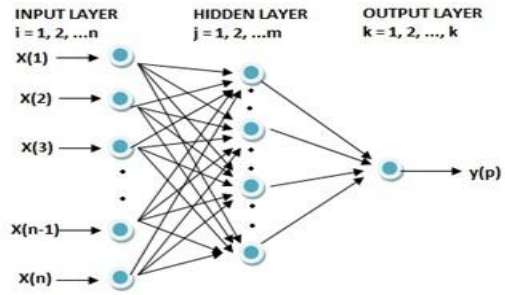


*Figure 4.4: ELM network model*

## V RESULTS

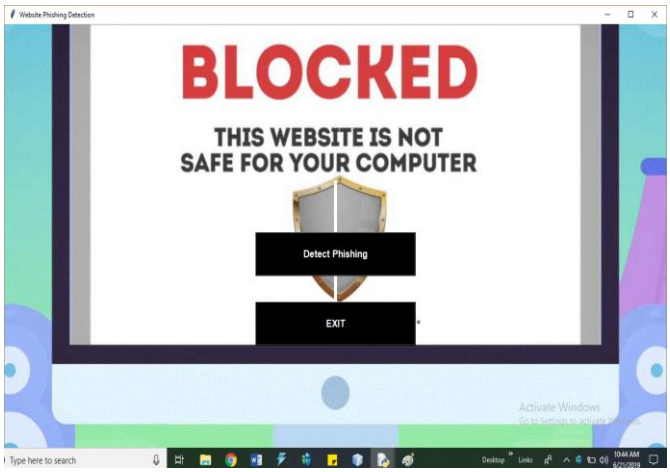


*Figure 5.1: Home Page*

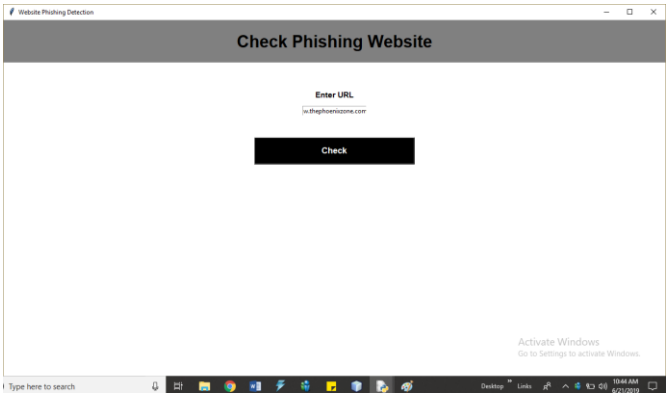


*Figure 5.2: Enter Website*

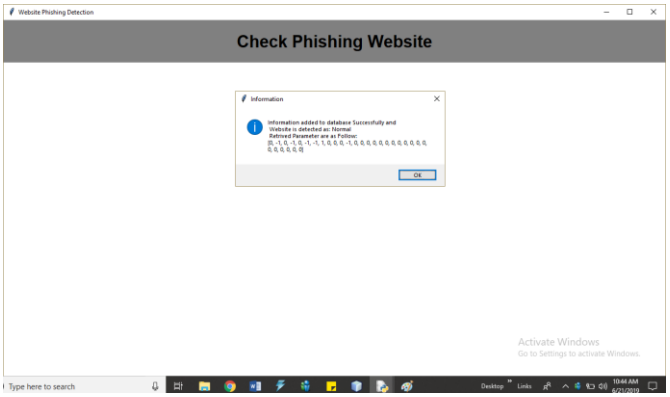


*Figure 5.3: Normal Website Result*

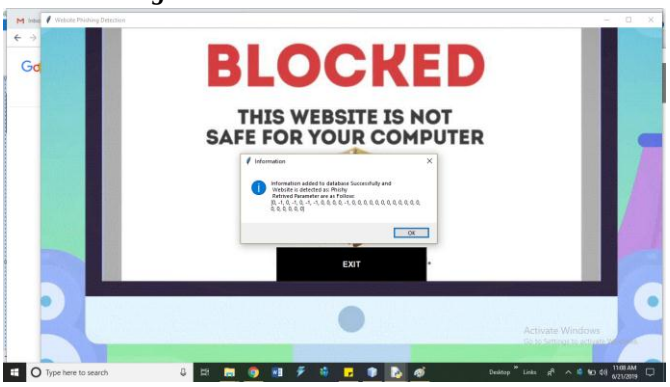


*Figure 5.4: Phishing Website Detection*

## VI CONCLUSIONS

Systems varying from data entry to information processing applications can be made through websites. The entered information can be processed; the processed information can be obtained as output. Nowadays, web sites are used in many fields such as scientific, technical, business, education, economy, etc. Because of this intensive use, it can be also used as a tool by hackers for malicious purposes. One of the malicious purposes emerges as a phishing attack. A website or a web page can be imitated by phishing attacks and using various methods. Some information such as users credit card information, identity information can be obtained with these

fake websites or the web pages. The purpose of the application is to make a classification for the determination of one of the types of attacks that cyber threats called phishing.
Extreme Learning Machine is used for this purpose. In this study, we used a data set taken from UCI website. In this dataset, input attributes are determined in 30, and the output attribute is determined in 1.Input attributes can take 3 different values which are 1, 0, and -1.Output attribute can take 2 different values which are 1, and -1. As a result of the study, the average classification accuracy was measured is 95.34%.

## REFERENCES

[1] Peng Yang, Guangzhen Zhao, Peng Zeng, "Phishing Website Detection based on Multidimensional Features driven by Deep Learning", IEEE Access, 2018.
[2] Chuan Pham, Luong A. T. Nguyenz, Nguyen H. Tran, Eui-Nam Huh, Choong Seon Hong, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks", IEEE Transactions on Network and Service Management, 2018.
[3] Tommy Chin, Kaiqi Xiong and Chengbin Hu, "PhishLimiter: A Phishing Detectionand Mitigation Approach Using Software-Defined Networking", IEEE Access, 2018.
[4] Christopher N. Gutierrezy, Taegyu Kimy, Raffaele Della Cortez, Jeffrey Averyyx, Dan Goldwassery, Marcello Cinquez, Saurabh Bagchiy, "Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks", Transactions on Dependable and Secure Computing, 2018.
[5] Longfei Wu, Xiaojiang Du, and Jie Wu, "Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms", IEEE Transactions on Vehicular Technology, 2018.
[6] Samuel Marchal, Giovanni Armano, Tommi Grondahl, Kalle Saari, NidhiSingh and N. Asokan, "Off-the-Hook: An Efficient and Usable Client-Side Phishing Prevention Application", Transactions on Computers, 2016.
[7] Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, "Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity", Open IEEE Access, 2017.
[8] Zuochao Dou, Issa Khalil, Abdallah Khreishah, Ala Al-Fuqaha, and Mohsen Guizani, "Systematization of Knowledge (SoK): A Systematic Review of Software Based Web Phishing Detection", Communications Surveys & Tutorials,2017.
[9] Ajaya Neupane, Nitesh Saxena, Jose O Maximo, and Rajesh Kana, "Neural Markers of Cybersecurity: An fMRI Study of Phishing, and Malware Warnings", Transactions on Information Forensics and Security, 2019.
[10] Xiayang Chen, Xingtong Liu, Lei Zhang, Chaojing Tang, "Optimal Defense Strategy Selection for Spear-Phishing Attack Based on a Multistage Signaling Game", IEEE Access, 2019.