Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)
IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

Variants of phishing attacks and their detection techniques

G. Jaspher Willsie Kathrine
Department of Computer Science and engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
kathrine@karunya.edu

A. Amrutha Rose
Department of Computer Science and engineering
Karunya Institute of Technology and Sciences
Coimbatore, India.
amruthaprema98@gmail.com

Paradise Mercy Praise
Department of Computer Science and engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
mercypraise97@gmail.com

Eligious Kalaivani. C
Department of Computer Science and engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
Eligioussharma1997@gmail.com

***Abstract*—Phishing is a treacherous effort to steal private data from users like address, aadhar number, PAN card details, credit/debit card details, bank account details, password for online shopping sites, etc. Pinching or phishing of private information on the web has caused havoc on a majority of users due to the lack of internet security. Phishing attacks make use of fake emails or websites, intended to fool users into revealing personal or financial information by posing as the trusted bank/shopping site. The various types of phishing attacks and the recent approaches to prevent the attacks are discussed. A framework to detect and prevent phishing attacks is also proposed. A combination of supervised and unsupervised machine learning techniques is used to detect known and unknown attacks.**

Keywords*— *Phishing, Internet, Security, Emails, Machine learning.

I. INTRODUCTION

Phishing is an attack which targets a user rather than the system. The term phishing was introduced in 1987 [1]. Phishing is to dishonestly carry out fraudulent financial transactions on behalf of users using a fraud email that contains a URL pointing to a fake website camouflaged as a government entity or an online bank [2]. Phishing attack can also be executed automatically an automated identity theft. This type of phishing attacks always takes advantage of nature of human and the vastness of internet to trick people. It may also result in the loss of large amount of money [3]. In more recent times it has become evident that phishing attacks now pose a significant threat to the global security. Due to the vast nature of phishing attacks, the attack detection identification begins by defining the problem and analyzing the phishing life cycle and effects. Characteristics of anti-phishing solutions from this have to be derived. Detection of this phishing website is based on blacklisted URL. Many other approaches for detecting phishing attack exist such as black list, white list based approach, Fuzzy rule based approach, machine learning approach, cantina based approach, image based approach and Heuristic approach [4], [5]. Phishing websites appear to be like a valid website and many users have difficulty in identifying the fraudulent website. Some anti-phishing tools are inbuilt in browsers [6].

II. TYPES OF PHISHING

There are various types of phishing attacks. The phishing attacks are based on their implementation methodologies. Some of the well known types are similar spoofing, instant spam spoofing, Hosts file poisoning, malware based phishing, Man-in-the middle, session hijacking, DNS based phishing, deceptive phishing, key loggers/loggers, Web Trojans, Data theft, Content-injection phishing, Search engine phishing, Email /Spam, Web based delivery, Link Manipulation, System reconfiguration, Phone phishing, Data shop lifting, black listing. The types of phishing attacks are shown in figure 1.

Figure 1. Types of Phishing Attacks

In similar spoofing, the phishers take great efforts in developing the web pages completely similar to the legal site [7]. The Uniform Resource Locator (URL) of phishing site and legal site will look similar and the user will be misguided. In instant spam messages the phishers send repeated messages to users notifying there is some problem with their internet banking or other financial applications and request them to change their password or entering sensitive information [8]. The link displayed in the mail looks similar to the legal URL because of which, the user may attempt to enter information. The information which is entered by the valid users will be misused by the phisher. Deceptive phishing is similar to instant spam messages technique since it also uses messages to obtain information such as bank account details, personal information from the user [9]. Messages are sent to victims to verify account information. The reason for the loss of data can be claimed to be system failure which necessitates the user to compulsorily re-enter all the personal details of the victim. Malware based phishing is a technique that is likely to install and run corrupt software or malicious code on the victim's machine [10]. This malicious code is linked to all the emails associated with financial transactions of the victim. When the user unknowingly clicks on the malicious link, it gets all online account information. Malware based phishing mostly affects small and medium business which do not always update their software applications. Key logger and screen logger based phishing endeavours to get victim information through keyboard inputs. This phishing technique involves recording all the key strokes of the keyboard without the complete knowledge of the victim. [10]. Key logger and screen logger type of phishing attack is an effective method where an attacker can repeatedly steal passwords and confidential information from the victims. Session hijacking also called as cookie hijacking utilizes the valid session of the user [10]. Session hijacking tries to gain access to the valid session key in order to perform theft or to perform pseudo authentication in the place of a valid user to a remote server [11]. Web Trojans [13] are malicious programs which are implanted in the user's system through valid means of access like emails, valid links, etc. In this technique, the attacker places a time bomb which might inadvertently explode when the user attempts to perform a normal operation like opening a file, accessing the internet, etc. These trojans collect the local records of the user and transmits them to the phisher. In host file poisoning attack, when the user types a URL to visit a website, the phisher has to intercept the URL before normal communication is established [14]. In Microsoft Windows operating system, there is a "hosts" file which is analyzed before resolving the Domain Name System (DNS). By "poisoning" the hosts file, the hackers have a bogus address attached. So, when a user unintentionally enters a valid URL, the link takes the user to a fake website where their information will be stolen. Since Small and Medium Business (SMB) use Windows operating system, the occurrence of such attacks is unavoidable [6] [15]. Social phone or email based phishing attacks are also common [15]. Insecure personal computers (PC) often contain sensitive information which is stored elsewhere on secure servers. Personal Computers are used to access such servers which can result in theft of personal data [8]. Sometimes sensitive information such as design documents, confidential communications, employee details etc can also be stored in the user PC. Attackers get profit by selling information to illegal person [14]. Content-injection phishing is an attack; where hackers change the whole or a part of the content of a genuine site with sham content to deceive or misguide the user into voluntarily give their classified information to the attacker [16]. This can be done by hackers inserting a malicious code to user's PC or operating network by which they can gather all information and transfer it to the hacker's phishing server. In man-in-the middle type of attack, the hackers place themselves between the common user and a genuine website [6]. They record the information being entered but do not interfere and continue to pass it on so that normal user transactions are not affected. Hackers can later leisurely have a glance at all the information they have recorded and sell them when the user is not active on the system [3]. Search engine phishing occurs when Phishers build websites which look attractive and also includes "special" offers which are also listed legitimately in the search engines. When a user attempts to order those "offer" products they are instructed to enter their credit card details [12]. These details will be easily collected by attackers. In Email /Spam based phishing attack, phishers try sending email to several people by making a request to enter their personal information, or update their information even include by asking people to renew their subscription, etc [13]. This information which is being filled and renewed is used by the attacker to perform illicit scheme [20]. In web-based delivery attack, the phisher detects all the details that are transferred during a transaction between the genuine website and the user [13]. As the user continues to pass information, it is been gathered by the phisher all without the user recognizing about it. In link manipulation attack, the attacker forwards a link to a malicious website. When the user clicks on the illusive link, it opens up the phisher's website rather than the website mentioned in the link. Placing the mouse over the link in order to view the genuine address will prevent users from avoiding link manipulation attacks.

III. APPROACHES OF PHISHING DETECTION

There are several methods to detect phishing attacks. These attacks are to be revised from time to time due to the innovativeness of the new phishing attacks. Figure 2 shows the different approaches existing to detect phishing attacks. In heuristics approach, the attempt is made to understand the analysis of phishing websites and recognize attacks based on the different features such as name of domain, domain age, spelling error, image source etc [14]. A spoofguard [23] is a browser plugin with certain web browsers like Internet explorer. Spoofguard examines the current domain name, the websites and analyzes the information to indicate originality of the website. In the blacklist approach, non-trusted URL's or

a list of banned websites is added in a list named as blacklist [7].

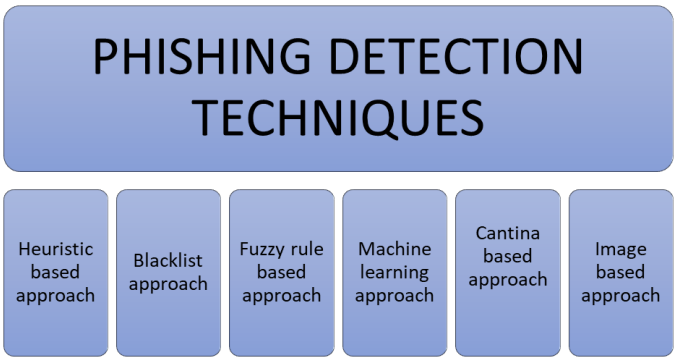

Figure 2. Approaches to detect phishing attacks

This method checks all the known phishing websites. The currently entered URL is checked with the malicious declared list. The contents are also analyzed and if the URL content matches, then the URL is blocked, and warning is issued to the user. The count of the total phishing pages is also listed in the blacklist [15]. In fuzzy rule-based approach, fuzzy logic-based data mining algorithm is used to experiment and find the phished websites [16]. In machine learning approach, different machine learning algorithms like, random forest algorithm, Support vector machine (SVM), swarm intelligence, genetic algorithm, etc re used. SVM has been used productively to solve many of the classification problems [17]. In cantina-based approach uses both Term Frequency and Inverse Document Frequency (TF-IDF) for recognizing phishing sites. TF-IDF is usually very well known as retrieval algorithm which is been used for classifying documents and comparison [18]. In Image based approach, there is a comparison of phishing and normal websites based on image based on visual similarity [19]. This technique breaks down web pages into block regions depending on 'visual cues'. Metrics such as layout similarity, block level similarity, etc are being used to calculate the visual similarity between phished and non-phished sites [21].

Table 1. Phishing Detection Techniques

S.NO.	APPROACH	TECHNIQUE	METRICS	ACCURACY
1.	Heuristic based approach	Decision tree algorithm	False positive: 5 True positive: 120 False negative: 3 True negative: 72	96.76%
2.	Blacklist approach	Simhash algorithm	False positive: 0	84.36%
3.	Fuzzy rule-based approach	Fuzzy data mining algorithm	Accuracy: 100%	100%
4.	Machine learning approach	Machine learning algorithms	False positive: 1.52% True positive: 98.39%	>98.4%
5.	Cantina based approach	TF-IDF information retrieval algorithm	False positive: 6%	97%
6.	Image based approach	Web logo technique	True positive: 99.8% True negative: 87%	98%

In Table 1, the various approaches of techniques of detecting phishing attacks have been discussed. The Fuzzy rule-based approach has been recorded with high accuracy of phishing attack detection.

IV. PROPOSED SCHEME

Phishing is occurring in various diverse formats in recent times. Due to the novelty in the attack techniques, not only novice users but also well qualified educated people are scammed by such attacks [23]. Business email compromise (BEC) is the main way that cyber attackers target their victims. Many techniques are prevalent for detecting the attacks after they occur. Some have been discussed in this paper. Education of the masses to prevent opening and providing sensitive information to all is to be strictly followed. The threat has now progressed to whaling attack [25] where

the attacker masquerades as a senior member of the organization and targets other members into divulging sensitive or financial information about the organization. The aim is always to steal money or private information. The proposed scheme analyses the existing attacks dataset to identify the behavior of a phishing scam and tries to prevent future attacks by blocking such unsolicited emails/messages. Figure 3 shows the proposed schema which utilizes a combination of both supervised and unsupervised machine learning to identify and prevent both old and new phishing attacks.

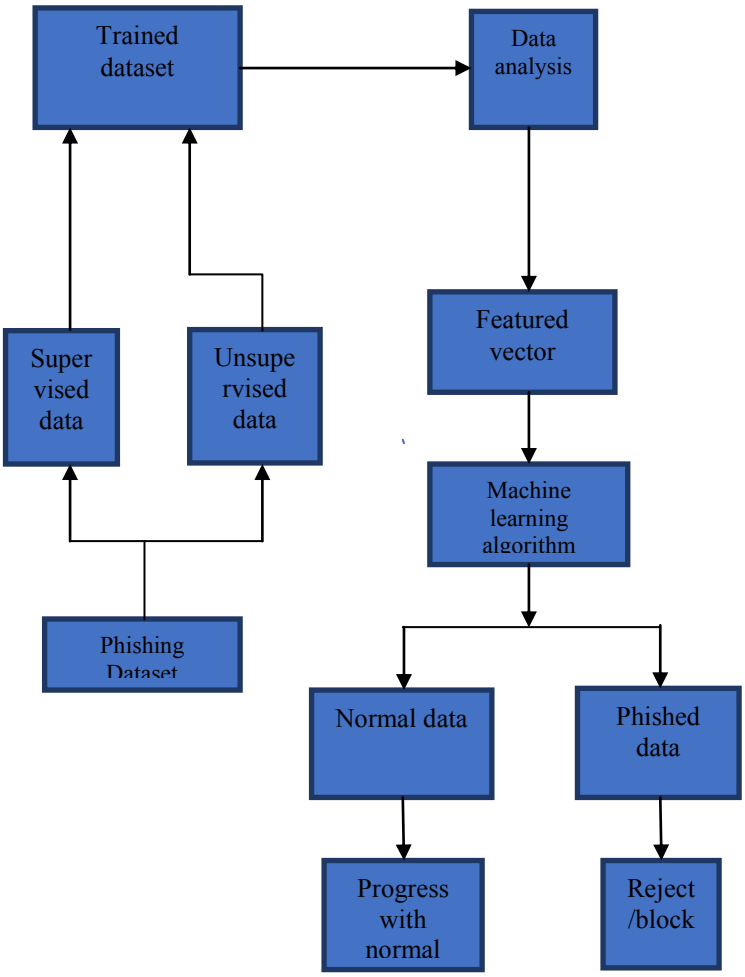

Figure 3. Proposed phishing attack detection and prevention Scheme

The supervised techniques try to identify all the known signatures, behaviors and channels of an attacker. The unsupervised techniques will catch hold of any abnormal behavior or trait which deviates from normal operation of the internet communication.

V. CONCLUSION

Phishing in its different forms of spear phishing and whaling is impacting medium to big organizations by mining all the private and very sensitive data of the organizations. Many ways of implementing the phishing attack have been discussed. The different approaches for identifying the true phishing attacks have also been discussed. The machine learning based approaches provide good identification of true positives. It has become impertinent that anti-phishing techniques be in-built with frequent updations to enhance security of every user. A phishing detection approach is proposed that increases the web security by verifying the links in the source code of the email or webpage.

References

[1] Ms. Neha R. Israni, Mr. Anil N. Jaiswal, "A Survey on various Phishing and Anti Phishing Measures", International Journal of engineering research and technology, Volume.4 (Issue 01), ISSN: 2278-0181, January 2015.
[2] V. Suganya, "A Review on Phishing Attack and Various Anti-Phishing Techniques", International Journal of Computer Applications, Volume.139 (Issue 01), ISSN: 0975-8887, April 2016.
[3] Vikas Sahare, Sheetal Kumar Jain, Manish giri, "Anti-Phishing Framework with Visual Cryptography on Cloud", International Journal of Advance Foundation and Research in Computer, Volume.2, ISSN: 2348-4853, January 2015.
[4] Dipesh Vaya, Sarika Khandelwal, Teena Habpawat, "A Review on Visual Cryptography", International Journal of Computer Applications, Volume.174 (Issue 05), ISSN: 0975-8887, September 2017.
[5] Chaitali Khatri, Supriya Bulkunde, Deepak Uplaonkar, Saurabh saoji, "Phishing detection system using visual cryptography", Multidisciplinary Journal of Research in Engineering and Technology, ISSN: 2348-6953, 2015.
[6] Prof. Gayathri Naidu, "A Survey on Various Phishing Detection and Prevention Techniques", International Journal of Engineering and Computer Science, Volume.5 (Issue 09), Page No. 17823-17826, September 2016.
[7] Mahmoud Khonji, Youssef Iraqi, Andrew Jones, "Literature Review on Phishing Detection", Institute of Electrical and Electronics Engineers Communication Surveys and Tutorials, Volume.15 (Issue 04), 2013.
[8] Ms. Pallavi D. Dudhe, Prof. P. L. Ramteke, "A Review on Phishing Detection approaches", International Journal of Computer Science and Mobile Computing, Volume. 4 (Issue 02), page no. 166-170, February 2015.
[9] Himani thakur, Dr. Supreet Kaur, "A survey paper on phishing detection", International Journal of Advanced Research in Computer Science, Volume.7 (Issue 04), ISSN: 0976-5679, July-August 2016.
[10] Minal chawla, Siddarth Singh Chouhan, "A Survey of phishing attack Techniques", International Journal of Computer Applications, Volume.93 (Issue 03), page no.0975 – 8887, May 2014.
[11] Ryan Russell," Hack Proofing Your Network", 2nd Edition, Syngress Publishers, ISBN: 978-1928994701, 2002.
[12] Dr. Radha Damodaram, "Study of Phishing Attack and Anti-Phishing Tools", International Research Journal of

Engineering and Technology, Volume.3 (Issue 01), ISSN: 2395-0056, January 2016.
[13] Pranal C. Tayade, Prof. Avinash P. Wadhe, "Review Paper on Privacy Preservation through Phishing Email Filter", International Journal of Engineering Trends and Technology, Volume.9 (Issue 12), March 2014.
[14] Kanchan Meena, Tushar Kanti, "A Review of Exposure and Avoidance Techniques for Phishing Attack", International Journal of Computer Applications, Volume 107 (Issue 05), ISSN: 0975 – 8887, December 2014.
[15] Pratik Patil, Prof. P.R. Devale, "A Literature Survey of Phishing Attack Technique", International Journal of Advanced Research in Computer and Communication Engineering, Volume.5 (Issue 04), April 2016.
[16] Anjum N. Shaikh, Antesar M. Shabut, M. Anwar Hossian, "A literature review on phishing crime, prevention review and investigation of gaps", International Conference on Software, 2016.
[17] Andronicus A. Akinyelu and Aderemi O. Adewumi, "Classification of Phishing Email Using Random Forest Machine Learning Technique", Hindawi Publishing Corporation Journal of Applied Mathematics, 2014.
[18] Swapan Purkait, "Phishing counter measures and their effectiveness – literature review", Emerald Group Publishing Limited, Volume.20 (Issue 05), page no. 382-420, 2012.
[19] Juan Chen, Chuanxiong Guo, "Online Detection and Prevention of Phishing Attacks", proceedings First International Conference on Communications and Networking in China, DOI: 10.1109/CHINACOM.2006.344718 2006.
[20] Gori Mohamed.J, M. Mohammed Mohideen, Mrs.Shahira Banu. N, "E-Mail Phishing - An open threat to everyone", International Journal of Scientific and Research Publications, volume.4 (Issue 02), ISSN: 2250-3153, February 2014.
[21] Dr. M. Nazreen Banu, S. Munawara Banu, "A Comprehensive Study of Phishing Attacks", International Journal of Computer Science and Information Technologies, Volume. 4 (Issue 06), page no. 783-786, 2013.
[22] Bhumika P Patel, Ghanshyam I Prajapati, "Phishing Attacks and its Detection", International Journal of Research and Scientific Innovation, Volume. IV (Issue VIS), ISSN: 2321–2705, June 2017.
[23] SpoofGuard, available at: https://crypto.stanford.edu/SpoofGuard/
[24] The biggest phishing attacks of 2018 and what companies can dot prevent them in 2019, available at: https://www.techrepublic.com/article/the-biggest-phishing-attacks-of-2018-and-what-companies-can-do-to-prevent-them-in-2019/
[25] What is whaling attack, Available at: https://me-en.kaspersky.com/resource-center/definitions/what-is-a-whaling-attack