ISSN: 2321-7782 (Online)
Volume 2, Issue 3, March 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Phishing Email Detection Techniques: A Review

Vishakha B. Pawar[1]
Department of Computer Science and Engineering
Sipna College of Engineering and Technology
Badnera Road, Amravati
Maharashtra – India

Pritish A.Tijare[2]
Department of Computer Science and Engineering
Sipna College of Engineering and Technology
Badnera Road, Amravati
Maharashtra – India

Abstract: ***Phishing email is one of the major problems of today's Internet, resulting in financial losses for organizations and annoying the users. Phishing is the process of enticing people into visiting fraudulent websites and persuading them to enter personal information such as usernames, passwords, addresses, social security numbers, personal identification numbers and any further information that can be made to seem plausible. This information is then used to impersonate the victim so as to empty their bank account, run fraudulent auctions, launder money, apply for credit cards, take out loans in their name, and so on. Although most current phishing attacks target the banks, phishing websites regularly appear for businesses as diverse as online auctions, payment sites, gambling websites, social-networking sites and merchants.***

Keywords: phishing email, filtering, classifiers, fraudulent.

I. INTRODUCTION

Phishing has increased enormously over the last years and is a serious threat to global security and economy. Phishing is a fraudulent way usually made through email, to steal your personal information. Phishing websites are forged websites created by malicious people to mimic real websites. Usually, the victim is convinced to perform a mouse click to download and install malicious code or access a fraudulent website without being aware of it. It is known that email is the most used Internet service nowadays, so it has been the prime resource for phishing [1].The SMTP (Simple Mail Transfer Protocol), used to send emails allows anyone to forge the sender address [2]. Also, most email clients support HTML (Hyper Text Markup Language), so all the resources of such may be used in a message. Those which supports HTML with hyperlinks (hypertext link used to divert to an invisible URL). Subsequently, through an embedded link within the email, the phisher attempts to redirect users to fake websites that are designed to fraudulently obtain financial data such as usernames, passwords, and credit card numbers. Majority of the present day phishing attacks employ e-mail as their primary carrier, in order to allure unsuspecting victims to visit the masquerade website.

Numerous approaches have been developed to filter phishing emails, yet the problem still lacks a complete solution. While the recent defense mechanisms focus on detection by validating the authenticity of the website, very few approaches have been proposed which concentrate on detecting e-mail based phishing attacks based on the structural properties inherently present in the phishing e-mail. There are a large number of possible countermeasures to phishing e-mail scams. As technical approaches like secure email authentication are not yet commonly used and require a high administrative overhead, there are countermeasures based on the contents of phishing emails.

II. TYPES OF PHISHING ATTACKS

Phishing is a type of spam that employs two techniques, malware-based phishing and deceptive phishing. For malware-based phishing malicious software is spread by deceptive emails and installed on user's machine [3]. Then the malware can capture user input and personal information can be sent to the phisher. While in deceptive phishing, phisher sends out deceptive emails pretending to come from a reputable and legitimate firm. The phisher urges the user to click a link to a fraudulent site

where the user is asked to disclose his private information. This information is used by the phisher to, e.g. to withdraw money from the users bank account [4].

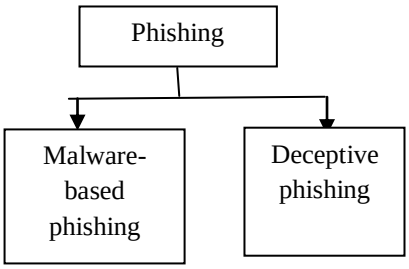

Fig.1 Types of phishing attacks

III. PHISHING FEATURES AND DETECTION TECHNIQUES

The phishing problem is treated as a pattern recognition problem, i.e. different features are extracted from emails to obtain a model that discriminates phishing messages from non-phishing ones. Hence it is a two-class pattern recognition problem.

Phishing features

The phishing detection techniques are based on identifying a set of features. The below mentioned are type of deceptive phishing. A large number of attacks are done through deceptive phishing.

The main characteristics related to phishing detection are as follows:

- Social engineering: To produce a tempting context and use of personalized information.
- Mimicry: The email and the linked website look like the official website. This includes the use of genuine design elements, trademarks, logos etc.
- URL hiding: Phishers attempt to make the URLs in the email and linked website to appear legitimate and hide the actual link address.
- Email spoofing: Phishers shows faked sender address to user and hides the actual sender's identity.
- Invisible content: Phishers insert information into the phishing mail or website, which is invisible to the user.
- Image content: Phishers create images that contain the text of the message only in graphical form.

Related work

Some of the related works to detect phishing emails is discussed in the following.

Chen proposed an image-based anti-phishing system, which is built on discriminative key point features in webpages. In that the visual similarity between the two web pages is detected and hence the legitimate and the suspicious pages are detected. One advantage of this approach is that the learning phase for the classifier is not necessary. However, if the phishing features change, the detection can fail [5].

Cook and his colleagues used a similar technique to the previous one. Some features need classification about its inclusion in the classification process, others do not provide a convincing explanation about its relevance regarding to phishing [6]. Also this approach needs to search for information on sources outside the email system (e.g. whois service), which can increase too much time needed to analyze each message.

Fette and his colleagues used a technique to detect phishing emails by including features specific to phishing [7]. They proposed 10 different features to identify a phishing email. Eight of these features can be extracted from an email itself. Of the other two features, the linked-to domain name has to be obtained with a whois query at the time the email is received. Such an approach may increase considerably the time to evaluate each message.

Camp developed a browser extension, "Net Trust" that extracts features that are not under the control of attackers to evaluate sites a user visits. In this approach, malicious websites are identified and communicated to users by gathering

information from browser history of user and inputs from social networks. Additionally, the reputation system proposed in this paper uses a rating mechanism that helps the user to assess the legitimacy of a webpage. For this purpose, the URL and the date of the initial visit are recorded for all sites a user visits. Because it is an incremental learning based method, Net Trust can flag a false negative when a user visits malicious sites for the first time.

Dhamiji presented an empirical study on the malicious strategies used by phishers when luring private credentials from online users. The researchers concluded that 90% of the participants found phishing websites attractive, and 23% of the participants ignored all of their web browsers security indicators, which led to an incorrect choice 40% of the time. However, in this research, the authors did not analyze the impact of phishing attacks that occur due to social engineering emails.

Shahriar and Zulkernine proposed a model to test suspected phishing websites based on a trustworthiness testing approach. In trustworthiness testing, the authors verified that the behavior (response) of websites matches the known behavior of a phishing or legitimate website to determine whether a website is phishing or legitimate.

Liu et al [8] detect phishing web pages or websites for a legitimate website URL. They develop an intermediate representation (in terms of features) of a legitimate page. Then suspicious URLs are detected based on heuristic rules (e.g. by replacing '0' with 'o') followed by downloading web pages from that URL. However, the heuristic might not generate all possible phishing URLs based on only this feature.

Yueand Wang [9] designed a component, "BogusBiter" that submits a large number of bogus credentials along with the actual credentials of users to nullify the attack.

A similar approach has been applied by Joshi et al. [10] (PhishGuard tool) who intercept user submitted credentials. However, to hide an actual supplied credential, they send another set of fake credentials at the end.

Krida and Kruegel [11] save a mapping between supplied credentials and corresponding trusted domains during the learning phase. In a detection phase, a submitted credential is matched with the saved credentials, and the current domain name is compared with the saved domain names. If there is no match, a website is suspected as phishing.

IV. A COMPARISON OF THE LITERATURE WORK TO DETECT THE PHISHING WEB PAGES

Name	Description	Input supply	Multiple Pages	Language independence
Chen	Image based anti-phishing technique	No	No	No
Cook	Image based anti-phishing technique	No	Yes	No
Fette	Detect phishing based on domain names	Yes	No	No
Camp	Detect through browser history	Yes	Yes	Yes
Dhamiji	Detect on social engineering i.e. online shopping	No	No	No
Shahriar and Zulkernine	Detect based on the behavior of phishing websites	Yes	No	No
Liu et atl	Detect fake web pages on visual similarities	No	No	Yes

Yue and Wang	Supply bogus credentials when a web page is detected as phishing to avoid information leakage	Yes	No	Yes
Joshi et al	Submit fake credentials before and after actual user credentials	Yes	Yes	Yes
Krida and Kruegel	Compare the domain of a visited URL with trusted domain names where a user previously submitted personal info	No	No	Yes

V. Conclusion

Phishing is a growing problem for internet users. The anti-phishing tools are proving beneficial to a certain extent to cope against this problem. Testing of phishing websites is challenging due to uncertainty of websites. However, there are still much limitation on accuracy or performance because the detection techniques are time consuming, costly. Moreover, most works were done on offline mode which requires data collection, data analysis, and a profile creation phase to be completed. Thus, there is still need of new techniques and technology that are able to solve all limitations related with phishing email detection.

References

1. Symantec, Spam decreasing, but social media phishing soaring (2010).
2. A. Herzberg, DNS-based e-mail sender authentication mechanisms: a critical review (2009) 731-742
3. Anti-Phishing Working Group, "Phishing Activities Trends REPORT," http://www.antiphishing.org/reports/apwg_report_Q1_2011.
4. M. Aburrous, et al., "Predicting Phishing Websites Using Classification Mining Techniques with Experimental Case Studies," in Seventh International Conference on Information Technology, IEEE Conf ,LasVegas, Nevada, USA, 2010, pp. 176-181
5. J. Chen, C. Guo, Online detection and prevention of phishing attacks, Communications and Networking in China (2006) 19–21
6. D. Cook, V. Gurbani, M. Daniluk, and Phishwish: a stateless phishing filter usingminimal rules Lecture Notes in Computer Science (2008) 182–186.
7. I. Fette, N. Sadeh, A. Tomasic, Learning to detect phishing emails, in: International World Wide Web Conference, 2007, pp. 649–656.
8. W. Liu, G. Huang, G. Huang, A. Fu, An antiphishing strategy based on visual similarity assessment, IEEE Internet Computing 10 (2) (2006) 58–65.
9. C. Yue, H. Wang, Anti-phishing in offense and defense, in: Proc. of the Annual Computer Security Applications Conference, ACSAC, Anaheim, California, December 2008, pp. 345–354.
10. Y. Joshi, S. Saklikar, D. Das, S. Saha, PhishGuard: a browser plug-in for protection from phishing, in: Proc. of the 2nd International Conference on Internet Multimedia Services Architecture and Applications, Bangalore, India, December 2008, pp. 1–6.
11. E. Krida, C. Kruegel, Protecting users against phishing attacks with AntiPhish, in: Proc. of the 29th Annual International Computer Software and Applications Conference, Edinburgh, Scotland, July 2005

