

Deep Learning Based Phishing Websites Detection

Vaibhav Handge¹, Shubham Pokale², Saurabh Lavhate³, Shubham Nalkol⁴, Prof G. B. Kote⁵

Students, Department of Computer Engineering^{1,2,3,4}

Guide, Department of Computer Engineering⁵

Pravara Rural Engineering College, Loni, Maharashtra, India

Abstract: *Phishing is a crime that involves the theft of confidential user information. Those targeted by phishing websites include individuals, small businesses, cloud storage providers, and government organisations and websites. The majority of phishing prevention techniques involve hardware-based solutions, although software-based options are preferred due of cost and operational considerations. There is no answer to the problem of zero-day phishing assaults from the present phishing detection approaches since there is no solution to the problem. The Phishing Attack Detector based on Web Crawler, a three-phase attack detection system, was designed to handle these issues and accurately detect phishing incidences using a recurrent neural network in order to resolve these issues. It covers the input aspects of web traffic, web content, and Uniform Resource Locator (URL) based on the classification of phishing and non-phishing pages, as well as the output features of phishing and non-phishing pages.*

Keywords: Recurrent Neural Network, Deep Learning, illegitimate URLs, cyber attacks

I. INTRODUCTION

Phishing is a type of cybercrime in which a person impersonating a legitimate agency contacts a victim or target via email, phone, or text message in order to entice the person to provide information, such as personal identification information, banking and credit card information, and passwords, among other things. Phishing is a serious offence. When an attacker creates the appearance of a legitimate website, he or she is inviting visitors to visit the site in order to obtain personal information from them, such as their login, password, financial information, account details, national security identifier, and so on. Phishing is a new term that was coined by combining the words 'fishing' and 'phishing.' In some cases, the information gathered is used for possible target advertisements or even identity robberies and attacks (such as money transfers from one's bank account). The most often utilised attack method is to send e-mails, messages that might lead to data theft or the theft of personal information from the recipient. Account on a social networking site Daily, people make mistakes while entering passwords, credit cards, or other sensitive information onto websites. Attackers use bogus websites to provide upgrades to their websites, entice you to comply with your personal information, and then change your information. As long as you are submitting your personal information, the attackers will be able to successfully collect it on your server side, and they will be able to carry out the following move with your information and utilise it to further their nefarious objectives. Phishing is defined as the replication of a legitimate business website in order to steal private information from customers, such as usernames, passwords, and structured savings numbers. Phishing is defined as the theft of private information from customers. Mail spammers can be classified based on who they are trying to reach. The majority of telemarketers are spammers, who send a few hundred or a big number of e-mail messages to customers who have sent them spontaneously. Spammers are divided into the following categories, each of which continues to send messages at random intervals, albeit with little enthusiasm. Frequently, they spam or push items that are about completely unrelated themes. Some of the examples include sightings, authoritative reports, and comments regarding meetings. Phishing is not a new concept, but fraudsters, also known as phishers, have been increasingly adept at using it in recent years to steal your personal information and commit economic and social crimes in the process. In the last four to five years, there has been a dramatic increase in the frequency of phishing assaults. Phishing is a common practise that is simple to carry out once you get at your location. Phishing is a method of attracting a victim by submitting a faked link to a bogus website, which is known as social engineering. It is possible to find the faked connection on often visited online pages or to send it to the victim via email. A false website is created in the same manner as the actual website. The victim's request is thereby routed directly to the attacker's website rather than to the legitimate web server.

II. REVIEW OF LITERATURE

In this paper [1], we did a comprehensive study on the security vulnerabilities caused by mobile phishing attacks, including the web page phishing attacks. Author 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. Existing schemes designed for web phishing attacks on PCs cannot effectively address the various phishing attacks on mobile devices.

To [2] an online user into elicit personal Information. The prime objective of this review is to do literature survey on social engineering attack: Phishing attack and techniques to detect attack. The paper discusses various types of Phishing attacks such as Tab-napping, spoofing emails, Trojan horse, hacking and how to prevent them. Every organization has security issues that have been of great concern to users, site developers, and specialists, in order to defend the confidential data from this type of social engineering attack.

Commercial and retail account [3] holders at financial institutions of all sizes are under attack by sophisticated, Organized, well-funded cyber criminals. Anomaly detection solutions are readily available, are deployed quickly and immediately and automatically protect all account holders against all types of fraud attack with minimal Disruption to legitimate online banking activity. Implementing anomaly detection will not only meet FFIEC Expectations, it will decrease the total cost of fraud, and will increase customer loyalty and trust.

This paper [4] gives an in-depth analysis of phishing: what it is, the technologies and security Weaknesses it takes advantage of, the dangers it poses to end users. In this analysis I will explain the concepts and technology behind phishing, show how the threat is much more than just a nuisance or passing trend, and discuss how gangs of criminals are Using these scams to make a great deal of money. Unfortunately, a growing number of cyber-thieves are using these same systems to manipulate us and steal our private information.

Author suggest in this paper[6] a technique called optimum RT-PFL for classifying malicious URLs detected on the websites from non-malicious URLs. In order to generate feature components, the data set should both be encoded as lexical and host functions for the URL. The function extraction method extracts those features. Optimum URL

Functions are chosen according to the proposed selection process, namely the Rough Set Theory algorithm based on Gray Wolf Optimizer. This proposed algorithm will define a minimal reduction in the attributes from the highly effective data collection, which in turn enhances the efficiency of classification systems. In order to decide whether the approved URL is good or malicious, the URL should be inserted into the classifier. The classification of URLs depends on the newly formulated fuzzy logical approach to particle filtering. The following categories are strengthened with the detection of a large number of suspicious URLs from malicious pages.

This paper [7] provides a detailed empirical analysis on 1529,433 malicious URLs in the last two years. Author evaluate tactical actions of attackers with respect to URLs and extract common capabilities. Author then divide it into three usable pools, so that the compromise levels of unknown URLs are calculated. Author use a similarity matching technique to leverage detection speeds. Author assume that the attackers' normal URL manipulation behaviors will classify new URLs. This method covers a wide range of malicious URLs with limited function sets. The exactness of the proposed method is rational (up to 70 percent) and the approach requires only analysis of the attributes of URLs. During preprocessing this model can be used to assess if input URLs are friendly or to estimate if an input URL is malicious as a web filter or a risk scaler.

This paper's [8] objective is twice. First, author will talk in depth about the history of phishing attacks and the motivation of attackers. Then, the different forms of phishing attacks are taxonomied. Second, to protect users from phishing based on the attacks found in our fiscalonomics, our services will provide taxonomies of many solutions suggested in the literature. In addition, we addressed the effects of Internet of Things phishing attacks (IoTs). We conclude our paper on several still existing literary issues and challenges that are relevant for the fight against phishing threats.

In this paper [9], author suggest a new method for protecting against phishing attacks by automatically updating the white list of legit sites visited by the user. Our solution proposed has high detection and short access time. The browser warns users not to reveal personal details when they attempt to open a page that is not available in the white list. In addition, we verify the validity of a website with hyperlinks. This is done by extracting hyperlinks from your website source code and using the proposed phishing detection algorithm. Our experimental results show the proposed solution to phishing as it has a true positive rate of 86.02%, whereas a false negative level of less than 1.48% is very successful.

III. PROPOSED METHODOLOGY

As the underlying concept behind the development of such a system, it is important to note that providing security to a customer's financial information is critical. As a result, banks and other financial institutions provide a variety of security mechanisms to reduce the risk of unauthorised access to their online customer accounts.

Online banking has become more dependent on transactions conducted through various application gateways, making it imperative that these transactions be protected against fraud and identity theft.

3.1 Architecture

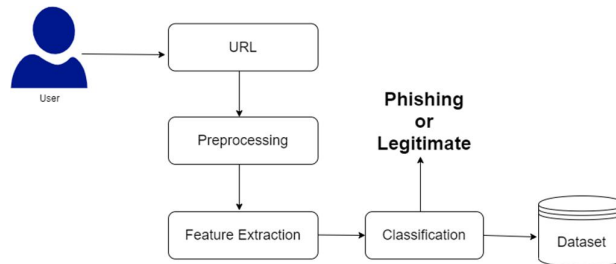


Figure 1: Proposed System Architecture

IV. CONCLUSION

Phishing is one of the most devastating types of web security risks available today. According to our research, we have developed a prediction model for the identification of Phishing websites, which is based on an analysis of the attributes of the attack. The Deep recurrent neural network's deep-seated learning model outperforms other machine learning models in terms of prediction and achieves the highest level of precision.

REFERENCES

- [1]. Surbhi Gupta et al., "A Literature Survey on Social Engineering Attacks: Phishing Attack," in International Conference on Computing, Communication and Automation (ICCCA2016), 2017, pp. 537-540.
- [2]. Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, "Phishing- Alarm: Robust and Efficient Phishing Detection via Page Component Similarity".
- [3]. Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen, "Web Phishing Detection Based on Graph Mining", Guardian Analytics, "A Practical Guide to Anomaly Detection Implications of meeting new FFIEC minimum expectations for layered security". Accessed: 08 Jan 2018.
- [4]. Ibrahim Waziri Jr., "Website Forgery: Understanding Phishing Attacks Nontechnical Countermeasures," in IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015, IEEE.
- [5]. Longfei Wu et al., "Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms," IEEE 2016, pp. 6678-6691.
- [6]. K. Rajitha and D. Vijayalakshmi, "Suspicious urls filtering using optimal rt-pfl: A novel feature selection based web url detection," in Smart Computing and Informatics, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer Singapore, 2018, pp. 227-235.
- [7]. S. Kim, J. Kim, and B. B. Kang, "Malicious url protection based on attackers' habitual behavioral analysis," Computers Security, vol. 77, pp. 790 - 806, 2018.
- [8]. B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," Telecommunication Systems, vol. 67, no. 2, pp. 247-267, Feb 2018.
- [9]. A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," EURASIP Journal on Information Security, vol. 2016, no. 1, p. 9, May 2016.