Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Review on Phishing Sites Detection Techniques

Oza Pranali P
M.E in Computer Engineering (Cyber Security)
Gujarat Technological University
Ahmedabad, Gujarat

Deepak Upadhyay
Assistant Professor
Gujarat Technological University
Ahmedabad, Gujarat

***Abstract* - Due to the rapid growth of the Internet, user interact with social network such as Facebook, Twitter, Linkedin and many more for communicate with each other. By using the unusual structure of the Internet, attacker set out new techniques, such as phishing, to lure the user to interact with the fake websites through social networks that appears similar to legitimate ones. The main motive behind this attack is to steal the sensitive information such as password, username, credit card details and many more details from the users. There are various platforms where phishing attack can occur like online payment sector, webmail, and financial institution, file hosting or cloud storage and many more. In this paper, presenting a comprehensive survey of phishing detection using different features of website and machine learning approach.**

Keywords - Social Network Security, Phishing attack, Detection Techniques

I. INTRODUCTION

Phishing is that the fraudulent plan to to obtain sensitive information like username, password, and creditcard details, often malicious purposes, by disguising as a trustworthy entity in an electronic communication. 'Phishing' recorded on 2nd January, 1996 according to Internet records.[6] Social media phishing is when attackers use social networking sites like Facebook, Twitter, and instagram rather than email to obatain your sensitive personal information or click.[4]

In this attack phishers use fake websites and emails to expose a user's sensitive private information. They plan to create a uniform false copy of an ingenious website. The rapid growth of Information Technology indeed created many connivances to us, but on the other hand it also resulted and increased security challenges to us to protect our information securely especially from social engineering attack now a days. According to Anti Phishing Working Group, The number of phishing attacks rise in the third quarter of 2019, to high level not seen since late 2016.Phishing is the major security threats faced by the cyber-world and could lead to financial losses for both industries and individuals. In this attack, Phisher makes a fake web page by copying contents of the legitimate page, so that a user cannot differentiate between phishing and legitimate sites. Social engineering schemes prey on unwary victims by fooling them into believing they are dealing with a trusted, legitimate party, such as by using deceptive email addresses and email messages.

In Q3 2019, 68 percent of web sites used for phishing were using SSL. But by the end of 2019, 74% of all phishing sites were using TLS/SSL. Attackers are using free certificates on phishing sites that they create, and are abusing the encryption already installed on hacked web sites.[2]

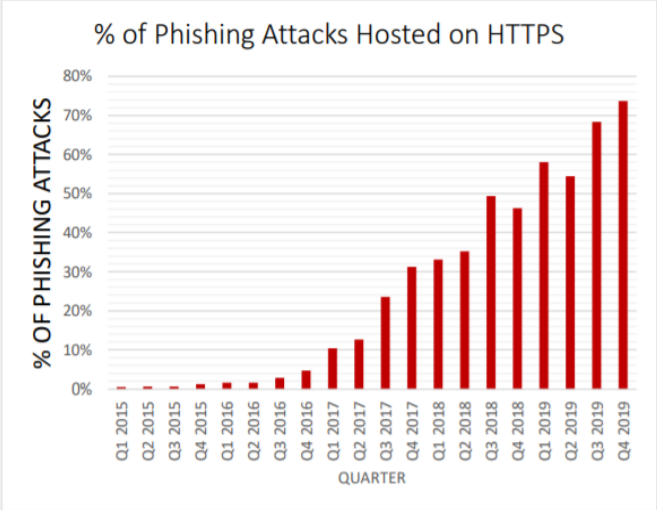

Figure 1 : % of Phishing Attacks Hosted on HTTPS [2]

30th October 2018 the Cybercrime cell finding a major inter-state racket of conning job expectant to getting money from their bank accounts through a phishing website such as cariourjobs.com, hindustanjobsonline.com, naukriira.com, jobsongoes.com, futurecareersolutions.com, indeed4jobs.com, and Hindustan-jobs.com.[5]

Types of Phishing Scams : [1]

Deceptive Phishing: Deceptive phishing refers to any attack by fraudsters impersonate a legitimate company and plan to steal people's personal information or login credentials.

Spear Phishing: It is method of sending a Phishing messages to a particular organization to gain organizational information for more targeted social engineering. For example, Social media sites like LinkedIn.

CEO Fraud: Phishers use an email address similar to that of an authority to request payments or data from others within in the company. For example, Use CEO Id.

Pharming: In this attack phisher hijack a website's domain name and use it to redirect visitors to a fake site. Pharmer targets a DNS server and changes the IP address.

Dropbox Phishing: Many people use Dropbox every day to backup, access and share their files. The attackers would

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

attempt to maximize the platforms popularity by targeting users with phishing email.

Google Docs Phishing: Phishers could choose to target Google Drive similar to the way they might prey upon Dropbox users. Specially, as Google Drive supports documents, spreadsheets, presentations, photos and even entire website, phishers can misuse the service to create a web page that mimics the Google account log-in screen.

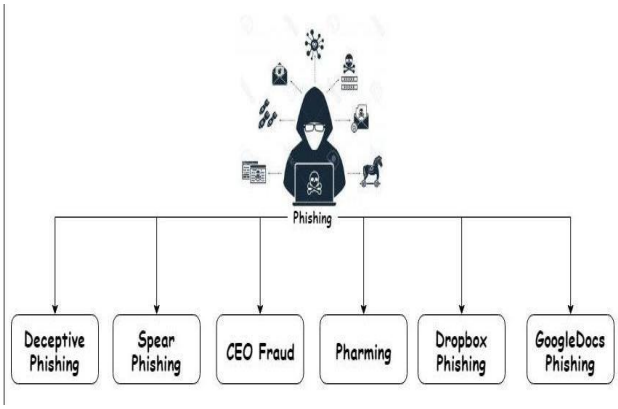

Figure 2 : Types of Phishing Scam [1]

Why Phishing Detection Required?

It has been approximately 23 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. [13]

II. HOW PHISHING WORKS

Planning: Phisher decide which business to target and determine how to get e-mail address for the customers of that business. They often use an equivalent mass-mailing and address collection techniques as spammers.

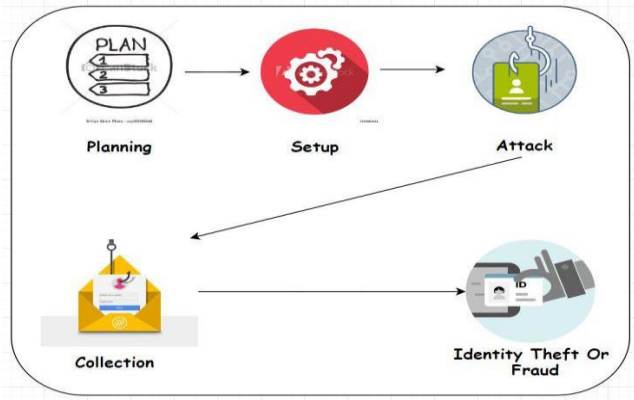

Figure 3 : How Phishing Works [3]

Setup: Once they know which business to spoof and who their victims are, phishers create methods for delivering the message and collecting the data. Most often, this involves e-mail addresses and Web page.

Attack: This is often the step people are most familiar with- the phisher sends a phony message that appears to be from a reputable source.

Collection: Phishers record the information victims enter into Web page or popup windows.

Identity Theft and Fraud: The phishers use the information they've gathered to make illegal purchases or otherwise commit fraud.[3]

III. LITERATURE SURVEY

Rathore, S., Sharma, P.K., Loia, V., Jeong, Y.S. and Park, J.H. [11] presents a comprehensive survey of different security and privacy threats that target every user of social networking sites. A Social Network Service (SNS) is a type of web service for establishing a virtual connection between people with similar interests, backgrounds, and activities. In recent years, SNSs become a well-liked medium of communication. The number of SNS users worldwide is continuously increasing every year. This paper separately focuses on various threats that arise due to the sharing of multimedia content within a social networking site. In this, describing three classes of threats – Multimedia Content Threats, Traditional Threats and Social Threats [11].

Pujara, P. and Chaudhari, M.B., [10] Phishing frauds might be the most popular cybercrime used today. This paper detailed literature survey and proposed new approach to detect phishing website by features extraction and machine learning algorithm. In this paper author describe different methodologies such as Blacklist method, Heuristic based method, Visual similarity and Machine learning for phishing detection. Blacklist method is used in which list of phishing URL is stored in database and then if URL is found in database, it is known as phishing URL and gives warning otherwise it is called legitimate. Heuristic based method is extension of blacklist and able to detect new attack as use features extracted from phishing site to detect phishing attack. Visual similarity approach deceive user by extracting image of legitimate site. Machine Learning approach works efficiently in large dataset [10].

Jain, A.K. and Gupta, B.B., [8] Attackers steal sensitive information like personal identification number (PIN), credit card details, login, password, etc., from internet users. In this paper, author proposed a machine learning based anti-phishing system based on Uniform Resource Locator (URL) features. To evaluate the performance of proposed system, author taken 14 features from URL to detect a website as a phishing or non-phishing. The proposed system is trained using quite 33,000 phishing and legitimate URLs with SVM and Naïve Bayes classifiers. Experiment results show quite 90% accuracy in detecting phishing websites using SVM classifier [8].

Sahingoz, O.K., Buber, E., Demir, O. and Diri, B., [12] tries to detect phishing site using url for preventing User's sensitive information. Computer users fall for phishing due to the five main reasons:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

- Users don't have detailed knowledge about URLs,
- Users don't know, which web pages can be trusted,
- Users don't see the whole address of the web page, due to the redirection or hidden URLs,
- Users don't have much time for consulting the URL, or accidentally enter some web pages,
- Users cannot distinguish phishing web pages from the legitimate ones.

In proposed system, author used NLP based features and Word features for classification of phishing and non-phishing sites. For classification used Decision Tree, Adaboost, K-star, kNN(n=3), Random Forest, SMO(Squential Minimal Optimization) and Naïve Bayes [12].

Machado, L. and Gadge, J. [9] Phishing sites are the fake websites created by phishers with intent of stealing user's personal information to carry out fraudulent activities. This paper proposes an efficient way for detection of the phishing website using C4.5 decision tree approach. The method proposed in this paper uses various URL features and also uses C4.5 decision tree approach for better results [9].

Jain, A.K. and Gupta, B.B., [7] presents a novel approach that can detect phishing attack by analyzing the hyperlinks found in the HTML source code of the website. A phishing attack is performed by taking advantage of the visual resemblance between the fake and the authentic web-pages. The proposed approach has divided the hyperlink specific features into 12 different categories and used these features to train the machine learning algorithms. Author evaluated the performance of proposed phishing detection approach on various classification algorithms using phishing and non-phishing website dataset [7].

IV. CONCLUSION

Phishing is a way to obtain user's private information via email or website.As technology increases, phishing attackers using new methods day by day. In this paper describe detailed literature survey about phishing website detection. Phishing websites are short-lived, and thousands of fake websites are generated every day. Therefore, there is requirement of real-time, fast and intelligent phishing detection solution. According to this, Machine learning is efficient technique to detect phishing. More features can be added to improve the accuracy of the proposed phishing detection system.

REFERENCES

[1] Godbole, N. and Belapure, S., 2011. Cyber Security, Understanding Computer Forensics and Legal Perspectives.
[2] https://apwg.org/trendsreports/.
[3] https://computer.howstuffworks.com/phishing.htm.
[4] https://inspiredelearning.com/blog/social-phishing/
[5] https://timesofindia.indiatimes.com/city/vadodra/multi-state-job-racket-busted-by-cybercrime-cell/articleshow/66421586.cms
[6] https://www.webopedia.com/DidYouKnow/Internet/phishing.asp.
[7] Jain, A.K. and Gupta, B.B., 2019. A machine learning based approach for phishing detection using hyperlinks information. *Journal of Ambient Intelligence and Humanized Computing*, *10*(5), pp.2015-2028.
[8] Jain, A.K. and Gupta, B.B., 2018. PHISH-SAFE: URL features-based phishing detection system using machine learning. In *Cyber Security* (pp. 467-474). Springer, Singapore.
[9] Machado, L. and Gadge, J., 2017, August. Phishing Sites Detection Based on C4. 5 Decision Tree Algorithm. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1-5). IEEE.
[10] Pujara, P. and Chaudhari, M.B., 2018. Phishing Website Detection using Machine Learning: A Review.
[11] Rathore, S., Sharma, P.K., Loia, V., Jeong, Y.S. and Park, J.H., 2017. Social network security: Issues, challenges, threats, and solutions. *Information sciences*, *421*, pp.43-69.
[12] Sahingoz, O.K., Buber, E., Demir, O. and Diri, B., 2019. Machine learning based phishing detection from URLs. *Expert Systems with Applications*, *117*, pp.345-357.
[13] Thakur, H. and Kaur, S., 2016. A survey paper on phishing detection. *International Journal of Advanced Research in Computer Science*, *7*(4).