Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/160609

Anti-Phishing Tools: State of the Art and Detection Efficiencies

M. F. Alghenaim[*]*, N. A. Abu Bakar and F. Abdul Rahim*

Advanced Informatics Department, Razak Faculty of Technology and Informatics Universiti Teknologi Malaysia 54100 Kuala Lumpur, Malaysia

Received: 22 Jun. 2022, Revised: 22 Sep. 2022, Accepted: 24 Sep. 2022.
Published online: 1 Nov. 2022.

Abstract: A phishing attack is a process of obtaining a customer's private data, whether by using phishing emails or fake websites. With every new development on the Internet, the attackers' means of phishing attacks develop, requiring more powerful phishing tools to counter these attacks. The Internet has become an essential part of the personal and social life of the public and governments, institutions, and companies all over the world. This means that Internet users need tools to protect against phishing attack types and web risks, which may cause personal, institutional, financial, and informational damage. Hence, this study reviews anti-phishing attack tools and shows their accuracy in addressing the current challenges of phishing attacks.

Keywords: Phishing attacks, Advanced phishing tools, Cyberattack, Internet security, Machine learning, Anti-phishing.

1 Introduction

Anti-Phishing is a term that describes measures to prevent phishing assaults. Anti-Phishing is a browser extension (programs) that seeks to protect inexperienced users from phishing attacks based on faked websites. Furthermore, it keeps track of a user's sensitive information and issues warnings if sensitive information is entered into a form on an untrustworthy website [1]. Websites' phishing attacks keep motiving many losses and damages to individual customers and corporations [2]. Further to monetary losses, phishing attacks pose additional safety threats to corporations, including malware and viruses [3]. Many cyberattacks are phishing as vectors to obtain credentials or coerce users into self-executing malware-infected files. Phishing is a hazardous attack because it reduces the number of channels attackers wants to get sensitive data.

Anti-phishing equipment is a safety technology designed to shield customers from phishing assaults based totally on faux or spoofed websites. Unluckily, the terrible effects of those equipment have affected their adoption and perceived usefulness; users do now not agree with their hints, although they are correct [4]. For this reason, there is current attention on improving the detection capabilities of anti-phishing equipment [5]. However, decreasing the impact of phishing in both private and organizational environments relies specifically on the security behavior of Internet users [6]. Hence, this paper aims to review the current anti-phishing tools and discuss their capabilities in addressing the challenges of phishing attacks.

The remainder of this paper is organized as follows: Section 2 discusses the literature review of the related studies. Section 3 presents and discusses non-traditional methods for detecting phishing websites. Finally, the conclusion is presented in Section 4.

2 Literature Review

2.1 Phishing

Phishing is social engineering and one of the top tools available for attackers to complete three (3) critical objectives at once:

(1) Use misleading links (generated with the help of URL shorteners) to have the attack occur and lure the end-user into clicking a specific link or downloading an infected file [7];

(2) Obtain all types of sensitive information from the end-user so that the attacker can exploit it later [7].

(3) Instil fear in the given end-users and force them to respond quickly and, most probably, pay money to grant permission to get their information back (which is never promised, depending on the attacker [8].

Ransomware infections remained stable year over year, according to a Proofpoint phishing report from 2021. At the same time, phishing-related malware infections were reported by 17% fewer people. In addition, 47% fewer people suffered direct financial losses. These findings imply that firms have put more strong countermeasures to these attacks [9]. According to [10], 52% of email clients failed to detect a legitimate phishing email, as shown in Figure 1.

[*]Corresponding author e-mail: aalghenaim@graduate.utm.my

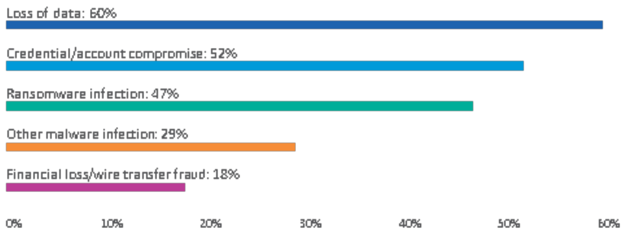

Fig. 1. Impacts of Successful Phishing Attacks

Phishing attacks and other cybersecurity dangers are rising, so this is a concerning trend in the cybersecurity domain. Financial institutions targeted 24.9% of phishing attacks in Quarter1 2021; as shown in Figure 2, 23.6% of threats come from social media. These two industries were the most vulnerable to phishing in the first three months of 2021.

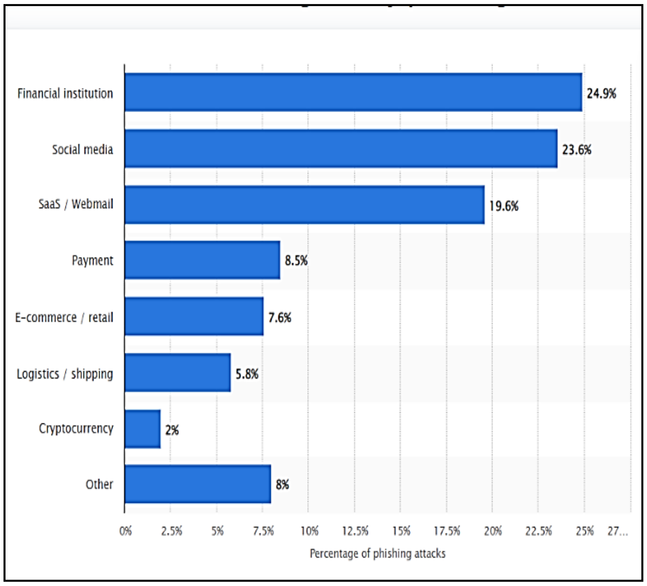

Fig. 2. Percentage of phishing attacks as of 1st Quarter 2021 [11].

2.2 Anti-Phishing Tools

2.2.1 Using Antiphish techniques

AntiPhish is a browser plugin that tracks sensitive information. Every time a consumer tries to enter sensitive information on a website, a warning is issued. This is very effective when someone accidentally enters financial institution credentials on a phishing website. However, AntiPhish has the problem of considering the reuse of valid certificates. To get around this difficulty of use, DOM AntiPhish is proposed. This technique compares the shapes of the analyzed pages' reporting elements to determine if the two pages are identical. If records are reused on a page, it is very similar to the original page (sensitive information)—suspicion of a phishing attempt. When facts are entered on a unique page, the system assumes legitimate reuse of statistics. Although DOM AntiPhish can effectively detect phishing pages, the biggest challenge is that the DOM tree is not necessarily a reliable feature for determining similarity between pages. In some cases, an attacker could also use various DOM elements to create a page-like appearance.

2.2.2 Passpet

Passpet simplifies logging in to websites by entering a username and password with the click of a mouse. We only have to remember one password, and Passpet generates a different password for each website. Even if one website is broken into, other accounts and passwords are safe. Passpet protects the user from attackers who try to trick the user into revealing passwords because each password is generated only for the website where the user initially set it. Passpet appears in the user Firefox toolbar as an animal icon. Everyone gets a random animal named, so the Passpet button is hard for a scammer to imitate. When the user launches Firefox for the first time, the user's Passpet is asleep. To wake it up, the user clicks on it and enters his/her main secret. When Passpet is awake, we can fill in the password automatically by clicking on the icon. If the user enters a caption in the text box, it will be displayed again when the user revisits the same page. To enter a password, Passpet calculates the password from the user label. So if the user wants to use Passpet on a specific website, enter the website's label in the field; when the user register for a new account on the website, the user clicks his/her Passpet to enter the new password [12].

2.2.3 SpoofGuard

SpoofGuard is an anti-phishing toolbar that uses a variety of heuristics to classify phishing web pages. It calculates a score for every web page in the form of a weighted total of the results of each heuristic collection and is characterized by higher false-positive rates. Since SpoofGuard does not use the blacklist approach, the URL cannot hash to another value or appear to come from a different domain name. This approach makes it receptive to content distribution network attacks. SpoofGuard uses the content of a website to test its authenticity. It waits for the web page to finish loading before classifying it as legitimate or otherwise, making it susceptible to page loading attacks [2].

2.2.4 PhishProof

PhishProof is an anti-phishing browser extension tool that uses blacklist and heuristic methods to help browser users distinguish between legitimate and phishing websites. The system notifies the user upon the identification of a phishing website. Also, the system profiles all known websites into scores based on their phishing characteristics, and this score is presented to the user when visiting the website. It uses a symmetric key algorithm as its security algorithm, enabling it to detect both known and unknown phishing sites more than other approaches. However, the shortcoming with this tool is that, since it performs multiple checks while authenticating, it may result in a slow response. Though this tool can protect users, it cannot protect them from malware [13].

3 Non-traditional methods for detecting phishing websites

There are standard methods for dealing with internet phishing, such as legal, educational, and awareness campaigns, blacklist methods, visual similarities methods, and search engine methods. On the other hand, non-traditional methods are more efficient and improve the accurate detection of phishing, as shown in many empirical studies. Non-traditional methods include content-based, heuristic, and artificial intelligence methods.

Table 1 summarizes non-traditional methods for phishing website detection, including the methodology employed, the datasets used, and the accuracy of the results. The researchers used various and extensive datasets to analyze the methods; however, the feature significance may be one-sided or incorrect. The URLs and site pages we focused on and emails because email is a popular means of delivering phishing URLs or malware.

Table 1: Non-traditional methods for phishing websites detection

Anti-Phishing Method	Ref	Accuracy
Meta-heuristics (Harmony Search (HS), and Super Vector Machine (SVM))	[14]	92.80%
TWSVM	[15]	98.05%
Modified TF-IDF	[16]	89%
Machine learning (XCS)	[17]	98.39%
Random Forest (RF)	[18]	97.98%
Convolutional Neural Networks (CNN)	[19]	99.98%
Artificial Neural Network (ANN)	[20]	98.77%
Feature Validity Value (FVV) and Neural Network	[21]	98.49%
Long short-term memory (LSTM) and CNN	[22]	97%
Adaptive Neuro Fuzzy Inference System (ANFIS)	[23]	98.3%
RF and FURIA	[24]	99.98%
neuro-fuzzy	[25]	98.36%
Deep Packet Inspection (DPI), Software Defined Networking (SDN) and ANN	[26]	98.39%
Particle Swarm Optimization (PSO) and back propagation (BP) neural network	[27]	98.95%
J48 algorithm and C4.5 algorithm	[28]	98.87%

In the Heuristic approach, Babagoli M. et al. [14] proposed a method for an anti-phishing website that used a nonlinear relapse algorithm and a dataset of 11055 legitimate and phishing site pages. To anticipate and discriminate against bogus locations, two meta-heuristic algorithms are used efficiently. The locations were ordered using the nonlinear relapse technique, with the proposed relapse model's bounds determined using the HS algorithm. Rao R. S. et al. [15] developed a heuristic technique by utilizing the TWSVM (twin support vector machine) classifier to identify dangerous enrolled phishing websites and websites facilitated on arrangement servers to overcome the above-noted constraints. This model looks at the sign-in page and the main page of the visiting site to distinguish phishing sites hosted in various domains. They used a variety of SVMs to create phishing web pages, and they discovered that the TWSVM outperforms various adaptations.

Following the content-based approach, A. K. Jain et al. [16] proposed a method for detecting phishing cyberattacks that revealed a client's qualifications to hostile substances, resulting in a security breach. The method aims to use a web search tool to match the website's space name under a microscope with the destinations that come up due to our search query. They initially noticed how the standard TF-IDF "Term Frequency–Inverse Document Frequency" works in classifying websites as phishing or legitimate. Then, by allocating various weights to various label information and controlling the TF-IDF result, implement a weighted heuristic proposed fully in their study to work on the presentation of their phishing indication. By using the machine learning methods, Yadollahi M. M. [17] proposed a flexible detection framework that can detect phishing sites from URLs by employing several web browser's unique properties alone, without third-party services. Their system comprises two main components named Feature Extractor and XCS, which operate as phishing detectors by detecting the style of upcoming phishing sites and forwarding a standard set of rules.

The deep learning methods have improved phishing detection efficiency significantly. Wei W. et al. proposed a deep neural network with convolutional layers was proposed by Wei W. et al. [19] for identifying phishing sites based solely on the URL address content. As a result, the approach is faster and can detect zero-day attacks. The network they demonstrated had been appropriately upgraded, so it can now be used even on cell phones without affecting its performance. Gajera K. et al. [20] developed a system to detect pharming in which they query a local and a global DNS to obtain IP addresses and compare them. If the results are the same, it suggests there will be no pharming, and they will proceed with the internet page examination. Zhu et al. [21] proposed an effective neural network-based phishing site detection methodology in another avenue. The proposed methodology was initially introduced to another metric, feature validity value (FVV), which evaluated the impact of delicate characteristics on phishing site detection. Following that, an estimation is planned to choose the best characteristics of phishing websites using the new FVV index. Wang W. et al. [22] suggested an approach for quickly detecting phishing sites that uses a bidirectional LSTM network to remove global components of the belt tensor and assign all string data to each letter in the URL. They use a CNN to determine which characters are thought to be important in phishing detection, capture the URL's essential segments, and pack the removed components into a fixed-length vector space.

For the Fuzzy Rule-based approach, Adebowale M. A. et al. [23] created an adaptable Neuro-Fuzzy Inference System (NFIS) for web phishing detection and insurance based on a solid design incorporating the integrated components of text, photos, and frames. Pham C. et al. [25] used URL attributes and online traffic data to detect phishing sites in conjunction with a proposed neuro-fuzzy structure called Fi-NFN. The trial results of their proposed technique, based on a large dataset of actual phishing incidents, have proven that their framework can effectively prevent phishing attacks and improve network security.

Chin T. et al. [26] proposed PhishLimiter, phishing detection, and mitigation approach, in which they first proposed a novel Deep Packet Inspection (DPI) strategy. Then, integrated it with Software Defined Networking (SDN) to spot phishing exercises via email and web-based communication. They developed an Artificial Neural Network (ANN) model to aggregate phishing attack markings and arrange real-time DPI for PhishLimiter to effectively recognize the aspects of phishing assault in the actual world. Another study by Chen W. et al. [27] suggested an approach for constructing a phishing site identification framework by combining Particle Swarm Optimization (PSO) and Back Propagation (BP) neural networks. They employed PSO advances neural network boundaries to further enhance the assembly execution of the neural network detection framework. Compared to the standard BP neural network approach, preliminary results show that this computation can improve the prediction speed and precision of identifying phishing sites by 3.7 percent.

Lastly, in the data mining domain, Smadi S. et al. [28] introduced an intelligent model for phishing message identification that depends on a preprocessing stage that removes some characteristics from various email portions. The J48 classification technique is used to arrange the separated characteristics. They tried out a total of 23 different features. For training, testing, and approval, a ten-fold cross-approval was used. The main goal of their work is to increase the overall benefits of email characterization by focusing on the preprocessing step and determining the optimal technique that can be used in this sector. The results demonstrate the benefits of removing characteristics from the dataset using their preprocessing step. Their model resulted in more accuracy.

4 Conclusion

Phishing is one of the most recent internet threats, and it has resulted in massive losses for online users, electronic businesses, and financial institutions. With the ongoing threat of phishing threats, the findings of this study show that the more accurate anti-phishing tools are, the more they help improve users' capacity to recognize phishing sources and respond quickly to electronic attacks. Phishing is a common strategy of deceiving internet customers and stealing their financial information by imitating their websites. This article provides a thorough evaluation of the most recent developments in anti-phishing. The study will aid academics and companies in assessing the present state of phishing detection and guiding them toward developing a unique strategy to encourage the most effective online phishing detection tactics.

Conflict of interest

The authors declare that there is no conflict regarding the publication of this paper.

References

[1] S. Back and R. T. Guerette, Cyber Place Management and Crime Prevention: The Effectiveness of Cybersecurity Awareness Training Against Phishing Attacks, *J. Contemp. Crim. Justice*, **37(3)**, pp. 427-451, (2021).

[2] Y. Zhang, S. Egelman, L. F. Cranor, and J. Hong, Phinding Phish: Evaluating Anti-Phishing Tools, *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS 2007)*, pp. 1-16, (2007).

[3] T. Dinev and Q. Hu, The centrality of awareness in the formation of user behavioral intention toward protective information technologies, *J. Assoc. Inf. Syst.*, **8(7)**, (2007).

[4] A. Abbasi, Z. Zhang, D. Zimbra, H. Chen, and J. F. Nunamaker, Detecting fake websites: The contribution of statistical learning theory, *MIS Q. Manag. Inf. Syst.*, **34(3)**, 435-461, (2010).

[5] W. Liu, X. Deng, G. Huang, and A. Y. Fu, An antiphishing strategy based on visual similarity assessment, IEEE Internet Comput., 10(2), 58-65, (2006).

[6] J. Park, J. Y. Son, and K. S. Suh, Fear appeal cues to motivate users' security protection behaviors: an empirical test of heuristic cues to enhance risk communication, *Internet Res.*, **32(3)**, (2021).

[7] R. Alabdan, Phishing attacks survey: Types, vectors, and technical approaches, *Future Internet*, **12(10)**, (2020).

[8] R. Bhakta and I. G. Harris, *Semantic analysis of*

dialogs to detect social engineering attacks, in Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC, 424-427, (2015).

[9] Proofpoint, 2020 State of the Phish, *Proofpoint*, pp. 1-48, (2020).

[10] T. Uy and C. Dimaano, Lessons learned from The Spirit Catches You and You Fall Down: Student perspectives on how cultural differences can lead to health disparities, *Health Educ. J.*, **79(1)**, 73–81, (2020).

[11] Statista, Online industries most targeted by phishing attacks, (2021).

[12] K. P. Yee and K. Sitaker, *Passpet: Convenient password management and phishing protection*, in ACM International Conference Proceeding Series, **149**, 32-43, (2006).

[13] T. Zahid, *An Anti-Phishing Tool: PhishProof*, University of Manchester, United Kingdom, (2012).

[14] M. Babagoli, M. P. Aghababa, and V. Solouk, Heuristic nonlinear regression strategy for detecting phishing websites, Soft Comput., **23(12)**, pp. 4315-4327, (2019).

[15] R. S. Rao, A. R. Pais, and P. Anand, A heuristic technique to detect phishing websites using TWSVM classifier, Neural Comput. Appl., **33(11)**, pp. 5733-5752, (2021).

[16] A. K. Jain, S. Parashar, P. Katare, and I. Sharma, "PhishSKaPe: A Content based Approach to Escape Phishing Attacks," *Pro*, **171**, 1102-1109, (2020).

[17] M. M. Yadollahi, F. Shoeleh, E. Serkani, A. Madani, and H. Gharaee, *An Adaptive Machine Learning Based Approach for Phishing Detection Using Hybrid Features*, in 2019 5th International Conference on Web Research (ICWR), Tehran, Iran, Apr, 281-286. (2019).

[18] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, Machine learning based phishing detection from URLs, *Expert Syst*, **117**, pp. 345-357, Mar. (2019).

[19] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, and M. Woźniak, Accurate and fast URL phishing detector: A convolutional neural network ap. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, **178(C)**, 107275, (2020).

[20] K. Gajera, M. Jangid, P. Mehta, and J. Mittal, *A Novel Approach to Detect Phishing Attack Using Artificial Neural Networks Combined with Pharming Detection*, in 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 196-200, (2019).

[21] E. Zhu, Y. Chen, C. Ye, X. Li, and F. Liu, OFS-NN: An Effective Phishing Websites Detection Model Based on Optimal Feature Selection and Neural Network, *IEEE Access*, **7**, 73271-73284, (2019).

[22] W. Wang, F. Zhang, X. Luo, and S. Zhang, PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks, *Secur. Commun. Netw.*, 1-15, (2019).

[23] M. A. Adebowale, K. T. Lwin, E. Sánchez, and M. A. Hossain, Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text, *Expert Syst*, **115**, 300-313, (2019).

[24] A. Abuzuraiq, M. Alkasassbeh, and M. Almseidin, *Intelligent Methods for Accurately Detecting Phishing Websites*, in 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 085-090, (2020)

[25] C. Pham, L. A. T. Nguyen, N. H. Tran, E.-N. Huh, and C. S. Hong, Phishing-Aware: A NeuroFuzzy Approach for Anti-Phishing on Fog Networks, *IEEE Trans. Netw. Serv. Manag.*, **15(3)**, 1076-1089, (2018).

[26] T. Chin, K. Xiong, and C. Hu, Phishlimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking, *IEEE Access*, **6**, 42516-42531, (2018).

[27] W. Chen, X. A. Wang, W. Zhang, and C. Xu, *Phishing Detection Research Based on PSO-BP Neural Network*, in Advances in Internet, Data & Web Technologies, The 6th International Conference on Emerging Internet, Data & Web Technologies, (EIDWT), **17**, L. Barolli, F. Xhafa, N. Javaid, E. Spaho, and V. Kolici, Eds. Cham: Springer International Publishing, 990-998, (2018).

[28] S. Smadi, N. Aslam, L. Zhang, R. Alasem, and M. A. Hossain, *Detection of phishing emails using data mining algorithms*, in 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Kathmandu, Nepal, Dec, pp. 1-8, (2015).

Mohammed Alghenaim: A Ph.D. student in Information Systems - Security Assurance at the University of Technology Malaysia (UTM). His research interests are security assurance, cybersecurity war, security penetration, decision molding, and quantitative method insurance, including designing advanced network security, data analysis, and software testing. He has published research articles in many international journals of

Information systems, Information security, and computer sciences. He is a referee at the 2nd International Conference on Emerging Technologies and Intelligent Systems (ICETIS 2022).

Nur Azaliah Abu Baka: A Ph.D. is a Senior Lecturer in the Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia. She has extensive experience in ICT and has served in the Public Sector as well as several multinational companies.

ICT Consultation Division at MAMPU, INTAN, JPA, Astro Malaysia, and Mydin Mohamed Holdings. She is a member of the OBI RG), MyAIS, AIS, and MBoT.Her topics of expertise and research interests include, but are not limited to, Informatics, Enterprise Architecture, Business Intelligence, ICT Strategic Planning, and Digital Government.

Fiza Abdul Rahim: A Ph.D. is a Senior Lecturer at the Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia. She is a Multimedia Protec Research Group (MPROTEC RG) member and the Malaysia Board of Technologists (MBOT).

She is also a Research Associate of the Institute of Informatics and Computing in Energy (IICE), Universiti Tenaga Nasional (UNITEN). Her topics of expertise and research interests include but are not limited to Information Security, Cybersecurity, Digital Forensics, and Informatics.