International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Literature Survey on Phishing Attack

Jayashree C. Pasalkar[1], Ankita S. Giri[2]

[1]Assistant Professor, Department of Information Technology, AISSMS's Institute of Information Technology, Pune-411001, INDIA
[2] TE. BE (Information Technology), AISSMS's Institute of Information Technology, Pune-411001, INDIA

A B S T R A C T

The paper is to detect the accuracy of the fake emails, it provides maximum accuracy and helps to determine the fake emails. Phishing is a deception technique that utilizes a combination of social engineering and technology.

A phishing attack is a method of tricking users into unknowingly providing personal and financial information or sending funds to attackers. The most common phishing attacks use some form of electronic messaging such as email to provide a link to what appears to be a legitimate site but is actually a malicious site controlled by the attacker. Phishing is a hybrid attack combining both social engineering and technological aspects and combatting phishing attacks requires dealing with both aspects.

In this study, a software called "Anti Phishing Simulator" was developed, giving information about the detection problem of phishing and how to detect phishing emails. With this software, phishing and spam mails are detected by examining mail contents. Classification of spam words added to the database by Bayesian algorithm is provided.

1. Introduction to Phishing Attack

1.1 Introduction

Phishing email messages, websites, and phone calls are designed to steal money or sensitive information. Phishing is a deception technique that utilizes a combination of social engineering and technology. Cybercriminals can do this by installing malicious software on your computer, tricking you into giving them sensitive information, or outright stealing personal information off of your computer. Nowadays, most transactions are made online. [2]Everything is done online, even sending money and paying bills. As soon as phishers have these stolen credentials, they can use these specifics to make a fake account of the victim, seriously jeopardizing their credentials or possibly preventing users from accessing their accounts. The Anti-Phishing Working Group (APWG Q2 2019) report [1] states that 182,465 phishing sites have been identified so far in 2019. Webmail and Software-as-a-Service are the industries that are most specifically targeted in this (SaaS). According to the Phish lab Phishing Report 2019. 84% of phishing attempts target financial services, shipping, cloud storage services, and payment services. The payment industry is phishing's most alluring target. Phishing kits mostly target businesses involved in banking, finance, retail, and consumer products, including Microsoft, PayPal, Amazon, Apple, and others. Many Internet users overlook the web browsers' protection indicator and improperly check the website's URL (Uniform Resource Locator). [1].

1.2 Motivation

i. Basically Phishing is a sub-topic of cyber-security.

ii. Cyber-security is forever trending topic.

iii. It's a Fascinating topic.

iv. It teaches us how to protect computer operating systems, networks, and data from cyber-attacks.

v. An IT Student should have the basic knowledge about the cyber-security.

1.3 Aim and Objective of the work

i. The aim of this project is to define phishing and identify various types of phishing scams.

ii. Recognize common baiting tactics used in phishing scams.

iii. Understand how to protect yourself from being hooked by a phishing scam.

Project objectives:

i. Detecting fake news on social media is important and also a technically challenging problem these days.

ii. We knew that Machine learning is helpful for building Artificial intelligence systems based on tacit knowledge because it can help us to solve complex problems due to real word data.

iii. On the other side we knew that Knowledge engineering is helpful for representing expert's knowledge which people aware of that knowledge project.

iv. This must be highly focused and feasible and should address the more immediate project outcomes

v. Guide must personnel check these aim and objectives and make students write these statements properly

2. Scope of ML in Phishing Attack

Applications of machine learning and statistical techniques have increased across a variety of fields in recent years as a result of the explosion in computing power and available data. The scientific process, from data analysis to modeling, has benefited from the application of these techniques in astronomy and space sciences.

2.1 Specifications

Types of Phishing

A. Social Engineering
   i. SMS phishing
   ii. Phishing
   iii. Deceptive phishing
   iv. Link manipulations
   v. In-session phishing

B. Technical subterfuge
   i. Malware based phishing
   ii. Key loggers and Screen loggers
   iii. Session Hijacking
   iv. Pharming

2.2 Features

For Detection of Phishing

i. HTML Email
ii. IP-based URL
iii. Age of Domain Name
iv. Number of Domains
v. Number of Sub-domains
vi. Presence of Form Tag
vii. Number of Links
viii. URL Based Image Source
ix. Matching Domains (From & Body)

x. Keywords

2.3 Illustration

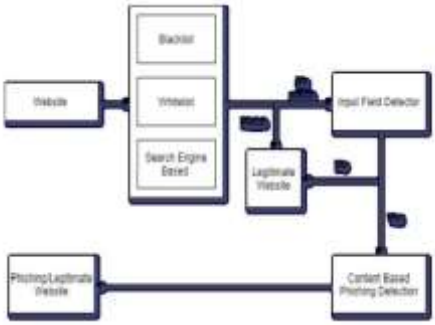

Figure 1 Phishing Detection system architecture

Figure 2 Processing cycle of phishing attack

2.4 Methodology

This architecture mainly aims to detect phishing web page on the client-side web browser. The main components of the architecture are shown in Figure 1. When a website is entered, it moves to a phase where three techniques of phishing detection are employed. As applying every technique in a single phase is time-consuming, will pick one technique at a given time based on condition. URL entered will be checked by the Blacklist method or whitelist method or search engine based technique. Figure 1. The blacklist consists of phishing URLs and their associated details. If entered URL is there in the blacklist, then that website is stated as phishing otherwise, the status of the web page is unknown marked as no match and moves to the next phase. As blacklist does not contain newly registered phishing websites. The whitelist consists of legitimate URLs and their associated details. If entered URL is present in the whitelist, the match occurs and that URL is declared legitimate. Popular and trusted websites are included in the whitelist. If the URL is not on the list, it is suspicious and it moves to the next level. In search engine based technique, the domain name is stripped from the URL. The title is retrieved from the website title tag. The explanation behind the domain name and title extraction is that these reflect the website's identity. The domain name on the phishing website is different from the claimed brand whereas the legitimate website uses the same domain name as the claimed brand. The approaches make the search query powerful. In top search results, an efficient query returns related domains. Title and domain name are extracted and fed in to search engine, then the domain name of the suspicious website is compared with corresponding domains returned from the results. If a match occurs, it is declared a legitimate website.

2.5 Enhancement

i. This architecture is efficient and effective but slow.

ii. It is somewhat difficult to understand.

iii. It is difficult to operate for a non-technical person.

iv. As this technology has less accuracy, by embedding ML algorithms it can be improved.

v. Hence it should be made easy to operate and understand.

3. Conclusion

One of the common hazards that is difficult to avoid is phishing. It is necessary to establish several forms of authentication for email networks. The phishing attempt is successful only when the victim clicks on the chosen link. The best way to prevent phishing attacks is to educate users about the different kinds of assaults that might occur within a network [2]. To prevent any type of data security vulnerabilities, choose the best security software solutions or programs, such as an anti-phishing browser extension. Another strategy to greatly reduce phishing is to update anti-phishing technologies. By examining the website's content, the system architecture presented here helps to lower the false positive rate [1]. This technique is effective at quickly identifying reliable websites. Legitimate website is filtered out in each phase without further moving to other phases.

References

[1]. Ram Basnet, Srinivas Mukkamala, and Andrew H. Sung,"Detection of phishing attack using ML", Springer, ISSN:978-3-540-77465-5_19, February 2008.

[2]. Athulya A A, Praveen K,"Towards the Detection of Phishing Attack",IEEE, ISSN:978-1-7281-5518-0, 27 July 2020

[3]. Muhammet Baykara, Zahit Ziya Gurel,"Detection of Phishing Attack",IEEE, ISSN:978-1-5386-3449-3, 27 July 2020

[4]. 2019 PHISHING TRENDS AND INTELLIGENCE REPORT, https://info.phishlabs.com/hubfs/2019%20PTI%20Report/2019%20Phishing%20Trends%20and%20Intelligence%20Report.pdf.

[5]. Peng, Peng, et al, "What happens after you leak your password: Understanding credential sharing on phishing sites." Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. 2019.

[6]. Mrdovic, Sasa, and Branislava Drazenovic, "KIDS–Keyed Intrusion Detection System." International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer, Berlin, Heidelberg, 2010.

[7]. SMS phishing, https://en.wikipedia.org/wiki/SMS_phishing