Review Paper on Privacy Preservation through Phishing Email Filter

Pranal C.Tayade[1], Prof. Avinash P.Wadhe[2]

[1]*ME (CSE) 2nd Sem, GHRCEM, Amravati India, S.G.B.A.U, Amravati.*
[2]*ME (CSE) 2nd Sem, GHRCEM, Amravati India, S.G.B.A.U, Amravati*

Abstract— **Phishing is the combination of social engineering and technical exploits designed which creates a replica of an existing Web page to fool users (e.g., by using specially designed e-mails or instant messages) into submitting sensitive information such as online banking passwords and credit card details, personal and financial information by masquerading as a trustworthy entity in an electronic communication. Phishing is a new type of network attack which constitutes more than half of all security incidents on the internet. Email based online phishing is critical security thread on the internet. More and more user are suffering from email based phishing attack over a last few year. Phishing email contains messages to lure victims into performing certain action, such as clicking on URL where phishing site is hosted. This paper present overview about phishing email attack, its classification and preventing approaches. Email phishing attacks fabricate the email's origin. Unfortunately, current email server systems cannot authenticate the genuineness of incoming emails.**

Keywords— **Phishing, Phishing Email, Attack, Security.**

I. Introduction

Phishing is a con game that use to collect personal information from unsuspecting users. The false e-mails often look surprisingly legitimate and even the Web pages where users are asked to enter their information may look real. Phishing email is one of the major problem of today's internet, resulting in a financial losses for organization and annoying individual users. Phishing is a form of online identity thief which is the way of attempting to acquire sensitive information such as usernames, passwords and credit card details by masquerading as a trustworthy entity in a communication network. This is similar to *Fishing*, where the fisherman puts a bait at the hook, thus, pretending to be a genuine food for fish. But the hook inside it takes the complete fish out of the lake. Communications purporting to be from popular social web sites, auction sites, online payment processors or IT administrators are commonly used to lure the unsuspecting public. Phishing is typically carried out by e-mail spoofing or instant messaging and it often directs users to enter details at a fake website whose look and feel are almost identical to the legitimate one. Phishing is an example of social engineering techniques used to deceive users, and exploits the poor usability of current web security technologies. Attempts to deal with the growing number of reported phishing incidents include legislation, user training, public awareness, and technical security measures.

The word 'Phishing' initially emerged in 1990s. The early hackers often use 'ph' to replace 'f' to produce new words in the hacker's community, since they usually hack by phones. Phishing is a new word produced from 'fishing', it refers to the act that the attacker allure users to visit a faked Web site by sending them faked e-mails (or instant messages), and stealthily get victim's personal information such as user name, password, and national security ID, etc. These information then can be used for future target advertisements or even identity theft attacks (e.g., transfer money from victims' bank account) [15].

Email has been an online 'killer application' utilized by people, businesses, Governments and different organizations for the needs of communicating, sharing and distributing data (MAAWG, 2011) [5]. Phishing email is a special type of spam message. Phishing scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication [13]. Such email is a criminal mechanism that relies on forged email claims purportedly originating from a legitimate company or bank. Subsequently, through an embedded link within the email, the phisher attempts to redirect users to fake Websites, which are designed to fraudulently obtain financial data such as usernames, passwords, and credit card numbers [1]. In these emails, they will make up some causes, e.g. the password of your credit card had been miss entered for many times, or they are providing upgrading services, to allure you visit their Web site to conform or modify your account number and password through the hyperlink provided in the email. You will then be linked to a counterfeited Web site after clicking those links. The style, the functions performed, sometimes even the URL of these faked Web sites are similar to the real Web site. It's very difficult for you to know that you are actually visiting a malicious site. If you input the account number and password, the attackers then successfully collect the information at the server side, and is able to perform their next step actions with that information (e.g., withdraw money out from your account).These fake Web sites are designed to obtain financial data from their victim fraudulently, including usernames, passwords, and credit card numbers, Occasionally, the phisher tries to misdirect the user to a fake website or to a legitimate one monitored by proxies (APWG, 2010).The damage caused by Phishing ranges from denial of access to e-mail to substantial financial loss. E-banking Phishing websites are forged websites that are created by malicious people to mimic real e-banking websites. Most of these kinds of Web pages have high visual similarities to scam their victims. Unwary Internet users may be easily deceived by this kind of scam. Phishing is a relatively new Internet crime in comparison with other forms, e.g., virus and hacking. More and more phishing Web pages have been found in recent years in an accelerative way [12]

This paper is not intended to cover related topics, such as spam, on which numerous studies have already been carried out. Phishing email is a different problem and, thus, needs more specific attention. Section 2 of the paper contains survey of phishing attack. Section 3 contains a background and overview of the phishing emails. Section 4 presents the analysis of phishing email attack and Section 5 concludes the paper.

II. LITERATURE REVIEW

Phishing emails pose a serious threat to electronic commerce because they are used to defraud both individuals and financial organizations on the Internet [1]. Phishing emails range from very simple to very complicated messages and are capable of deceiving even the clever Internet users. Fraudulent emails can steal secret information from the victims, resulting in loss of funds. This threat has led to the development of a large number of techniques for the detection and filtering of phishing emails. The many approaches proposed in the literature to filter phishing emails, may be classified according to the different stages of the attack flow, e.g. network level protection, authentication, client side tool, user education, server side filters and classifiers, etc.

The APWG Phishing Activity Trends Report analyzes phishing attacks reported to the APWG [4].The number of phishing sites detected jumped almost 30 percent, from 38,110 in June 2013 to 49,480 in July 2013, and stayed at the higher rate through the third quarter. The total number of phish observed in q3 was 143,353, a 20 percent increase over q2's 119,101. The rise is generally attributable to an increased number of attacks against money transfer sites and retail sites

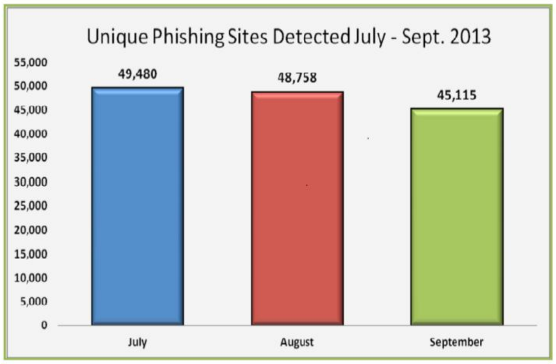

Figure 1: "Phishing Sites Detected in July-Sept 2013[4]".

According to an ecrime trends report [8], phishing attacks are increasing at a rapid rate. For example, phishing in Quarter 1 (Q1) of 2011 grew by 12% over that in Quarter 1 (Q1) of 2010. A survey by Gartner [6] on phishing attacks shows that, approximately 3.6 million clients in the US alone had lost money to phishing attacks and total losses had reached approximately US$ 3.2 billion Dollar. The number of victims increased from 2.3 million in 2006 to 3.6 million in 2007, an increase of 56.5%. Among all complaints received by the Federal Trade Commission in 2009 from Internet users, identity theft attributed to phishing email ranked first. It accounted for 21% of the complaints and cost consumers over 1.7 billion US dollars [12].

According to Luis Corrons, Panda Labs Technical Director and Trends Report contributing analyst, Trojans were once again the most prevalent type of malware, accounting for 76.85 percent of all new samples identified, and continued to be the weapon of choice for malware writers. It is worth noting the slight increase in the number of adware and spyware infections, at 6.05 percent. In the third quarter of 2013, 31.88 percent! Of computers worldwide appeared to be infected with some sort of malware or adware/spyware, almost a full point lower than in the second quarter. As for individual countries, China once again topped the list 59.36 percent of computers there appeared to be infected according to Panda Labs, a record high. China was followed by Turkey (46.58%), Peru (42.55%), and several other Latin American countries [4].

Among automatic phishing detection mechanisms, two commonly used techniques are blacklists and heuristics [2]. Blacklist-based techniques generate close-to-zero false positives and can detect most phishing attacks [4, 9]. For example, Ludl et al. demonstrated that blacklists provided by

Google (used by Firefox 2) can recognize almost 90% of live phishing sites. However, because some phishing sites may not be added into blacklists and the so called zero-day attacks may occur, researchers have proposed various heuristic-based techniques to identify phishing sites in real time [9]. These heuristic-based techniques have obtained very encouraging results. For example, CANTINA, a content-based detection tool proposed by Zhang et al. can identify 90% of phishing pages with only 1% false positives. A URL-based classifier pro- posed by Garera et al. [11] is another tool which can catch 95.8% of phishing pages with only 1.2% false positives. Currently, Firefox 2 primarily employs blacklist-based techniques while Internet Explorer (IE) 7 uses both kinds of techniques [16]. Because Bogus Biter's design is independent of any specific detection scheme, it can leverage advances in both blacklist-based techniques and heuristic based techniques to combat the majority of phishing attacks. According to the Anti-Phishing Working Group there were on average 996 unique phishing websites detected by APWG per day in 2007 [6]. On average 141 unique brands were hijacked per month in 2007[4]. While most phishing campaigns target financial institutions like banks, APWG reports an increasing number of attacks against government agencies like the US Internal Revenue Service or tax authorities in the UK and Australia. A study by Dhamija et al. [11] supports that many internet users have problems to detect phishing attacks. Even if users have the explicit task to identify phishing scams, many of them cannot distinguish a legitimate website from a spoofed website. In the study, the best phishing site was able to deceive more than 90% of the participants. In addition, users often do not know which security signs indicate the trustworthiness of a website, e.g., the padlock symbolizing secure transmission. The user study of Wu et al. comes to the same conclusions [11].

III. ANALYSIS OF PHISHING ATTACK

Internet has changed the life of human significantly by increasing the comfort of human life; on the other hand it also increases the need for security measures over the communicating network. As result of an increasing number of researchers and practitioners are attempting to quantify risks and degrees of vulnerabilities in order focus protective measures. In this section, we give a brief overview of the different types of phishing attacks to familiarize the user with the threat.

A. Types of Attack

Phishing attacks fall into several categories. The earliest form of phishing attacks were e-mail- based and they date back to the mid 90's. These attacks involved spoofed e-mails that were sent to users where attackers tried to persuade the victims to send back their passwords and account information [13]. Exploit-based phishing attacks are technically more sophisticated and make use of well-known vulnerabilities in popular web browsers such the Internet Explorer to install malicious software (i.e., mal- ware) that collects sensitive information about the victim.

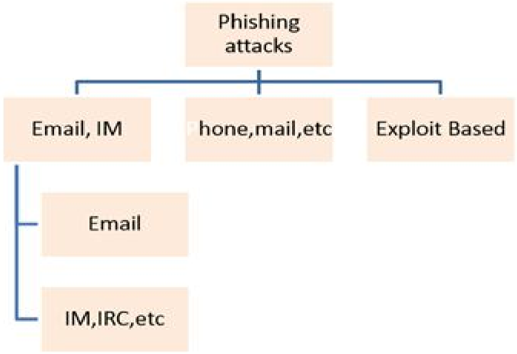

Figure 2: "Types of Attack [2]."

For malware-based phishing a malicious software is spread by deceptive emails or by exploiting security holes of the computer software and installed on the user's machine. Then the malware may capture user input, and confidential information may be sent to the phisher [6]. The deceptive phishing, in which a phisher sends out deceptive emails pretending to come from a reputable institution, e.g., a bank. In general, the phisher urges the user to click a link to a fraudulent site where the user is asked to reveal private information, e.g., passwords. This information is exploited by the phisher, e.g., by withdrawing money from the users bank account.

B. Basic Features Used in Email Classification.

Basic feature are those feature which are used for detection of phishing emails. The features can be categorized are as follows:

Structural features: Structural features can be extracted from an HTML tree, emails are send as either plain text, HTML or combination of both which explains the structure of email body such as the MIME standard that explains the message format number. These features include discrete and composite body parts with the number of alternative body parts [1].

Link features: Link features represent different features of URL links embedded in an email, such as the number of links with IP-based URL, number of deceptive links (URL visible to the user), number of links behind an image, number of dots in a link and text link "Click here to restore your account access [1]."

Element features: Element features represent the type of Web technology used in an email such as HTML, scripting, particular JavaScript, and other forms. JavaScript is used for many things from creating popup windows to changing the status bar of web browser or email client [8].

Spam filter features: A Spam feature in general has two Boolean values (0, 1). Many mail clients already have a spam filter in place, and as such it seems natural to leverage the ability of existing solutions in combating the phishing problem. Most of researchers use the Spam Assassin (S.A.) tool [1], which has more than 50 features to determine whether an email is classified as spam or not. By default, a message is considered as spam, if it scores more than 5.0.

Word list features: A list of words may possibly characterize a phishing email which can be classified by Boolean features, whether the words occur in the email or not. Word stems such as account, update, confirm, verify, secure, log, click, Update, Confirm, User, Customer, Client Suspend, Restrict, Hold , Verify, Account , Login, Username, Password and so on [1] . Phishing emails contain number of frequently repeated keywords such as suspend, verify, username, etc. We use word frequency (Count of keyword divided by total number of words in an email) of a handful of most commonly used keywords by phishers. This feature is continuous.

C. Procedure of Phishing Attacks

Phishers use e-mail as their major method to carry out phishing attacks [15]. In general, phishing attacks are performed with the following four steps:

1) Phishers set up a counterfeited Web site which looks exactly like the legitimate Web site, including setting up the web server, applying the DNS server name, and creating the web pages similar to the destination Web site, etc.

2) Send large amount of spoofed e-mails to target users in the name of those legitimate companies and organiza- tions, trying to convince the potential victims to visit their Web sites.

3) Receivers receive the e-mail, open it, click the spoofed hyperlink in the e-mail, and input the required information.

4) Phishers steal the personal information and perform their fraud such as transferring money from the victims' account.

D. Approaches to Prevent Phishing Attacks

There are several technical and non technical ways to prevent phishing attacks. Educate users to understand how phishing attacks work and be alert when phishing-alike e-mails are received; use legal methods to punish phishing attackers; use technical methods to stop phishing attackers.

Detect and block the phishing Web sites in time: If we can detect the phishing Web sites in time, we then can block the sites and prevent phishing attacks. It's relatively easy to (manually) determine whether a site is a phishing site or not, but it's difficult to find those phishing sites out in time [15].

Network level protection: Network level protection is implemented by not allowing a range of IP addresses or a set of domain forms entering the network. IP allows a website admini-strator to block message from those system that usually send spam or phishing email [1].

Authentication: Authentication based approaches to filter the phishing emails are design to conform wheather the email was send by a valid path or a valid domain name is not being spoofed.Authentication increases the security at both uses level which is employed by password as credentials and domain level which is implemented on providers site [1].

Block the phishing e-mails by various spam filters: Phishers generally use e-mails as 'bait' to allure potential victims. SMTP (Simple Mail Transfer Protocol) [15] is the protocol to deliver e-mails in the Internet. It is a very simple protocol which lacks necessary authentication mechanisms. Information related to sender, such as the name and email address of the sender, route of the message, etc., can be counterfeited in SMTP. Thus, the attackers can send out large amounts of spoofed e-mails which are seemed from legitimate organizations.

Install online anti-phishing software in user's computers: Anti-phishing software consists of computer programs that attempt to identify phishing content contained in websites and e-mail. Despite all the above efforts, it is still possible for the users to visit the spoofed Web sites. As a last defense, users can install anti-phishing tools in their computers. It is often integrated with web browsers and email clients as a toolbar that displays the real domain name for the website the viewer is visiting, in an attempt to prevent fraudulent websites from masquerading as other legitimate web sites.

VI. CONCLUSION AND FUTUR WORK

While surfing to the internet most of the users who are inexperienced falls victim to phishing activity. Phishing is a form of online identity theft that aims to steal sensitive information such as online banking passwords and credit card information from users. It is a means of obtaining confidential information through fraudulent emails that appear to be legitimate. Phishing scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication. Phishing has become a serious threat to global security and economy. Because of the fast rate of emergence of new phishing websites and because of distributed phishing attacks. Although phishing scams have received extensive press coverage, phishing attacks are still successful because of many inexperienced and unsophisticated Internet users. Attackers are employing a large number of technical spoofing tricks such as URL obfuscation and hidden elements to make a phishing web site look authentic to the victims. This paper present a survey on phishing email attack, feature of email classification, and procedures of phishing attack and its prevention approaches. Antiphishing is the solution over a phishing problems. The number of antiphishing software are present but no single technology will completely stop phishing. This survey improves the understanding of the phishing emails problem, the current solution space, and the future scope to filter phishing emails.

ACKNOWLEDGMENT

I would like to thank to all the people those who have help me to give the knowledge about these research papers and I thankful to my guide with whose guidance I would have completed my research paper and make it to published, finally I like to thank to all the website and IEEE paper which I have gone through and have refer to create my review paper successful.

REFERENCES

[1] Ammar Almomani, B. B. Gupta, Samer Atawneh, A. Meulenberg, and Eman Almomani., A survey of phishing email filtering techniques, IEEE communications surveys & tutorials, vol. 15, no. 4, fourth quarter 2013.

[2] Jyoti Chhikara, Ritu Dahiya, Neha Garg, Monika Rani, Phishing & Anti-Phishing Techniques: Case Study, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013 ISSN: 2277 128X.

[3] Aryan Chandrapal Singh1, Kiran P. Somase^2, Keshav G. Tambre3, Phishing: A Computer Security Threat, International Journal of Advance Research in Computer Science and Management Studies Volume 1, Issue 7, December 2013 ISSN: 2321-7782.

[4]APWG, Phishing Activity Trends Report, 3rd Quarter 2013. http://www.apwg.org.

[5] Ammar ALmomani; B. B. Gupta; Tat-Chee Wan; Altyeb Altaher; Selvakumar Manickam, Phishing Dynamic Evolving Neural Fuzzy Framework for Online Detection “Zero-day” Phishing Email, Indian Journal of Science and Technology, Vol: 6 Issue: 1 January 2013 ISSN:0974-6846.

[6] Jingguo Wang,Tejaswini Herath,Rui Chen,Arun Aishwanath, and H. Raghav Rao, Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email,” IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, VOL. 55, NO. 4, DECEMBER 2012.

[7] Ram Basnet, Srinivas Mukkamala, and Andrew H. Sung,” Detection of Phishing Attacks: A Machine Learning Approach”, Soft Computing Applications in Industry, STUDFUZZ 226, pp. 373–383, 2008.

[8] Ian Fette, Norman Sadeh, Anthony Tomasic, Learning to Detect Phishing Emails, Track: Security, Privacy, Reliability, and Ethics WWW 2007.

[9]S. Garera, N. Provos, M. Chew, and A. D. Rubin., A framework for detection and measurement of phishing attacks. In Proceedings of the WORM, 2007.

[10] W. Liu, X. Deng, G. Huang and A. Y. Fu, An Antiphishing Strategy Based on Visual Similarity Assessment, Published by the IEEE Computer Society 1089-7801/06 IEEE, INTERNET COMPUTING IEEE, 2006.

[11] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 581–590, 2006.

[12] Engin Kirda, Christopher Kruegel, Protecting Users Against Phishing Attacks, Published by Oxford University, The Computer Journal Vol. 00 No. 0, 2005.

[13] Engin Kirda and Christopher Kruegel, Protecting Users Against Phishing Attacks with AntiPhish, Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC’05)IEEE 0730-3157/05 $20.00 © 2005.

[14] Gunter Ollman. The Phishing Guide - Under- standing and Preventing Phishing Attacks. White Paper, Next Generation Security Software Ltd., 2004.

[15] Juan Chen, Chuanxiong Guo, Online Detection and Prevention of Phishing Attacks, The National Natural Science Foundation of China (NSFC) under contract No. 60503049.