A Literature Survey of Phishing and Its Countermeasures

Joshua Vega, Daniel Shevchyk, Yuan Cheng

California State University, Sacramento, U.S.A.

Abstract

Phishing remains one of the most prominent threats to individuals and organizations in the public and private sectors today. It is a gateway attack that can lead to more severe attacks, including identity theft, ransomware attacks, and denial-of-service attacks. Unfortunately, people and their poor choices, ignorance, and lack of attention to detail combine to make phishing prevalent and effective. This paper gives an overview of the phishing problem and introduces the motives behind phishing and common attack vectors used in phishing attacks. We then review several existing phishing detection and mitigation solutions in three categories: education, machine learning, and blacklists/whitelists. Finally, we also discuss the challenges found in the solutions we surveyed.

1. Introduction

Phishing refers to a form of cyber-attack where attackers spoof messages from reputable sources and attempt to defraud recipients' personal information. The term was coined because the attackers are "fishing" for personal information. The most common information targeted in phishing attacks includes passwords, bank account information, credit card number, and personally identifiable information, such as name, address, social security number, and mother's maiden name. A recent report published by the Anti-Phishing Work Group (APWG) revealed that there were approximately 260,642 phishing attacks in July 2021, which was the highest monthly in APWG's reporting history [1]. The number of phishing attacks has also doubled since early 2020.

In this survey, we first give an overview of phishing and its prevalence. We also discuss the motives behind phishing and highlight the most common ones. We elaborate on different attack vectors used in phishing attacks. Most of the attack vectors of phishing are related to the Internet, but phishing can also be launched through traditional communication media. Next, we move on to the literature review of existing countermeasures of phishing. There are many solutions to mitigating phishing attacks. We narrow down the different approaches to three broad categories: education-based, machine learning-based, and blacklists/whitelists-based. We then summarize a few example solutions from each of the three categories. At last, we identify some challenges in this line of research and discuss some potential directions of improvement on this topic.

The remainder of the paper is organized as follows. Section 2 gives an overview of phishing, summarizes the motives behind phishing, and classifies the common types of phishing attacks. In Section 3, we review several existing phishing detection and mitigation techniques. Section 4 discusses the open challenges in this field, and Section 5 concludes the paper.

2. Overview of Phishing

With the first well-known incident dating back to the mid-1990s [2], phishing is still one of the most prevalent threats that we cannot put back in the bottle to this day. APWG reported an increase of 5753% of average phishing attacks per month over 12 years from 2004 to 2016 [3].

Phishing attacks initially targeted user accounts on web portals such as AOL and Yahoo!, but have gradually moved towards more profitable targets, such as online banking, payment, and e-commerce services. Individual users, usually the weakest link in the security chain, are the primary victims of phishing attacks. It was found that even after anti-phishing education, 29% of individuals still fall for phishing according to a study by Sheng et al. in 2010 [4].

Phishing is an art of deception. Today, more sophisticated phishing attacks are crafted to appear more legitimate than ever in their history. From ordinary people to political leaders in presidential campaigns, such as John Podesta, the Chairman of Hillary Clinton's 2016 Presidential Campaign [5], phishing can have a significant impact on anyone who is vulnerable.

2.1. Motives

Yu et al. concluded eight motives behind phishing from an attacker's perspective, shown as follows [6]:

1. Financial gain
2. Identity theft
3. Identity trafficking

4. Industrial espionage
5. Malware distribution
6. Harvesting passwords
7. Fame and notoriety
8. Exploits security holes

Financial gain is undoubtedly the most common motivation for phishing. Attackers can spoof the websites of financial institutions by setting up visually similar websites and gain access to the accounts of victims. Identity theft can also result in financial benefits for attackers. Stolen identities can be sold to interested parties via underground market or used for committing subsequent criminal activities such as fraud or more phishing attacks. Industrial espionage is another interesting motivating factor as the ultimate victims of phishing are no longer limited to individual users. Highly sophisticated phishing attacks are launched against specific individuals or groups within an organization to spy on the victims. Spear-phishing is a typical example of such context-aware focused attacks. Understanding these motives behind phishing attacks can help us identify prevention or mitigation techniques against phishing and build defensive infrastructures accordingly.

2.2. Attack Vectors

Phishing usually begins with a forged email sent to a victim. A good phishing email is sent from a legitimate entity not obviously unrelated to the content of the email. Correct spelling, grammar, and relevant email content, such as company logos embedded in the email, are also characteristics of a good phishing email. Part of the email's content will be a link to redirect the victim to a forged website. This website should seem legitimate and prompt the user to log in using the sensitive credentials the attacker is phishing for. These credentials are captured in several different ways. One way is to trick a victim into downloading and running a keylogger to capture any input from the victim [7]. Another way is to redirect a victim to enter the credentials on a remote server that the attacker hosts or has access to.

Other types of phishing may still be initiated via a forged email. Instead of seeking sensitive information directly, the email may contain malware embedded in an attachment that a victim is prompted to click on to download. A popular type of malware included in phishing email attacks is ransomware. This type of attack aims to gain money from the victim, most commonly via bitcoin or other cryptocurrencies.

Websites are commonly used as a definitive vector for attackers to deploy their baits in email-based phishing attacks. Victims usually give out their personal information when visiting a sophisticatedly crafted website that is visually similar to legitimate ones. In addition to mimicking authentic login pages, clickjacking is another popular means of webpage manipulation in attackers' arsenal. Victims are tricked into performing actions unknowingly via the compromised user interface (UI). Drive-by-download is another technique for attackers to inject malware into a victim's machine. It is more secretive and effective than explicit email attachments since most modern email clients are equipped with malware scanners to help users block potential malicious attachments. Another reason why click-jacking and drive-by-download are highly successful is that users tend to be less attentive to malicious websites than malicious emails, according to a survey conducted by Jakobsson [8].

Instant messaging is another common attack vector with which many recorded phishing attacks were affiliated. Users can send texts, emojis, images, files, etc., and even initiate audio and video calls on instant messaging apps. Therefore, it is not surprising that attackers will take advantage of this convenient channel to launch effective phishing attacks.

Online social networks (OSNs) have emerged since the early 2000s and have become an integral part of people's daily lives, allowing them to connect closer than ever before. Unfortunately, while it facilitates a fast sharing of information for authentic users, it also will enable attackers to deploy large-scale phishing attacks efficiently.

Besides emails, instant messaging, OSNs, and webpages, phishing can also be carried out via traditional communication media, such as Short Message Service (SMS) and voice [9]. In fact, the term "phishing" is probably inspired by the earlier hacking culture against telephone networks, "phreaking." Attackers approach their victims via text messages ("smishing") or voice calls ("vishing") to lure victims for further exploitation.

3. Countermeasures of Phishing

There are technical and non-technical solutions to preventing and mitigating phishing attacks. This section examines these different approaches based on three major categories: education, machine learning, and blacklists/whitelists.

3.1. Education

One of the primary non-technical solutions is education. It is crucial to raise awareness of the ever-growing phishing threat. Educational training and seminars are being held across the globe to inform more people about phishing and what to do when phishing is encountered. For example, APWG has partnered with Carnegie Mellon University to educate users through a phishing education program. Other educational efforts involve creating games and simulated training to warn users of phishing. Kumaraguru et

al. and Dodge et al. designed training experiments that send simulated malicious emails to users to measure users' phishing awareness [10, 11]. Users are informed about their vulnerability to phishing either at the end of the training or immediately after they click on the baits. Sheng et al. developed a game, namely *Anti-Phishing Phil*, to teach users some basic knowledge about phishing, including browser address bars, phishing pages, etc [12]. An experiment showed that this game improved novice users' ability to identify phishing by 61%. Arachchilage et al. built a mobile game and used it as a method for raising user awareness [13, 14]. They evaluated the users' learning curve when playing this game and argued that it effectively educates users compared with traditional training approaches.

The effectiveness of education programs over time remains an open question. Reinheimer et al. attempted to address this question through a survey in a public sector organization in Germany [15]. This survey evaluated the effectiveness of a newly deployed phishing education program and four different types of reminders. They concluded that it is necessary to remind users half a year after the initial education. They also found that videos and interactive examples are the most effective reminders, which extend users' awareness for another six months.

Not only can interactive training cause a longer-lasting effect on users, but interactive warnings can also heed users' attention more often than passive warnings, according to Egelman et al.'s study [16].

If users could be more careful when browsing the Internet and dealing with their emails, the problem of phishing could be substantially minimized. However, novice Internet users still do not possess such knowledge; and even if they do, many of them are reluctant to employ the best security practices to counter this daily threat, according to a survey conducted by researchers from Carnegie Mellon University [17]. Moreover, while users are becoming more educated, attackers are becoming smarter in inventing new, more sophisticated, and harder-to-detect phishing techniques. Thus, using education alone is not sufficient to mitigate the problem.

3.2. Machine Learning

Machine learning-based approaches to countering phishing attacks have been extensively studied in the past two decades since they are natural solutions to classification problems like phishing detection. Table 1 shows a summary of a few machine learning-based phishing detection approaches along with reference details, classifier used, number of features used, and accuracy.

Chandrasekaran et al. proposed a Support Vector Machine (SVM)-based approach that utilizes multiple features of emails, such as URLs, IP addresses, subjects, etc., to identify phishing emails and achieves an accuracy of 95% [18]. Fette et al. adopted ten features in their filter and tested the performance of several classifiers, including Random Forests, SVM, decision trees, etc [19]. They found that their proposed filter can correctly identify 96% of the phishing emails. Huang et al. used a specific type of SVM with an associated learning algorithm for classification and regression analysis and achieved an accuracy of 99% on a dataset [20]. Xu et al. presented a cross-layer approach to detect malicious websites using both network-layer traffic and application-layer web contents [21]. Among the four machine learning algorithms they used, the decision tree-based J48 classifier achieves an accuracy of over 99% with a total of 124 cross-layer features. Barraclough et al. focused on optimizing a neural-network-based strategy of building a model using "Neuro-Fuzzy" logic and a five-input scheme, which results in an accuracy of 98% [22]. Sahingoz et al. compared their novel phishing detection system with other published work [23]. They proposed a real-time anti-phishing system that uses seven different classification algorithms and Natural Language Processing (NLP) based features, reaching an accuracy rate of 98%. Research from Moghimi et al. proposed two novel feature sets and used a rule-based approach to implement a browser extension, called PhishDetector, with an accuracy of 99% [24]. More recent work from Chiew et al. focuses on dimensionality reduction in machine-learning techniques used in phishing detection [25]. In this work, Cumulative Distribution Function gradient (CDF-g), a novel algorithm, is used to mark the feature cut-off rank. The feature selection framework Hybrid Ensemble Feature Selection (HEFS) and the CDF-g algorithm enabled the identification of the best feature subset that contributes significantly towards the phishing detection rate.

3.3. Blacklists and Whitelists

Many machine learning-based techniques need to use reliable datasets of URLs or domain names. These datasets are commonly compiled as a blacklist - a set of malicious URLs or domain names. Whitelists, on the contrary, are collections of legitimate URLs or domain names. Whitelisting is stricter than blacklisting in the sense that only URLs that are proven to be safe are explicitly allowed; the default would be to deny access to all other URLs.

Google Safe Browsing API allows applications to validate the presence of a given URL in the blacklists [26]. However, it constantly relies on Google updating its blacklists, which introduces a performance bottleneck. Phish-

[1]The accuracy values here are claimed by the authors. It is not an apples-to-apples comparison.

[2]It uses seven different classifiers. We list the one with the best performance here.

Table 1. A comparison of ML-based solutions

Reference	Classifier	Number of features	Accuracy [1]
Chandrasekaran et al., 2006 [18]	SVM	25	95%
Fette et al., 2007 [19]	Random Forest	10	96%
Huang et al., 2012 [20]	SVM	23	99%
Xu et al., 2013 [21]	Decision Tree	124	99%
Barraclough et al., 2013 [22]	Neuro Fuzzy	288	98%
Moghimi et al., 2016 [24]	SVM	50	99%
Sahingoz et al., 2019 [23]	Random Forest [2]	40 (NLP features)	98%
Chiew et al., 2019 [25]	Random Forest	48	95%

Net is a tool that counteracts the *exact match* limitation in traditional blacklists [27]. Any change in a URL, for example, would result in a "no match" in the blacklist. Empowered by URL variation heuristics, PhishNet dramatically improves the reliability of blacklists in detecting minor differences in phishing URLs. Automated Individual White-List (AIWL) relies on a whitelist of Login User Interfaces (LUIs) [28]. There are critical features that determine whether an LUI is trusted or not, such as the URL, the corresponding IP address, the structure of the Document Object Model path for the user credentials, to name a few. To automatically maintain the whitelists, the authors use a classifier to decide whether an LUI should be added to the trusted lists or not. Sonowal et al. proposed a phishing detection model with multiple filters, one of which is a whitelist filter [29]. Although whitelisting tend to offer higher accuracy, updating the list is a major issue. This model mitigates the problem by automatically updating a URL to the whitelist if it passes through other filter layers.

4. Challenges and Discussions

Most above-mentioned technical solutions claimed to achieve very high accuracy in phishing detection. However, using the machine learning technique alone is not sufficient for mitigating the ever-growing phishing threat. Phishing is a multi-variable problem that requires a multi-variable solution. If executed properly and with the proper funding, it is argued that the risk can be reduced to manageable levels. However, it will never be fully diminished if humans are involved in processes with inherent risk.

Many technical solutions use security warnings to alert users and rely on users' proper responses to the warnings [16]. If users ignore such warnings, the solutions will be rendered in vain. Some users may be reluctant to learn; and even if they learn, the success of educational approaches depends on whether users can retain such knowledge or not. Although education is considered helpful and necessary and has been carried out extensively, a number of researchers argued that it is not as useful as we thought [30, 31]. For example, Stefan Gorling pointed out that the regulation of user behavior is dependent on many aspects other than education alone [30].

Machine learning algorithms for phishing detection depend on many features and distinctive characteristics in labeled datasets that define phishing webpages and emails. The majority of machine learning-based solutions we surveyed belong to supervised learning. Deep learning is a refinement of machine learning. It, however, does not rely on explicit human supervision. There is an ongoing trend of applying deep learning to malicious URL detection to mitigate phishing [32–34]. We believe that there is plenty of room for exploring deep learning and its applications in phishing email detection.

5. Conclusion

As phishing continues to be a highly prevalent security vulnerability, we aimed to summarize recently published research in phishing mitigation and prevention. We categorized different approaches in this research field and highlighted the findings that stood out to us. Education-based approaches aim to teach users the significance of phishing and properly handle phishing threats. On the other hand, machine learning-based approaches focus on selecting the effective features from phishing vectors and using the right classifier to detect phishing threats. Blacklists/whitelists-based approaches suffer from their own challenges, such as constant maintenance costs, scalability, and staleness. We argued that none of these approaches alone is the silver bullet for solving the phishing problem.

References

[1] APWG. Phishing activity trends report - 3q 2021, 2021.

[2] Rekouche K. Early phishing. ArXiv 2011;abs/1106.4692.

[3] APWG. Phishing activity trends report - 4q 2016, 2017.

[4] Sheng S, Holbrook M, Kumaraguru P, Cranor LF, Downs J. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010; 373–382.

[5] The phishing email that hacked the account of John Podesta. https://www.cbsnews.com/news/the-phishing-email-that-hacked-the-

account-of-john-podesta/, 2016. [Online; accessed 12-Jan-2022].

[6] Weider DY, Nargundkar S, Tiruthani N. A phishing vulnerability analysis of web based systems. In 2008 IEEE Symposium on Computers and Communications. IEEE, 2008; 326–331.

[7] Qabajeh I, Thabtah F, Chiclana F. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. Computer Science Review 2018;29:44–55.

[8] Jakobsson M. The human factor in phishing. Privacy Security of Consumer Information 2007;7(1):1–19.

[9] Rader M, Rahman SSM. Exploring historical and emerging phishing techniques and mitigating the associated security risks. ArXiv 2015;abs/1512.00082.

[10] Kumaraguru P, Rhee Y, Sheng S, Hasan S, Acquisti A, Cranor LF, Hong J. Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. In Proceedings of the APWG 2nd annual eCrime researchers summit. 2007; 70–81.

[11] Dodge Jr RC, Carver C, Ferguson AJ. Phishing for user security awareness. Computers Security 2007;26(1):73–80.

[12] Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J, Nunge E. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In Proceedings of the 3rd Symposium on Usable Privacy and Security. 2007; 88–99.

[13] Arachchilage NAG, Cole M. Design a mobile game for home computer users to prevent from "phishing attacks". In International Conference on Information Society (i-Society 2011). IEEE, 2011; 485–489.

[14] Arachchilage NAG, Love S. A game design framework for avoiding phishing attacks. Computers in Human Behavior 2013;29(3):706–714.

[15] Reinheimer B, Aldag L, Mayer P, Mossano M, Duezguen R, Lofthouse B, Von Landesberger T, Volkamer M. An investigation of phishing awareness and education over time: When and how to best remind users. In Proceedings of the Sixteenth Symposium on Usable Privacy and Security. 2020; 259–284.

[16] Egelman S, Cranor LF, Hong J. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008; 1065–1074.

[17] Downs JS, Holbrook M, Cranor LF. Behavioral response to phishing risk. In Proceedings of the APWG 2nd annual eCrime researchers summit. 2007; 37–44.

[18] Chandrasekaran M, Narayanan K, Upadhyaya S. Phishing email detection based on structural properties. In NYS Cyber Security Conference, volume 3. Albany, New York, 2006; 2–8.

[19] Fette I, Sadeh N, Tomasic A. Learning to detect phishing emails. In Proceedings of the 16th International Conference on World Wide Web. 2007; 649–656.

[20] Huang H, Qian L, Wang Y. A SVM-based technique to detect phishing URLs. Information Technology Journal 2012; 11(7):921.

[21] Xu L, Zhan Z, Xu S, Ye K. Cross-layer detection of malicious websites. In Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy. 2013; 141–152.

[22] Barraclough PA, Hossain MA, Tahir M, Sexton G, Aslam N. Intelligent phishing detection and protection scheme for online transactions. Expert Systems with Applications 2013;40:4697–4706.

[23] Sahingoz OK, Buber E, Demir O, Diri B. Machine learning based phishing detection from URLs. Expert Systems with Applications 2019;117:345–357.

[24] Moghimi M, Varjani AY. New rule-based phishing detection method. Expert Systems with Applications 2016; 53:231–242.

[25] Chiew KL, Tan CL, Wong K, Yong KS, Tiong WK. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Information Sciences 2019;484:153–166.

[26] Google. Google safe browsing API. https://developers.google.com/safe-browsing/?csw=1. [Online; accessed 12-Jan-2022].

[27] Prakash P, Kumar M, Kompella RR, Gupta M. Phishnet: predictive blacklisting to detect phishing attacks. In 2010 IEEE INFOCOM. IEEE, 2010; 1–5.

[28] Cao Y, Han W, Le Y. Anti-phishing based on automated individual white-list. In Proceedings of the 4th ACM Workshop on Digital Identity Management. 2008; 51–60.

[29] Sonowal G, Kuppusamy K. Phidma–a phishing detection model with multi-filter approach. Journal of King Saud University Computer and Information Sciences 2020; 32(1):99–112.

[30] Gorling S. The myth of user education. In Proceedings of the 16th Virus Bulletin International Conference. 2006; 11–13.

[31] Gaffney G. The myth of the stupid user, 2011.

[32] Bahnsen AC, Bohorquez EC, Villegas S, Vargas J, González FA. Classifying phishing URLs using recurrent neural networks. In 2017 APWG symposium on electronic crime research (eCrime). IEEE, 2017; 1–8.

[33] Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. Advances in Neural Information Processing Systems 2015;28:649–657.

[34] Yang P, Zhao G, Zeng P. Phishing website detection based on multidimensional features driven by deep learning. IEEE Access 2019;7:15196–15209.

Address for correspondence:

Yuan Cheng
Department of Computer Science
6000 J Street, Sacramento, CA 95819
yuan.cheng@csus.edu