WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comparative Analysis of Phishing Attacks and their Remedies

[1]Chanchal Shekhawat, [2] Dr. Subhash Chandra

[1]Research Scholar, [2]Associate Professor

[1]chanchalshekhawat15@gmail.com, [2] subhashccjat@yahoo.com

[1,2]Department of Computer Science & Engineering, Rajasthan College of Engineering for women, Jaipur, Rajasthan 302026

ABSTRACT

Phishing is a security attack that uses websites or emails to get a user's personal information such as passwords, credit card numbers, or other account information. Phishing websites resemble authentic websites, making it difficult for the average person to tell the difference. Phishing is a type of internet fraud that involves the theft of a user's personal information and credentials. It's a type of fraud in which the perpetrator has complete access to the private information of others. An competent designer may easily construct a false website that looks just like the real, making detecting the website as fake difficult. To avoid being identified, almost all phishy URLs employ HTTPS and redirects. This study gives a comprehensive review of the strategies for detecting phishing websites. A comparison of the anti-phishing tools in use was completed, and their limitations were acknowledged. Our key addition is an analysis of previous URL-based features to improve their definitions in light of the current scenario.

Index Terms—Phishing attacks, web attacks, web domains, https Protocols.

I. INTRODUCTION

PHISHING- Phishing is a scam in which someone impersonating a genuine and well-known company contacts a targeted individual in order to attract them into revealing personal information such as financial information, credit card numbers, and passwords. Individuals' personal information is then utilized to get access to their accounts, resulting in identity theft and financial loss. In 2004, a California teenager was charged with phishing after making a spoof of the website "America Online." Sensitive information was obtained from duped consumers through this replicated website, and credit card details were accessed to withdraw money from their accounts.

Communications purporting to be from popular social web sites, auction sites, banks, online payment processors or IT administrators are commonly used to lure the unsuspecting public. Apart from the internet phishing, phishing artist also make phone calls and send text messages with lucrative offers and entice the targeted individual to provide sensitive information.

Phishing can be done in a variety of ways:

a) Email-to-email: In this method, an email is sent to a specific user, urging the user to supply sensitive information, which is then forwarded to the phishing artist.

b) email-to-website: In this case, an email is sent to web users with a link that, when clicked, redirects the user to a fake website.

c) website-to-website phishing: This sort of phishing occurs when a user visits a phishing website after clicking on an advertisement link on a valid website.

d) browser-to-website: This happens when a web user is negligent and misspells a valid website, which then leads to a phishing website.

There are two ways to protect users against phishing attacks on these websites. The first is to ban phished websites. The URL of each website is verified in phishing detection engines such as phish tank and others using this way. However, because there are many freshly generated web sites that are not in this pool, this will not provide total protection against phishing websites.

Heuristic-based approach is the second approach. This method tries to detect phished pages based on a number of characteristics such as URL authenticity and so on. The content of online sites can also be utilized to detect phishing. For the same objective, there are also several approaches based on machine learning. For the detection of phished pages, some artificial intelligence methods have been introduced. The usage of security toolbars is another such technique. Our study will concentrate on detecting phished pages using online browser security toolbars, which are an integral element of every advanced browser.

Literature Survey

Phishing[1] is a violation of the Confidentiality, Integrity, and Availability (CIA) rule. Many methods are being developed to detect phishing assaults, yet they are still feasible and pose a hazard to people. Hardware devices can detect phishing websites with great accuracy, but they are highly

expensive, hence software-based alternatives are favored. Blacklist and whitelist[2]are used to identify phishing with great accuracy, but they require list maintenance because the list of phishing website URLs must be manually updated. As a result, automatic detection technologies such as machine learning and heuristic approaches are presently applied to solve the difficulties in manual list update[3]. For identifying the phishing URL, many machine learning approaches use web structure or web content-based methods; Cantina+[4] is a well-known heuristic-based solution. Sunil et al. developed a phishing detection strategy simply based on Google's PageRank value; however, because many new legal websites or blogs with low rank can be misunderstood as phishing websites, a combination of several factors can increase the performance and identification rate. For phishing detection, a web structure approach[5] based on page ranking was developed. The use of a browser plug-in to minimize phishing assaults [6] is possible, but the user alert is created solely on the basis of blacklists, which does not address zero-day phishing attacks. The phishing detection approach used by Netcraft[2] is based on both blacklists and heuristics. It swiftly detects phishing websites, but it is ineffective when the perpetrators plan the attack to prevent detection. Dhamija and colleagues proposed Security skin is a dynamic method of generating a visual hash for browser window customisation to show that the site is secure[7]. Keywords, IP addresses, grammar checkers, URLs, and other features are combined in rule-based filters[8]. These characteristics are used to generate rules for detecting phished emails. For improved identification, the rules are changed on a regular basis. A web crawler, often known as a web spider or an online bot, is a script that is used to automatically scan webpages. The use of web crawlers in phishing research is very new, as crawlers are often employed for data extraction. It also addresses zero-day phishing attacks that aren't covered by many other existing solutions. Web crawlers[4] are used by the majority of search engines to harvest pages from the internet for indexing in search engines. It collects all of the useful web pages and links that connect them. Zero-day phishing websites are brand new assaults launched by attackers within the last day or hour. It is extremely risky, as there will be no case reports of such attacks. Existing methods are effective in identifying phished websites, but they are ineffective in detecting zero-day phishing attacks. A web crawler-based phishing detection system has been offered as a solution to this problem.

LPD, a client-side based web page phishing detection system, was proposed by Varshney et al [9]. The strings from a web page's URL and page title are retrieved and searched on the Google search engine. The web page is regarded to be authentic if the domain names of the top T search results and the domain name of the given URL match. Their evaluations resulted in a true positive percentage of 99.5 percent.

Smadi et al. [10] suggested a reinforcement learning-based neural network model that can adapt to the dynamic nature of phishing emails. Using an updated offline database, the suggested model can manage zeroday phishing assaults and offset the problem of a limited dataset. On fifty features taken from a dataset of 12,266 emails, their experiment yielded a high accuracy of 98.63 percent.

ML-based approaches to detect phishing websites are a hot topic in study, with a variety of supervised classification techniques being used to classify phishing websites. For phishing detection, Feng et al. propose a new neural network [11].

They boost the network's generalization ability by using a risk minimization principle. The proposed network's performance is measured against a UCI repository 1 that contains 11,055 phishing/legitimate samples. The dataset also includes 30 features for each website, which are divided into four categories: address bar-based, abnormal-based, HTML/Javascript-based, and domain-based. Rao and Pais offer a new approach that combines machine learning and image verification to detect phishing websites. They also harvest information from URLs, websites, and third-party services [12]. It's worth noting that, while employing third-party service features can lengthen detection times, it improves detection accuracy in practice [13]. They test the suggested method on 1407 legitimate and 2119 phishing websites from the PhishTank 2 and Alexa databases 3 databases, respectively.

For detecting phishing websites, Mohammad et al. present a unique self-structuring neural network [14]. They provide 17 features for 600 genuine and 800 phishing websites from the PhishTank and Millersmiles 4 archives, some of which were scraped from third-party services. Their findings show that the neural network has a high degree of generalizability and ability in phishing detection. They propose a feed forward neural network trained by back propagation to classify websites in another paper [15]. For 859 legitimate and 969 phishing websites, respectively, 18 features are provided. To detect phishing websites, Jain and Gupta present a machine learning-based technique that uses solely client-side information [16]. They extract 19 features from the online pages' URL and source code and test their method on 2,141 phishing web pages from

1. https://archive.ics.uci.edu/ml/datasets/phishing+websites
2. http://www.phishtank.com/
3. http://www.alexa.com/
4. http://www.millersmiles.co.uk/

PhishTank and Openfish 5 were used, as well as 1,918 valid pages from the Alexa database, as well as various online payment and banking services.

Although all of the research listed above suggest distinct qualities for detecting phishing websites, some of these features may not be discerning enough to detect phishing cases [17]. Only a little amount of research has gone into determining the most effective elements for detecting phishing websites. To detect phishing activities, Rajab suggests utilizing Correlation Feature Set (CFS) and Information Gain (IG) to choose the most influential features [17]. The findings of the three UCI repositories with 30 features provided for 11,055 samples show that IG and CFS picked 11 and 9 characteristics, respectively. A data mining method called RIPPER is used to evaluate the classification performance using selected characteristics. Babagoli et al., using a comparable data set, suggest feature selection using decision trees and the wrapper technique [18], resulting in the selection of 20 features [19]. They use an unique meta-heuristic-based nonlinear regression approach to assess phishing detection performance. These studies' feature selection approaches, however, are data-dependent and require user-specified threshold values that should be set heuristically. These thresholds can have an impact on the classification algorithm's overall performance, especially when features are chosen based on out-of-sample training data in practice.

Conclusion

Although many attempts have been made to detect and contain phishing assaults using various anti-phishing algorithms, the attacks still require adequate attention. For freshly registered domains, the blacklisting strategy for detecting spam proved to be less effective. Every day, a new set of

bogus websites is established. These websites cannot be discovered if the bogus URL is not included in the blacklist. When there are no rules for a particular attribute, the heuristics-based approach can fail. As a result, the attribute goes undiscovered, and we must ensure that all of the rules or heuristics are included in the system.

References

[1] S. Marchal, J. François, R. State and T. Engel, "PhishStorm: Detecting Phishing With Streaming Analytics," in IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458-471, Dec. 2014.

[2] Netcraft, Netcraft Anti-Phishing Toolbar. Visited: Nov 20, 2006. http://tool bar.netcraft.com/

[3] Checking Page Rank. Accessed: Sep. 2016. [Online]. Available: https://www.prchecker.info/check_page_rank.php

[4] A. Naga Venkata Sunil and A. Sardana, "A PageRank based detection technique for phishing web sites," 2012 IEEE Symposium on Computers & Informatics (ISCI), Penang, 2012, pp. 58-63. doi: 10.1109/ISCI.2012.6222667

[5] A. N. V. Sunil and A. Sardana, "A PageRank based detection technique for phishing Web sites," in Proc. IEEE Symp. Comput. Informat. (ISCI), Penang, Malaysia, 2012, pp. 58–63.

[6] Blake Ross, Collin Jackson, Nick Miyake, Dan Boneh and John C. Mitchell. A Browser Plug- In Solution to the Unique Password Problem. http://crypto.stanford.edu/PwdHash/, 2005.

[7] R. Dhamija and J.D. Tygar, "The Battle against Phishing: Dynamic SecuritySkins", Proc. Symp. Usable Privacy and Security, 2005,pp 77-88. Mobile Marketing Statistics. Accessed: Mar. 2017.

[8] C. N. Gutierrez et al., "Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks," in IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 6, pp. 988-1001, 1 Nov.-Dec. 2018. doi: 10.1109/TDSC.2018.2864993

[9] G. Varshney, M. Misra, and P. K. Atrey, "A phish detector using lightweight search features," Computers & Security, vol. 62, pp. 213– 228, 2016.

[10] S. Smadi, N. Aslam, and L. Zhang, "Detection of online phishing email using dynamic evolving neural network based on reinforcement learning," Decision Support Systems, vol. 107, pp. 88–102, 2018.

[11] F. Feng, Q. Zhou, Z. Shen, X. Yang, L. Han, and J. Wang, "The application of a novel neural network in the detection of phishing websites," Journal of Ambient Intelligence and Humanized Computing, 2018.

[12] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," Neural Computing and Applications, pp. 1–23, 2018.

[13] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," Expert Systems with Applications, vol. 117, pp. 345–357, 2019.

[14] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Neural Computing and Applications, vol. 25, no. 2, pp. 443–458, 2014.

[15] R. Mohammad, T. McCluskey, and F. A. Thabtah, "Predicting phishing websites using neural network trained with back-propagation." World Congress in Computer Science, Computer Engineering, and Applied Computing, 2013.

[16] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, no. 4, pp. 687–700, 2018.

[17] M. Rajab, "An anti-phishing method based on feature analysis," in Proceedings of the 2nd International Conference on Machine Learning and Soft Computing. ACM, 2018, pp. 133–139.

[18] D. Rodrigues, L. A. Pereira, R. Y. Nakamura, K. A. Costa, X.-S. Yang, A. N. Souza, and J. P. Papa, "A wrapper approach for feature selection based on bat algorithm and optimum-path forest," Expert Systems with Applications, vol. 41, no. 5, pp. 2250–2258, 2014.

[19] M. Babagoli, M. P. Aghababa, and V. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," Soft Computing, pp. 1–13, 2018.