ISSN No. 0976-5697

Volume 7, No. 4, July-August 2016

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A Survey Paper On Phishing Detection

Himani Thakur
Computer Science Engineering
PURCITM ,Mohali,India

Dr.Supreet kaur
Computer Science Engineering
PURCITM ,Mohali, India

***Abstract*—** With the evolution of internet in the last few years , phishing scams have also rapidly grown which is posing a great threat to Internet security globaly. Phishing is one of the most common and serious security threat over Internet where cyber attackers generally try to deceive users into revealing confidential information or financial credentials by using either malwares or some other social engineering platforms. Because of huge damage and financial loss caused by such attacks, detection of phishing is of great importance and also has been an area of great interest. No one phishing detection method is adequate enough due to the fact that there are several ways to carry out such an attack and this has led to various phishing detection techniques. In this paper two major concerns related to phishing have been addressed. Firstly we have addressed the history and motivation of hackers that led them to these attackes. We also provide classification of various types of phishing attacks. The second issue addressed is related to the various solutions that have been proposed to detect and defend from phishing attacks. It provides a better understanding of the existing problem, available solution space and scope of future work against such attacks.

***Keywords*—** *Anti-phishing;Data mining;Internet security*

I. INTRODUCTION

One of the most money-making offence since past is ‘‘identity theft’’, which means to filch any person’s identity. In a conventional term [1], criminals perform these either by homicide the victim and pretend to be that person or steal private information from the garbage by entering information from remaining letters, financial record, electricity bills and many others bills which are discarded without shredding them properly [2]. The term ‘‘phishing’’is imitative from the similarity of ‘‘fishing’’ for victims’ passwords and documentations in the web. The expression ‘‘ph’’comes from ‘‘phone phreaking’’, which was very general technique that bothered telephone systems during 1970s.

The phrase ‘‘phishing’’ was used for the first time over the Internet by a group of hackers in 1996, who shawl America Online (AOL) accounts by traping unaware AOL users into disclosing their

passwords [1].Phishing can be referred to as an computerized identity theft, which takes the benefit of human nature and the Internet to trap millions of people and take a great amount of money. It has been pragmatic that in last few years phishing attacks have grown speedily posturing a real threat to global security. The main endeavour of these campaigns is to develop the vulnerabilities present in the system, which may be either technical or due to user lack of knowledge, which means that researchers have to afford defense against these bothers at both the user level and the technical. Researchers have tried to realize the former by employing various loom, and the final can be possible by increasing consciousness and educating the Internet users. Phishing campaigns effort to pull out secret data from the victims, which may lead to significant financial losses. Studies have shown that one-third of all the phishing attempts in 2013 were proposeed toward bank accounts or to gain other financial information [3].

Since 2012, financial phishing assaults were increased by 8.5 % as contrasted to 2011, an all-time high accountable for phishing attack [4]. In spite of causing severe financial damages to the users across the Internet, spam and phishing are immobile growing at a faster rate, and it will carry on to do so as long as 1 out of 100,000 recipients actually counters to the phrases like ‘‘Click here’’ in spam emails. According to the Anti- Phishing Working Group (APWG) reports, phishing tricks will keep growing with the use of more superior technologies, and it will become the main risk over Internet,beating spam behind, as phishing scams are increasing 56 % per month [5]. In this paper, present an general idea of phishing attacks and many possible defense schemes. This review gives a broader classification of fense methods, and we provide a set of description used for phishing finding associated with these qualities ranked according to their talents to classify the phishing emails successfully. We also provide nomenclature of various solutions proposed in the literature that can detect and defend from phishing attacks. In accumulation,there is a discussion of various issues and challenges to agreement with phishing attacks. There is also review of various tools and datasets used by the researchers for evaluating of their looms. The respite of the paper is managed as follows: Section 2 of the paper presents statistics of phishing attacks. Section 3 describes motivation of hackers. Section 4 presents classification of various types of phishing attacks. Section 5 explains phishing defense mechanism. Section 6 presents open issues and challenges against phishing detection, and finally, Section 7 concludes the paper.

Figure 1 (4th Quarter 2015 Phishing Activity Trends Summary)[6]

II. PHISHING STATISTICS

The word ‘‘phishing’’ was utilized for the first time over the Internet by a group of hackers in 1996, who shawl America Online (AOL) accounts by traping uninformed AOL user into giving their passwords [1]. Fig.1 confirms 4th Quarter 2015 phishing activity trends.According to the APWG report [6], the entirety number of exclusive phishing sites were 158,574. APWG noted a large point in phishing from November to December 2015, with an augment of over 21,000 phishing sites noticed during the holiday period. USA lingers the most targeted country for these attacks [6]. Most of the phishing movements used spitefully registered domains and subdomains. The number of domains has raised from 260 million in April 2013 to 272 million in November 2013 [7, 8]. The annoys embattled 82,163 unique domain names, which is again significantly larger than 53,685 throughout the first half of 2013. Out of the 22,831 registered deception domains, 1541 used well-known kind names. Moderately than using domain names, some of the attack attempts used IP address, and figures showed about 2400 such molests used approximately 840 IP addresses [9]. as a outcome to a survey in 2013 [10], 62 % associations were instigate to be a casualty of spear phishing, whereas the review by InfoSecurity [11] showed that 42 % organizations had faced these attacks. Overall 20 % (18 % in RSA and 32 % in InfoSecurity) said that they have not looked such assaults and 21 % did not decide whether that occured or not. The organizations with more than 1000 employees have a higher possibility to become a butt of spear phishing [10, 12].

The United States Computer Emergency Readiness Team composed protection incident reports from federal, state and local government agencies and development 107,655 incident reports in 2011, with 43,889 of them relating federal agencies. After dispensation these incident reports, they found that more than half of those event reports (Approx. 51.2 %) came from phishing. Therefore, for attainment a foot into the door of a government network, most popular means is phishing to the hackers by a extensive scope [5]. The study is that gauges of phishing attacks by means of eCrime Trend Reports. In fourth quarter of 2013, According to [13], .com is the mostly compose use of domain for phishing attacks with 41 %, tracked by.net with 6 %, .org with 5 %, .br with 4 % and residual IP address based with 4 % .We also found that USA is the most admired country for hosting phishing websites with 45 %. Next mainly popular phishing websites hosting country is Germany with 6 %, Canada with 3 %, France with 5 %, UK with 4 %, Brazil with 3 %, Russia with 2 % and Poland with 2 % .

III. MOTIVATION AND WHY PEOPLE FALL FOR PHISHING ATTACKS

The motivation for phishing attacks has amended over the time, and will carry on to go forward into the future as well. Most of these attacks consequences into economic loss and is key motivation behind these attacks.

One of the main reason why phishing attacks are successful and people fall for such attacks is lack of awareness and ignorance among human beings about warning messages. One of the earliest investigating where members were invited to identify various Web sites as legitimate or forge demonstrated that 90% of members were tricked by good phishing sites Many members incorrectly judged sites based on Web pages content without understanding that these were copied.

Further studies demonstrated that women and younger members (ages 18 to 25) were more vulnerable to such attackes than men, mainly due to having less exposure to technical knowledge, less online familiarity, and not much exosed to trainings on phishing. Hence, a deeper and thoughtful understanding of hacker’s motivations, beliefs, and psychological models of people exposed to phishing is critical for the phishing protection society to figureout effective counter measures. considering what hackers are after is a measure step in being able to stop them.

Information Theft – When the attacker aims to obtain information owned by the target and/or stored in the target’s network. This information may be in the form of client information, business-critical information, or intellectual property.

The assailants managed to infiltrate the security company’s network through carefully-crafted spearphishing mail, which conceded malware that exploited certain Adobe Flash Player vulnerabilities. From there the attackers stole all the data they can find.

Espionage – When the purpose of the attacker is to observe the activities of the targets and steal information that these targets may have—such as information that could conciliation national security.

Sabotage – When the aim of the attacker is the devastation, denigration or blackmail of its targets.

IV. CLASSIFICTION OF PHISHING ATTACKS

On broader perspective phishing attacks can be classified into two categories: social engineering or deceptive phishing and malware-based phishing attacks.

Social engineering phishing attacks generally engage psychological exploitation of users or tricking company employees into handing over their private data. [14–16] These attacks occurs through fake emails, which seems legitimate otherwise or some other social platforms that appeals to certain emotions in the victim, where victim ends up in click a malicious link, or releasing sensitive information, The users with less technical expertise fell easily for social engineering attacks, so endeavors must put efforts to educate employees against these attacks, in order to stay two steps ahead of hackers and prevent these attacks from succeeding

Similarly, malware-based phishing engages running malicious software or unnecessary programs on the user’s machine. This is a general threat for small and medium businesses (SMBs).

Further these attackes can be classified as: key loggers/screen loggers, Man-in-the-Middle Phishing, session hijacking, host file poisoning, DNS phishing, Search Engine Phishing and content injection.

Some of the methods or measures used to carry on these phishing attacks are summarized in the following paragraphs.

V. PHISHING DEFENCE MECHANISMS

Phishing emails are being sent with purpose of stealing confidential information from the victims.

Most people fell victim to phishing attacks because of abandon and careless internet browsing. Companies should inform their employees about the traps and plans of phishers.

In this section there will be discussing the machnisms to oppose phishing attacks and features used to recognize

phishing. Some spam filters utilize hundreds of features to filter out phishing emails. These features[17] for detection of phishing emails can be categorized as:

1. Body-based features: These characteristics are extracted from the email body. They contain binary characteristics such as occurrence of forms, HTML or certain phrases and links in the email body.
2. Subject-based features: Some characteristics are extracted from the subject of an email such as whether it is a reply to some preceding mail, or the presence of certain words like verify, debit.
3. URL-based features: These characteristics check whether an IP address is used instead of domain name, the presence of @ in the links, number of images, external and internal links in the email text, the count of periods in the links, etc.
4. Script-based features: These characteristics check for the occurrence of JavaScript, pop-up window code, onClick events, etc., in the email.
5. Sender-based features: These characteristics consist of sender's details such as dissimilarity between the sender's address and the reply to address.

A. Classification of protection against phishing

1) User education:

User education refers to extending awareness and instruction about phishing among Internet users. Education-based approaches offer online information about threat of such attacks and their avoidance techniques [18]. Some approaches also offer online training and testing to the users.

2) Software-based defence approaches:

a) Protection at network level:

In this loom,certain range of IP addresses or a set of domain is not authorized to enter the network. DNSBLs[19] utilize the DNS protocol and are generated and updated frequently by observing the network traffic. An open-source software grunt can also be used at the network level although these require continuously updated.

b) Authentication-based mechanisms:

In this loom, it is inveterated whether or not the message was sent by a suitable path and domain name and can be utilized at both the user and domain level. These methods improve the security of email communication. The confirmation schemes are quite simple and can be done at the domain level or by digitally signing the document before sending.

c) Client-side tools :

These contain user profile filter and browser-based toolbars. Other techniques are domain checks, URL examination, etc. These tools also depend on blacklisting and whitelisting techniques where a list of detected phishing or legitimate websites is downloaded with updates at standard intervals. The limitations of these techniques are their fail to detect zero-day attack.

VI. OPEN ISSUES AND CHALLANGES

Various solutions to manage phishing attacks have been given in the literature. Though, we can articulate that no result is a ''bullet of silver'' against phishing. With time, phishing intimidation is increasing and becoming a general trick to do e-crime. Every time, when researchers come with any scheme to control this problem, phishers change their assail strategy by developing vulnerabilities found in the current solution. Therefore, we can be say that it is a very rigid race between phishers and researchers. Phishing tricks could be committed either by social engineering or by using malicious codes. In social engineering method, phisher utilized either spoofed emails or false websites to trick the users and do fraud. Therefore, solutions are also based on these surveillances. The blacklisting and whitelisting approaches have low FP rates and are very incompetent for the detection of zerohour phishing attacks, i.e., these looms are able to identify only about 20 % of such attacks. They also require communication over the network, which lessers the performance.

PhishNet [21] requires high bandwidth so as to increase the blacklist. The Google safe browsing API [20] aims to lower the bandwidth requirements. In case of AIWL[22], the efficiency totally depends on how the customer trains his/her browser. The instrument learning and data mining looms provide the best consequences in phishing detection. Chandrashekharan et al. [23] applied structural characterstics with SVM to detect phishing attacks with 95 % accuracy. However, this loom is very time-consuming, even for a small dataset. The accuracy of the system using SVM can be increased up to 97 %. PILFER[27] also gives about 95 % accuracy. But the FP and FN rate show that substantial number of emails is not well classified.

Similarly, robust classifier model[28] is 99.8 % correct. But, it is a time-consuming method as it imposes due to its five stages and used datasets are not criterion. Phishing detection by heuristics also gave good results. But some of them have very high FP rates, e.g., Spoof- Guard [24] and PhishWish[25]. In Phishwish, since there are 11 rules to be followed, it is not adaptive to changes in the situation. CANTINA [26] also has high FP rate in addition to its time-consuming processing. Another challenge with these looms is the frequent update time which makes it quite costly. User understanding is an significant issue, for defense against phishing hits. Along with an increase in the user education, some other remedies could be improvement in the user interfaces, i.e., giving active warnings and automatically detecting malicious messages.

In recent times, one of the latest areas, i.e., IoT, has also become a victim of phishing attacks. IoT is a very fast evolving architecture these days connecting every day-today object making our lives more comfortable. But, due to incomplete resources existing to the IoT devices, their security mechanism is not very strong which makes them a very easy target for the attackers [29-31]. In January 2014, Proofpoint unleashed the first spam and phishing attacks on IoT devices such as refrigerators and smart TVs; the attackers utilize these devices as a medium to send about 100,000 emails holding malwares. Once ruined, the IoT devices are needed to be bought offline to remove malware and those which were not are immobile infected. In the year 2013, 20 billion gadgets were connected to Internet, and this numeral will enhance to 32 billion by the year 2020. Smart fixation are the future, and everyone is appreciating it but these devices are also making the job of attackers easy [32-34].

VII. CONCLUSION

It has been approximately 20 years since the phishing problem was acknowledged. But, still it is used to steal personal

information, online documentations and credit card details. There are diverse solutions offered, but whenever a result is proposed to overcome these attacks, phishers come up with the vulnerabilities of that solution to maintain with such an attack. Phishing attacks can be classified generally into two categories: Social engineering, which refers to obtaining user's testimonial using emails or fake websites, and malware attacks, which use malicious code or software to obtain the data required. There are several approaches to shield the user from email and website phishing and were examined in this document.The appraisal helps new researchers to identify with the history, current inclinations of attacks and failure of various accessible solutions. Defense against phishing attacks is one of the hardest confronts faced by the network security these days.

A good defense mechanism should be capable to identify phishing attacks with low false positives. The defense techniques discussed in this study are blacklisting, data mining and heuristics, machine learning and soft computing algorithms. Blacklisting techniques have negligible FP rates but consume a lot of bandwidth and should be avoided if there is a risk of zero-hour attacks. The heuristic and data mining techniques have high FP rates than blacklists with high computational costs but better at identifying zero-hour attacks. The machine learning techniques give the best results as evaluates to other techniques as they are able to alleviate zero-hour phishing attacks enhanced than the other. Some of the machine learning methods [27, 28] are able to identify TP up to 99 %.We know that be deficient in the awareness among the users is also a factor that relates to victory of phishing attacks. Thus, educating the user is also a necessity to lesser the phishing attacks, besides improvements in the interfaces that give warnings or the automatic exclusion of malicious content before the end-users would be a more promising approach. After the classification, there is also described various issues and challenges in present solutions to recognize new researcher about the idea for future study by protecting against phishing attacks.

VIII.REFERENCES

[1] The Phishing Guide Understanding & Preventing Phishing Attacks By: Gunter Ollmann, Director of Security Strategy, IBM Internet Security Systems, 2007

[2] Phishing: Cutting the Identity Theft Line Published by Wiley Publishing, Inc. 10475 Crosspoint Boulevard Indianapolis, IN 46256 www.wiley.com, 2005, Rachael Lininger and Russell Dean Vines

[3] Anti-Phishing Working Group (APWG), ''Phishing activity trends report—first quarter 2013.http://antiphishing.org/reports/apwgtrendsreportq12013.pdf, accessed September 2014

[4] Aloul F (2010) The need for effective information security awareness. Int J Intell Comput Res 1(3):176–183

[5] James L (2005) Phishing exposed. Syngress Publishing, Burlington 6. Anti-Phishing Working Group (APWG) (2015) Phishing activity trends report- fourth quarter 2015 https://docs.apwg.org/reports/apwg_trends_report_Q4_2015.pdf.

[6] Anti-Phishing Working Group (APWG) (2014) Phishing activity trends report—first quarter 2014. http://antiphishing.org/reports/apwgtrendsreportq12014.pdf. Accessed Sept 2014

[7] Anti-Phishing Working Group (APWG) (2014) Phishing activity trends report—fourth quarter 2013.http://antiphishing.org/reports/apwgtrendsreportq42013.pdf. Accessed Sept 2014

[8] Anti-Phishing Working Group (APWG) (2014) Phishing activity trends report—second quarter 2013.http://antiphishing.org/reports/apwgtrendsreportq22013.pdf. Accessed Sept 2014

[9] Anti-Phishing Working Group (APWG) (2014) Global Phishing Survey—second half 2013. http://antiphishing.org/reports/apwgglobalphishingreport2h2013.pdf. Accessed Sept 2014

[10] IT Business Edge (2014) Spear phishing, targeted attacks and data breach trends. http://www.itbusinessedge.com/slideshows/spear-phishing-targeted-attacks-and-data-breach-trends-04.html. Accessed on Sept 2014

[11] Pierluigi Paganini (2014) Phishing: a very dangerous cyber threat http://resources.infosecinstitute.com/phishingdangerouscyber-threat/2012. Accessed on Sept 2014

[12] Krebs B (2014) HBGary federal hacked by anonymous. http://krebsonsecurity.com/2011/02/hbgary-federal-hacked-by-anonymous/2011. Accessed Sept 2014

[13] eCrime Trends Report: Fourth Quarter (2013) http://Internetidentity.com/resource-tags/quarterly-ecrime-reports/. Accessed Sept 2014

[14] Jakobsson M, Myers S (2007) Phishing & countermeasures:understanding the increasing problem of electronic identity theft. Wiley, New York

[15] Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J, Nunge E (2007) Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: Proceedings of the SOUPS, Pittsburg, pp 88–99

[16] Markus Jakobsson SM (2007) Phishing and countermeasures, Microsoft's anti-phishing technologies and tactics. 18 MAY 2007, pp 551562.

[17] Dhamija R, Tygar JD, Hearst MA (2006) Why phishing works,''in proceedings of the 2006 conference on human factors in computing systems (CHI). ACM, Montre´al, Que´bec, Canada, pp 581–590.

[18] Chou N, Ledesma R, Teraguchi Y, Mitchell JC (2004) Clientside defense against web-based identity theft. In: NDSS. The Internet Society.

[19] Levine J (2008) DNS blacklists and whitelists, IRTF anti-spam research group, Nov 2008, Internet Draft draft-irtf-asrg-dnsbl-08.txt.

[20] Google (2014) Google safe browsing lookup API.https://developers.google.com/safebrowsing/lookup_guide/. Accessed Oct 2014.

[21] . Prakash P, Kumar M, Kompella RR, Gupta M (2010) Phishnet: predictive blacklisting to detect phishing attacks. In: Proceedings of the 29th conference on information communications INFOCOM'10. IEEE Press, Piscataway, NJ, USA, pp 346–350

[22] Cao Y, Han W, Le Y (2008) Anti-phishing based on automated individual white-list. In DIM'08: proceedings of the 4th ACM workshop on digital identity management. ACM, New York, NY, USA, pp 51–60

[23] Netcraft (2014) Netcraft toolbar, 2006. http://toolbar.netcraft.com/. Accessed Sept 2014.

[24] Likarish P, Dunbar D, Hansen TE (2008) Phishguard: a browser plug-in for protection from phishing. In: 2nd International conference on internet multimedia services architecture and applications, IMSAA, Bangalore, India, pp 1– 6

[25] Zhang Y, Hong JI, Cranor LF (2007) Cantina: a content-based approach to detecting phishing web sites. In: Proceedings of the

16th international conference on World Wide Web, ser. WWW'07. ACM, New York, NY, USA, pp 639–648

[26] Chou N, Ledesma R, Teraguchi Y, Mitchell JC (2004) Clientside defense against web-based identity theft. In NDSS. The Internet Society.

[27] Angelov PP, Filev DP, Kasabov N (2010) Evolving intelligent systems: methodology and applications, vol 12. Wiley, NewYork

[28] ALmomani A, Wan T, Al-Saedi K, Altaher A, Ramadass S, Manasrah A (2011) An online model on evolving phishing E-mail detection and classification method. J Appl Sci 11(18):3301–3307.

[29] Atzori L, Iera A, Morabito G (2010) The internet of things: absurvey. Comput Netw 54:2787–2805

[30] Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of things (IoT): a vision, architectural elements, and future directions. Future Gener Comput Syst 29(7):1645–1660

[31] Roman R, Najera P, Lopez J (2011) Securing the internet of things. Computer 44(9):51–58

[32] Koroneous GL (2015) Enterprise tech spotlight: IoT tipping point, phishing scams, retail breaches.http://news.verizonenterprise.com/2015/08/iotretail-breaches-phishing-security/

[33] Bertlucci J. Internet of thingbots: the new security worry http://www.informationweek.com/ big-data/big-ata-analytics/internetof-thingbots-the-new-ecurity-worry/d/d-id/1234973

[34] Gorman M. The internet of things isn't safe: thousands of smart gadgets hacked to send spam and phishing emails .http://www.engadget.com/2014/01/17/internet-of-things-hacked-maliciousemail-phishing/

