IT in Industry, Vol. 7, No.3, 2019

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

A Review on Anti-Phishing Framework

Preeti Chaudhary
Department of Comp. Sc. & Info. Tech.,
Graphic Era Hill University, Dehradun, Uttarakhand, India 248002,
priti.chaudhary1989@gmail.com

Abstract
Phishing is an assault that is typically carried out by combining foundations of social engineering with ever-evolving technical approaches. Phishing is also known as spear phishing. This masking of the phony site to work as if it were the real one induces the user to divulge their personal details such as the passwords and bank accounts associated with such accounts. As a result, in modern times, conducting an exhaustive study on previous attacks is obligatory in order to adequately prepare ourselves to avoid becoming victims of such dangers. The purpose of this study article is to provide a better knowledge on the working principles of such threats, to promote the development of anti-phishing measures in the future, and to provide a brief discussion on prior and ongoing attacks. The fundamental purpose of the work that is being given is to raise people's levels of awareness and teach them on how to protect themselves from attacks of this kind. In addition, the purpose of this assessment is to offer assistance to policy makers and software developers so that they may arrive at the best decisions and help create an environment free of viruses.

Keywords— Detection techniques, Phishing attacks, Prevention approaches, Security threats

Introduction

Phishing attacks are typically combined with the most common forms of social engineering assaults, which are carried out with the assistance of modern technology for the financial benefit of the individual conducting the attack. This tactic is utilized by the hacker, who then places the user in the precarious position of having to divulge his identity. Users who are unaware of this attack put themselves at risk of becoming victims, which could result in the disclosure of their financial statements, mobile OTPs, and mail ID passwords. These kinds of attacks are carried out against specific users with the purpose of gaining political advantage [1]. It has been observed that the frequency of this attack has dramatically increased in tandem with the fast proliferation of internet services that are being obtained by each household. Standard operating procedures and protocols have been devised and developed by the Department of Information Security in order to protect online information sharing and money laundering in an effort to combat the growing problem of cybercrime. Because of this, every organization has also established its own security procedures to not only avoid being in a scenario where they are under assault, but also to prevent the servers from being attacked in the future [2]. In this kind of scenario, the working personnel of an organization examine the weaknesses in their system on an annual basis and devise methods to prevent them. However, social engineering can be used to accomplish an infinite number of different types of attacks, all of which are successful in duping victims into disclosing their private information. The phishing assault is the one that is carried out the most frequently. The victim is tricked into opening the attack by receiving fraudulent emails that are designed to look identical to those sent by legitimate bank accounts. In addition to this, users also receive cash winning emails, which can be activated by clicking particular links inside the emails themselves. When a person clicks on the link that is provided to them, they are taken to the homepage of the bogus website that is being promoted on social networking sites [3]. These

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

kinds of baiting approaches are frequently utilized by high-end attackers who force their victims to reveal their card numbers in order to obtain user details. This procedure of sending forged e-mails is carried out all throughout the internet with a thousand different targets, with the expectation that a small percentage of those people will fall for the scam. These faked emails are purposefully crafted to look professional in order to trick recipients into believing they came from a legitimate source. Therefore, it is believed that the primary goal of this assault is to focus on persons with little to no knowledge of online transfers; as a result, these individuals may believe that the response is coming from a legitimate organization; hence, the attacker is able to take advantage of the weakness of the user [4].

For instance, phishing attempts were not seen for the first time until 1995. During these assaults, users were coerced into disclosing their AOL account information to the general public. Since that time, this attack has consistently been on the rise, and significant advances have been made to create a trap for users that is as fool proof as possible [5]. In recent years, it has come to everyone's attention that phishers have been concentrating on webmail attacks rather than other types of social engineering attempts [6]. In 2018, assaults on webmail accounted for a total of 33 percent of all attacks, followed by attacks on security organizations which accounted for 29 percent of all attacks. It has been discovered that "gift cards" have also been utilized to encash money from users in relation to some financial components of phishing. This was done in order to steal their money. It has been demonstrated that an attack of this kind takes place at the earliest possible stage. However, according to estimates provided by the FBI, there are around 26,371 persons who have fallen victim to such frauds [7]. The estimated value of the net loss that was incurred by sectors such as the health care industry, the business, financial services, and educational institutions was approximately one hundred million dollars [8]. In order to get working workers' login credentials, the con involved sending web-based emails to the victims. After gaining access to this information, the phishers exploited it to obtain access to the company's payroll systems, where they then executed programs to prevent employees from receiving notifications when there are changes made to their accounts. Later on, the phishers will alter the card details of the employee and transfer the funds from the employee's account into their own account [9].

Phishing scams were not only seen in the health care industry but also on online gaming websites, indicating that they are not restricted to only that sector. By luring gamers in with the promise of a "free skin gateway" that could only be accessed by clicking on a link that sent them to a phishing website, these cons were carried out with the intention of stealing the login credentials of players. The gamers were placed in an unidentified vulnerable circumstance, which required them to provide their information by scrolling through a false chat box [10]. In order to give the gamers, the idea that the process is legitimate, it was carried out in a manner that was an exact reproduction of the original website. Gamers were prompted to log in through an application known as Steam, which turned out to be a fraudulently produced site in fact. The users' entered credentials were saved, and a mechanism of two-factor authentication was utilized so that validity could be guaranteed for those credentials.

This attack also had the objective of producing one of the most devastating circumstances in Ukraine in 2015; specifically, spear phishing attacks were directed at the administrative department that managed electricity distribution for the country. This assault utilized a Word document that was infected with a virus and was activated when the document was downloaded. This simple action taken by software developers led to the installation of malicious files on the server system, which resulted in a shutdown of the entire organization for a period of 23 hours and a loss of electricity across the country. This campaign, on the other hand, provided a model of what a well-planned

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

and effective phishing attack can look like. The study that was completed before this one reveal that it is of the utmost importance to understand the scope of attacks like this and devise strategies to defend against them. As a result, this paper provides a concise analysis of already carried out attacks and discusses various strategies that might be utilized to stop future attacks.

Literature Review

Because the perpetrator of a phishing attack executes such attacks by altering the personal information of a user, this type of attack is considered to be one of the most serious types of cybercrime. As a result, this worry has developed into one of the most significant security concerns in the business world and in larger organizations that are accountable for providing security services. There are several methods that can still be used to attack a server system; similarly, there are a variety of toolbars that can be installed on internet browsers to warn consumers about websites that contain phishing scams. The following are some of the research studies that have been undertaken by numerous authors. Each of these authors has contributed their own study to the overall effort to detect such assaults and build strategies to prevent them.

The authors of [11] described a number of distinct sorts of phishing schemes, each with their own unique detection system. The authors made the observation that the cyber chain continued to extend because malware actors took advantage of the user's vulnerabilities and targeted to target only those persons who had less understanding regarding the operation of websites asking for login credentials. This allowed the malware actors to continue their spread of the infection. Their investigation included an examination of the methods utilized by phishers to compromise the safety of an organization by focusing on weaknesses in the server infrastructure. In addition, the researchers provided a detailed analysis of the ways in which machine learning algorithms might be utilized to detect and counteract the effects of attacks of this nature.

A comprehensive analysis on the assaults, tactics, and approaches that can be used to identify phishing risks in a server system was presented in [12]. The review helped improve software developers' development of anti-phishing strategies and shifted their focus toward the design of such strategies. The writers went into great length about the ways in which a premeditated attack might bring about the complete dismantling of an organization. Through their research, they attempted to educate developers and raise awareness about the ways in which a virus could contaminate a system without the user's knowledge. In general, the study recommends leaning more toward anti-phishing strategies in order to prevent getting into a situation where you are under assault. In [13], a taxonomy was developed for a much more comprehensive classification of different phishing attack tactics. The writers concentrated solely on web-mail attacks that were responsible for gaining unauthorized access to the personal information of users. In the course of their investigation, they looked into the weaknesses of the current system as well as the ways in which these weaknesses could be improved to better protect users from phishing scams. In addition to providing a comprehensive analysis of a number of distinct types of assaults, the study breaks down those assaults into six distinct stages. It was demonstrated that the comparative taxonomy that the authors developed is superior to the classification methods that are already in use.

The authors of [14] proposed a method for developing those kinds of systems, which are prone to having inadequate security frameworks. The authors made an effort to center their attention on phishing attacks that take place on android phones and involve the use of vishing assaults by the perpetrator, who is attempting to get access to user credentials. This work also contributed to the development of appropriate application programming interfaces (APIs), which helped prevent

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

middle-man attacks. The authors proposed a practical technique to demonstrate how particular features, such as password managers on instant apps, contribute to mobile phishing attempts.

Different phishing attack variations were discussed in [15]. Their research focused on the methods that threat actors could use to obtain personal information from PAN cards and Aadhar cards by convincing an individual to enter their information after being tricked into doing so. In addition, in order to detect the presence of unknown malwares in a software system, the authors created a model that was based on supervised and unsupervised learning approaches. The writers also emphasized how grave the problem of malware attacks is, with the intention of preventing individuals from falling victim to such assaults and drawing attention to the issue overall. In [16], the authors provided their perspectives on the threats posed by the increasingly common practice of conducting business online. The research also demonstrates how obliviously innocent people can get themselves caught up in this web. The authors' primary focus was on the role that machine learning algorithms played in the process of determining whether or not a system was under assault by malware. They also discussed strategies and methods that can be used to prevent assaults of this nature in the future.

[17] presented a taxonomical methodology that cantered on countermeasures and vector outputs of a phishing system. This methodology was proposed. The article went on to provide specifics of the ways in which an organization's vulnerable systems were susceptible to such assaults. The major objective of this body of research was to outline a course of action for software developers to follow in the development of anti-phishing technologies that could thwart attacks of this nature. The study included machine learning algorithms to evaluate and detect malware that was present in the server system, as well as descriptions of attacks that were taking place on social networking websites.

Detection Techniques

All of the currently available traditional methods for detecting phishing attacks are only able to detect twenty percent of the total number of attacks that are being made on the system server. In addition, these detection methods performed with a poor level of accuracy. As a result, there was an immediate demand for the development of strategies and methods to overcome the obstacles and constraints that were already there. Nevertheless, during the past few years, the authors have developed machine learning algorithms that have provided outcomes that are optimized and have improved accuracy.

This way of identifying assaults using ML tactics is often embraced by software developers, despite the fact that the process is time demanding and typically works on a smaller dataset. Additionally, the datasets that it works on are typically much smaller. In addition to machine learning, certain examples of techniques that make use of deep learning are also presented below. The construction of CNN architectures is typically required for deep learning techniques. This is due to the fact that CNNs facilitate the straightforward execution of detection and classification strategies by utilizing several layers of the architecture in question. In addition to the application of machine learning and deep learning approaches, heuristic-based detection tactics are also put into action. All of the research projects that are now being carried out as part of the studies that are being conducted are centred on adjusting to different techniques that have the potential to assist in improving the precision of the entire system. On the other hand, a relatively small number of researchers have been successful in reaching this goal. The capacity of machine learning concepts to effectively detect and classify labels is one of the primary benefits of using these ideas. This ability remains until an optimized model is reached. The role of machine learning algorithms in the detection of phishing is to target websites that are responsible for injecting such attacks. In order to carry out this method,

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

the dataset that has been gathered must first undergo training in such a way that it is able to recognize all of the characteristics that are associated with the phishing website.
In addition, machine learning is able to adapt to numerous strategies at once, making use of particular classifiers and feature selection algorithms that are effective in analysing attacks. The standard workflow of an ML for the detection and classification of malware is depicted further down in this section.

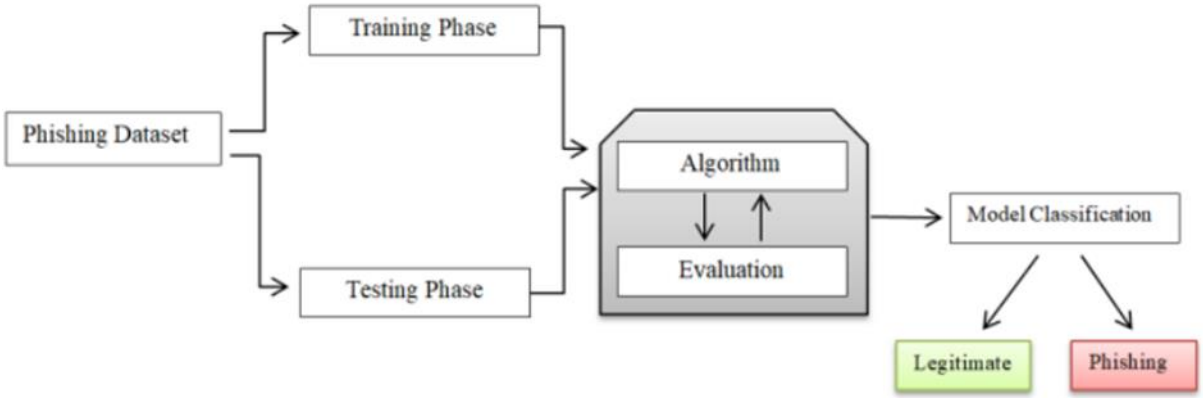

Figure 1: Workflow of machine learning based detection

Phishing detection methods were carried out by the authors of [16], who utilised Random Forest as a classifier in conjunction with a mixture of binary classifiers. The authors downloaded the dataset from the Mendeley site and then used trained RF to run the detection algorithm on it. These techniques for picking features led to the selection of useful features, which were then categorised using the same algorithms that were used for feature selection.
Bagging, boosting, and stacking approaches were combined in the implementation of the detection methodology that was presented in [17]. Because of this model, an ensemble learning model was created, and it was able to pull a total of 30 features from the dataset. However, the dataset was retrieved from the UCI repository, and additional implementation was carried out in order to achieve the level of accuracy that was sought.
The authors of [18] applied detection methods to a URL dataset gathered from the phishing website. This dataset included 34,600 genuine URLS in addition to 38,149 fabricated and false URLS. The dataset underwent additional classification using the NLP methodology. Additionally, the authors in [19] contributed their effort towards URL analysis and detection. The primary objective of the study was to carry out a comparative analysis making use of four different classifiers. Despite this fact, the researchers made use of the dataset that was retrieved from the UCI repository. In the research described in [28], machine learning techniques were utilised to identify and differentiate between legal files and phishing assault files. ANN, SVM, and Random Forest classifiers were utilised throughout the execution of the authors' study.

Anti-Phishing Methodologes

As was discussed in the section that came before this one, a variety of methods and algorithms have been put into place in order to identify the presence of phishing assaults. All of the study academics have had the same basic goal in mind, which is to improve the general effectiveness of the system model in order to thwart any attacks that might be launched against it. Nevertheless, even though there are a number of different anti-phishing solutions available, there are still instances of breaches and spoofing that take occur. The prevalence of such attacks has led to a significant reduction in the amount of monetary assets that have been lost. Not only has this led to a loss of trust, but

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

it has also caused organisations to abandon their previous financial strategies. As a result, it is clear that more advanced methods have to be utilised if one is to be successful in overcoming such challenges. The utilisation of both technical and non-technical detection approaches has been discussed in the parts that came before this one. This subsection of the thesis discusses the many classifications of methodologies that fall within the technical category.

The technique of anti-phishing might begin with the categorization of the many stages of the phishing lifecycle, and then move on to the various tools that can be used to correct and prevent phishing attacks. These types of corrective approaches are able to monitor the various stages of a phishing assault's life cycle and, as a result, prevent the attack from taking place on the system. For example, the cookies that are already installed on the page are a type of web crawler that can be utilised in order to keep track of the component that initially prompted the attack. On the other hand, the prevention methods have the ability to stop attacks from happening in the future and hence avoid their occurrence. In addition to corrective and preventative measures, there are a variety of detecting instruments that can determine whether or not something is the same.

The techniques of blacklist and whitelist are collectively referred to as the phishing detection tool that has gained the most widespread acceptance. In addition to the utilisation of such tools, the utilisation of security toolbars can also be utilised to detect the same. On the user's side of the web server, security toolbars can be installed if they so choose. After the toolbars have been installed, the user has the option of downloading virus scanners to further safeguard his machine against a variety of attacks of this kind. The following is a list of the most popular preventative measures used against phishing attacks:

In the field of intrusion detection systems, concepts from deep learning are also utilised in order to achieve optimal results. The fundamental concept behind deep learning is to operate in a manner analogous to that of the human brain. As a result, the execution of its algorithm involves neurons. The workflow of a neural network performed better than machine learning techniques and assisted in achieving higher levels of accuracy, as was also seen by several researchers. This was one of the findings that emerged from the study. However, because the design of a neural network depends on a wide variety of learning parameters and other elements, carrying it out can be a laborious endeavour. The fundamental idea behind CNNs and RNNs is typically incorporated into the design of deep learning algorithms.

Toolbar	Description
NetCraft	This toolbar functions on the basis of information provided to it through blacklists and whitelists
eBay Toolbar	This toolbar is responsible to protect the users from fake online shopping web pages and websites that might direct them to a phished site
GeoTrust	This toolbar depicts the legitimacy of the web page by using web crawlers and the browser history of the phisher
Cloudmark	This toolbar is responsible to capture attacks using the ratings given by the customer. If the ratings are given by a bot; this toolbar detects that and further gives it a red signal; identifying that it is a phished site
DOMAntiphish	This toolbar is responsible to capture attacks taking place over Document Object Models. It performs the verification check and finally declares the web page as either legit or phished

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

Conclusions

The survey that was carried out as part of this research primarily concentrates on the various ways that are already in use to identify phishing assaults and also discusses the difficulties that are encountered by software developers when attempting to create such strategies. In addition to this, the authors have seen a rise in the frequency of such attacks carried out by malware operators. As a result, the fundamental goal of our research activity is to bring about a better understanding of the functioning method of such attacks and to promote awareness among people who use the internet. The review contributes to the development of a more comprehensive knowledge of the fundamental architecture of phishing assaults. In addition to this, the authors compared a variety of other methodologies already in use by several researchers. It was also brought to everyone's attention that the most prevalent and commonly used methods by hackers to get the credential information of internet users were spear phishing and mail phishing attacks. As a result, the authors came to the conclusion that it was necessary to make recommendations about detecting strategies in order to avoid the general situation of such attacks. In the end, the authors were successful in developing ML and DL detection strategies that were able to further assist in minimizing the number of attacks that were taking place on a legitimate website.

References

[1] Jakobsson, M.; Myers, S. Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft; Wiley: Hoboken, NJ, USA, 2006

[2] ICC (IC3)/Federal Bureau of Investigation (FBI). Internet Crime Report 2018. 2018

[3] Seals, T. Elder Scrolls Online Targeted by Cybercrooks Hunting InGame Loot. Threatpost 2019

[4] Zetter, K. Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid. WIRED 2018

[5] S. Aonzo, A. Merlo, G. Tavella, and Y. Fratantonio, "Phishing attacks on modern android," Proc. ACM Conf. Comput. Commun. Secur., pp. 1788– 1801, 2018, doi: 10.1145/3243734.3243778

[6] G. J. W. Kathrine, P. M. Praise, A. A. Rose, and E. C. Kalaivani, "Variants of phishing attacks and their detection techniques," Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019, no. Icoei, pp. 255–259, 2019

[7] M. V. Kunju, E. Dainel, H. C. Anthony, and S. Bhelwa, "Evaluation of phishing techniques based on machine learning," 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019, no. Iciccs, pp. 963–968, 2019, doi: 10.1109/ICCS45141.2019.9065639

[8] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," Comput. Secur., vol. 68, pp. 160–196, 2017, doi: 10.1016/j.cose.2017.04.006

[9] K. A. Moul, "Avoid phishing traps," Proc. ACM SIGUCCS User Serv. Conf., no. August 2017, pp. 199–208, 2019, doi: 10.1145/3347709.3347774

[10] Q. Cui, G. V. Jourdan, G. V. Bochmann, R. Couturier, and I. V. Onut, "Tracking phishing attacks over time," 26th Int. World Wide Web Conf. WWW 2017, pp. 667–676, 2017

[11] M. N. Banu and S. M. Banu, "A Comprehensive Study of Phishing Attacks," Int. J. Comput. Sci. Inf. Technol., vol. 4, no. 6, pp. 783– 786, 2013

[12] T. Churi, P. Sawardekar, A. Pardeshi, and P. Vartak, "A secured methodology for anti-phishing," Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017, vol. 2018- Janua, pp. 1–4, 2018

[13] James, J., Sandhya, L., & Thomas, C. (2013). Detection of phishing urls using machine learning techniques. In 2013 International conference on control communication and computing (ICCC) (pp. 304–309)

[14] Subasi, A., Molah, E., Almkallawi, F., & Chaudhery, T. J. (2017). Intelligent

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.812

phishing website detection using random forest classifier. In 2017 International conference on electrical and computing technologies and applications (ICECTA) (pp. 1–5). IEEE

[15] Hutchinson, S., Zhang, Z., & Liu, Q. (2018). Detecting phishing websites with random forest. In International conference on machine learning and intelligent communications (pp. 470–479). Springer

[16] Mao, J., Bian, J., Tian, W., Zhu, S., Wei, T., Li, A., et al. (2018). Detecting phishing websites via aggregation analysis of page layouts. Procedia Computer Science, 129, 224–230

[17] Jain, A. K., & Gupta, B. B. (2018). Towards detection of phishing websites on client-side using machine learning based approach. Telecommunication Systems, 68(4), 687–700 Li, Y

[18] ., Yang, Z., Chen, X., Yuan, H., & Liu, W. (2019). A stacking model using url and html features for phishing webpage detection. Future Generation Computer Systems, 94, 27–39

[19] Document validation and verification system S Shivadekar, SR Abraham, S Khalid Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET) 5 (3)

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731