Aslam Khan *et al*, International Journal of Computer Science and Mobile Applications,
Vol.6 Issue. 4, April- 2018, pg. 33-37 **ISSN: 2321-8363**
Impact Factor: 5.515

A Survey Paper on Detection of Phishing Website by URL Technique

Aslam Khan[1], Rahul Sharma[2]

[1]R.K.D.F School of Engineering, Indore (M.P) India, aslamkhanashu@gmail.com
[2]R.K.D.F School of Engineering, Indore (M.P) India, sharma.rahul5656@gmail.com

Abstract: While the World Wide Web has become a killer application on the Internet, it has also brought in an immense risk of cyber-attacks. Adversaries have used the Web as a vehicle to deliver malicious attacks such as phishing, spamming, and malware infection. For example, phishing typically involves sending an email seemingly from a trustworthy source to trick people to click a URL (Uniform Resource Locator) contained in the email that links to a counterfeit webpage. Therefore, phishing is form of identity theft that combines social engineering techniques and sophisticated attack vectors to harvest financial information from unsuspecting consumers. Often a phisher tries to lure her victim into clicking a URL pointing to a rogue page. In this paper, we gives an overview of the state of URL phishing techniques and describes the various terminology which is extensively relative to survey the phishing specification according need of current generation and concerning of end user applications.
Keywords: Phishing, Data Mining, Security, Classification, Information Filtering, Social Network, URL phishing.

1. INTRODUCTION

The Web has become a platform for supporting a wide range of criminal enterprises such as spam-advertised commerce, financial fraud and as a vector for propagating malware. The precise commercial motivations behind these schemes may differ but the common thread among them is the requirement that unsuspecting users visit their sites. These visits can be driven by email, web search results or links from other Web pages, but all require the user to take some action, such as clicking, that specifies the desired Uniform Resource Locator (URL) and obtains sensitive information [1].

In order to overcome this problem, the security community has responded by developing blacklisting services encapsulated in toolbars, appliances and search engines that provide an alert or warning precisely as feedback. Many malicious sites are not blacklisted either because they are too new, or never evaluated, or not evaluated incorrectly. In order to address this problem, some client-side systems analyze the content or behavior of a Web site as it is visited which causes runtime overhead due to browser based vulnerabilities [2] [3].

The Web has become a platform for supporting a wide range of criminal enterprises such as spam-advertised commerce financial fraud and as a vector for propagating malware. Although the precise commercial motivations behind these schemes may differ, the common thread among them is the requirement that unsuspecting users visit their sites. These visits can be driven by email, Web search results or links from other Web pages, but all require the user to take some action, such as clicking, that specifies the desired Uniform Resource Locator (URL) [4].

The advent of new communication technologies has had tremendous impact in the growth and promotion of businesses spanning across many applications including online-banking, e-commerce, and social networking. In fact, in today's age it is almost mandatory to have an online presence to run a successful venture. As a result, the importance of the World Wide Web has continuously been increasing. Unfortunately, the technological advancements come coupled with new sophisticated techniques to attack and scam users. This paper is a review of URL phishing and their techniques and describes some of the commonly used techniques to solve complex problems of security attacks and vulnerability.

2. BACKGROUND

The background of a study is an important part of our survey paper. It provides the context and purpose of the study. Hence there is need for background study that contribute to prepare different aspects of the URL Phishing detection.

2.1 What is phishing?

Phishing is the process of committing an internet fraud. In this process, an adversary sends a mail to a user posing as a financial institution. The user is tricked by the adversary, as the website put up by him looks exactly like the original web site. It can contain logos and images also. There are three ways in which phishing attacks are performed namely Impersonation, Forwarding and Pop-up [5].

Phishing is a deception technique that utilizes a combination of social engineering and technology to gather sensitive and personal information, such as passwords and credit card details by masquerading as a trustworthy person or business in an electronic communication. Phishing makes use of spoofed emails that are made to look authentic and purported to be coming from legitimate sources like financial institutions, ecommerce sites etc., to lure users to visit fraudulent websites through links provided in the phishing email. The fraudulent websites are designed to mimic the look of a real company webpage [6].

In general, phishing attacks are performed with the following four steps [7]:

- A fake web site which looks exactly like the legitimate Web site is set up by phisher
- Phisher then send link to the fake web site in large amount of spoofed e-mails to target users in the name of legitimate companies and organizations, trying to convince the potential victims to visit their web sites.
- Victims visit the fake web site by clicking on the link and input its useful information there.
- Phishers then steal the personal information and perform their fraud such as transferring money from the victims' account.

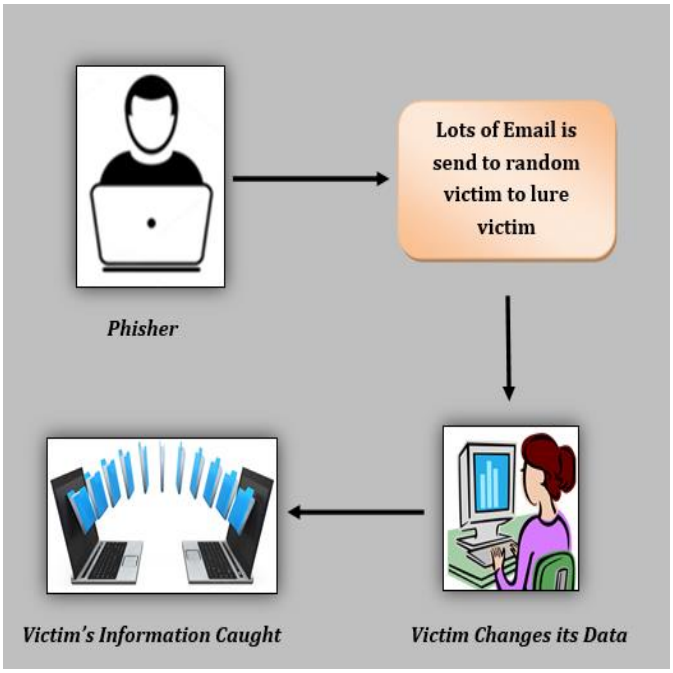

Figure 1: Process of Phishing

2.2 URL Phishing

Detection of phishing URLs has become increasingly difficult due to the evolution of phishing campaigns and their efforts to avoid mitigation by blacklists. The current state of cybercrime has made it possible for phishers to host campaigns with shorter lifecycles, diminishing blacklist effectiveness [8]. The act of acquiring sensitive information by convincing the users to reveal their personal information such as usernames, passwords, credit card credentials, etc. by pretending as a trusty source in an electronic transmission is known as phishing. It is a criminal offense which targets both social engineering and technical tricks to steal personal identity or financial account information of user and is an automated form of identity theft. Phishing websites are affecting both individuals and financial organizations on the Internet, leading to a serious threat to electronic commerce.

Every URL has this common syntax-

< protocol >://< hostname >< path >

A phishing URL is created with a malicious purpose to download malware, to perform phishing attacks or to manipulate search engine's results. The technical experience of criminals is increasing to build more survivable infrastructures that support phishing activities.

2.3 Phishing URL Types

In a phishing attack, an adversary typically lures the victim into clicking a URL pointing to the phishing site. The adversary usually obfuscates this URL through varied methods. To determine the popular obfuscation techniques currently in use, we examined a black list of phishing URLs maintained by Google. This black list is used to provide phishing protection in Firefox [9] [10]. The prominent obfuscation techniques are:

Type I: Obfuscating the Host with an IP address. In this form of attack the URL's hostname is replaced with an IP address, and usually the organization being phished is placed in the path. Very often the IP address is also represented in hex or decimal rather than the dotted quad form.

Type II: Obfuscating the Host with another Domain. In this form of attack the URL's host contains a valid looking domain name, and the path contains the organization being phished. This form of attack usually tries to imitate URLs containing a redirect so as to make it appear valid.

Type III: Obfuscating with large host names. This form of attack has the organization being phished in the host but appends a large string of words and domains after the host name.

Type IV: Domain unknown or misspelled. Here there is no apparent relationship to the organization being phished or the domain name is misspelled.

2.4 Phishing Example

Phishing websites use a number of different techniques to hide the fact that they are not authentic including overwriting or disguising the true URL shown in the browser, overlaying the genuine web site with a crafted pop-up window, drawing fake padlock images on top of the browser window to give the impression that SSL is enabled, and registering SSL certificates for domain names similar to the real organization etc. In practice, these tricks make it extremely difficult for the average user to distinguish a phishing site from a genuine one [11].

Figure 2: Phishing Websites

Above example show the example of phishing website. The rapid development of online financial services and e-commerce, phishing website attacks have become one of the most dangerous and prevalent threats on Internet, causing inestimable damage. More and more phishing web pages have been found in recent years in an accelerative way. To avoid phishing websites, both online financial organizations and their consumers have to understand phishing and anti-phishing technologies and take security actions.

3. LITERATURE SURVEY

In spite of lot of work that has been done on implementing better and efficient tools on phishing detection and prevention. Various approaches have been proposed for preventing the website or link from phishing attack. In this section, describes related study of earlier work done for identifying phishing URLs.

Hassan Y. A. Abutair et al. [12] introduce a Case-Based Reasoning (CBR) Phishing Detection System (CBR-PDS). It mainly depends on CBR methodology as a core part. The system is highly adaptive and dynamic as it can easily adapt to detect new phishing attacks with a relatively small data set in contrast to other classifiers that need to be heavily trained in advance. Authors test their system using different scenarios on a balanced 572 phishing and legitimate URLs. Experiments show that the CBR-PDS system accuracy exceeds 95.62%, yet it significantly enhances the classification accuracy with a small set of features and limited data sets.

R. Gowtham et al. [13] presented a novel approach that not only overcomes many of the difficulties in detecting phishing websites but also identifies the phishing target that is being mimicked. They have proposed an anti-phishing technique that groups the domains from hyperlinks having direct or indirect association with the given suspicious webpage. The domains gathered from the directly associated web pages are compared with domains gathered from the indirectly associated web pages to arrive at a target domain set. On applying Target Identification (TID) algorithm on this set, we zero-in the target domain. Authors then perform third-party DNS look up of the suspicious domain and the target domain and on comparison identify the legitimacy of the suspicious page.

Routhu Srinivasa Rao et al [14] implemented a desktop application called PhishShield, which concentrates on URL and Website Content of phishing page. PhishShield takes URL as input and outputs the status of URL as phishing or legitimate website. The heuristics used to detect phishing are footer links with null value, zero links in body of html, copyright content, title content and website identity. PhishShield is able to detect zero hour phishing attacks which blacklists unable to detect and it is faster than visual based assessment techniques that are used in detecting phishing. The accuracy rate obtained for PhishShield is 96.57% and covers a wide range of phishing web sites resulting less false negative and false positive rate.

Monire Norouzi et al. [15] presented a data mining classification approach to detect malware behavior. They proposed different classification methods in order to detect malware based on the feature and behavior of each malware. A dynamic analysis method has been presented for identifying the malware features. A suggested program has been presented for converting a malware behavior executive history XML file to a suitable WEKA tool input. To illustrate the performance efficiency as well as training data and test, author apply the approaches to a real case study data set using WEKA tool. The evaluation results demonstrated the availability of the proposed data mining approach. Also our proposed data mining approach is more efficient for detecting malware and behavioral classification of malware can be useful to detect malware in a behavioral antivirus.

Yujie Fan et al. [16] proposed an effective sequence mining algorithm to discover malicious sequential patterns, which is based on the instruction sequences extracted from the file sample set, propose and then All-Nearest-Neighbor (ANN) classifier is constructed for malware detection based on the discovered patterns. The developed data mining framework composed of the proposed sequential pattern mining method and ANN classifier can well characterize the malicious patterns from the collected file sample set to effectively detect newly unseen malware samples. A comprehensive experimental study on a real data collection is performed to evaluate our detection framework. Promising experimental results show that framework outperforms other alternate data mining based detection methods in identifying new malicious executables.

Aaron Blum et al. [17] explores the possibility of utilizing confidence weighted classification combined with content based phishing URL detection to produce a dynamic and extensible system for detection of present and emerging types of phishing domains. The system is capable of detecting emerging threats as they appear and subsequently can provide increased protection against zero hour threats unlike traditional blacklisting techniques which function reactively.

4. CONCLUSION

Phishing techniques have not only grown in number, but also in sophistication. Phishing recognition techniques are rapidly varying to keep up with the novel techniques used by phishers. The detection of Phishing Websites system offers security for the user that highly helps to achieve better transaction needed for the users, in their own interest. The phishing detection system reduces mainly the detection time and

informs the users, whether it is a phishing website or legitimate website. This paper deal with different terminology of URL phishing paradigm. Additionally, in this paper we described various phishing techniques and example by which people can understood for the email threats.

REFERENCES

[1] "Sahoo, Doyen, Chenghao Liu, and Steven CH Hoi."Malicious URL detection using machine learning: A survey." arXiv preprint arXiv: 1701.07179 (2017)."
[2] Dhanalakshmi, R. "Studying And Improving Techniques to Mitigate EMail Threats." (2014).
[3] Gayathri Naidu, "A Survey on Various Phishing Detection and Prevention Techniques", Volume 5, Issue 09, September 2016, pp. 17823-17826.
[4] Ma, Justin, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. "Beyond blacklists: learning to detect malicious web sites from suspicious URLs." In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1245-1254. ACM, 2009.
[5] Laxman Muthiyah, "What is Phishing? How to Create Phishing Page, Facebook Example", available online at: https://www.7xter.com/2016/08/phishing.html. Last Modified: March 20, 2017.
[6] "Chapter 4: Phishing URL Detection", available online at: http://shodhganga.inflibnet.ac.in/bitstream/10603/195819/12/12_chapter%204.pdf
[7] Hicham Tout, William Hafner "Phishpin: An identity-based anti-phishing approach" in proceedings of international conference on computational science and engineering, Vancouver, BC, pages 347-352, 2009
[8] Gundel, T. (2005) Phishing and Internet Banking Security, Technical Security report, IBM Crypto Competence Center.
[9] V. Suganya, "A Review on Phishing Attacks and Various Anti Phishing Techniques", International Journal of Computer Applications (IJCA) Volume 139 – No.1, April 2016.
[10] Choi, Hyunsang, Bin B. Zhu, and Heejo Lee. "Detecting Malicious Web Links and Identifying Their Attack Types." WebApps 11 (2011): 11-11.
[11] "Phishing Examples: What to Watch For", available online at: https://www.safecomputing.umich.edu/be-aware/phishing-and-suspicious-email/examples
[12] Hassan Y. A. Abutair, Abdelfettah Belghith, "Using Case-Based Reasoning for Phishing Detection", 8th International Conference on Ambient Systems, Networks and Technologies, Procedia Computer Science 109C (2017), pp. 281–288
[13] R.Gowtham, Dr.Ilango Krishnamurthi, K.Sampath Sree Kumar, "An efficacious method for detecting phishing webpage through Target Domain Identification", Decision Support Systems November 30, 2013
[14] Routhu Srinivasa Rao and Syed Taqi Ali, "PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) pp.147 – 156
[15] Norouzi, Monire, Alireza Souri, and Majid Samad Zamini. "A data mining classification approach for behavioral malware detection." Journal of Computer Networks and Communications 2016 (2016): 1.
[16] Fan, Yujie, Yanfang Ye, and Lifei Chen. "Malicious sequential pattern mining for automatic malware detection." Expert Systems with Applications 52 (2016): 16-25.
[17] Blum, Aaron, Brad Wardman, Thamar Solorio, and Gary Warner. "Lexical feature based phishing URL detection using online learning." In Proceedings of the 3rd ACM workshop on Artificial intelligence and security, pp. 54-60. ACM, 2010.

