WEB PHISHING DETECTION : A LITERATURE SURVEY

ABSTRACT:

This article surveys the literature on the detection of web phishing attacks. Website Phishing costs internet users billions of dollars per year. Phishers steal personal information and financial account details such as usernames and passwords, leaving users vulnerable in the online space. Phishing attacks, which have existed for several decades and continue to be a major problem today, constitute a severe threat in the cyber world. Attackers are adopting multiple new and creative methods through which to conduct phishing attacks, which are growing rapidly. Therefore, there is a need to conduct a comprehensive review of past and current phishing approaches.

INTRODUCTION:

Phishing is one of the social engineering: techniques that scam artists use to manipulate human psychology. Social engineering techniques include forgery, misdirection and lying—all of which can play a part in phishing attacks. On a basic level, phishing emails use social engineering to encourage users to act without thinking things through.

Like many common threats, the history of phishing starts in the 1990s. When AOL was a popular content system with internet access, attackers used phishing and instant messaging to masquerade as AOL employees to trick users into divulging their credentials to hijack accounts.

In the 2000s, attackers turned to bank accounts. Phishing emails were used to trick users into divulging their bank account credentials. The emails contained a link to a malicious site that mirrored the official banking site, but the domain was a slight variation of the official domain name (e.g., paypai.com instead of paypal.com). Later, attackers pursued other accounts such as eBay and Google to hijack credentials, steal money, commit fraud or spam other users.

In 2018, the FBI received around 100 complaints, with the most commonly targeted industries being healthcare, education, and air travel, which resulted in a combined net loss of approximately USD 100 million dollars. This scam involved the use of phishing emails to target employees and discover their login credentials. These were then used to gain access to the payroll system, after which rules were implemented by the phishers so that employees no longer received notifications about changes made to their accounts. The phisher was then able to change account holders' direct debit information to funnel the funds into their own account, which in this instance involved a prepaid card.

Types

Phishing has evolved into more than simple credential and data theft. How an attacker lays out a campaign depends on the type of phishing. Types of phishing include:

Email phishing: The general term given to any malicious email message meant to trick users into divulging private information. Attackers generally aim to steal account credentials, personally identifiable

information (PII) and corporate trade secrets. However, attackers targeting a specific business might have other motives.

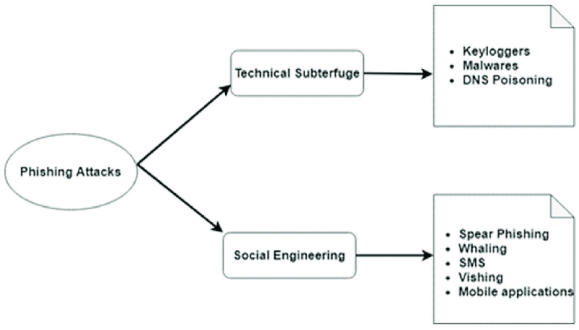

Fig:1 Multiple forms of phishing attacks

Spear phishing: these email messages are sent to specific people within an organization, usually high-privilege account holders, to trick them into divulging sensitive data, sending the attacker money or downloading malware.

Link manipulation: messages contain a link to a malicious site that looks like the official business but takes recipients to an attacker-controlled server where they are persuaded to authenticate into a spoofed login page that sends credentials to an attacker.

Whaling (CEO fraud): these messages are typically sent to high-profile employees of a company to trick them into believing the CEO or other executive has requested to transfer money. CEO fraud falls under the umbrella of phishing, but instead of an attacker spoofing a popular website, they spoof the CEO of the targeted corporation.

Content injection: an attacker who can inject malicious content into an official site will trick users into accessing the site to show them a malicious popup or redirect them to a phishing website.

Malware: users tricked into clicking a link or opening an attachment might download malware onto their devices. Ransomware, rootkits or keyloggers are common malware attachments that steal data and extort payments from targeted victims.

Smishing: using SMS messages, attackers trick users into accessing malicious sites from their smartphones. Attackers send a text message to a targeted victim with a malicious link that promises discounts, rewards or free prizes.

Vishing: attackers use voice-changing software to leave a message telling targeted victims that they must call a number where they can be scammed. Voice changers are also used when speaking with targeted victims to disguise an attacker's accent or gender so that they can pretend to be a fraudulent person.

"Evil Twin" Wi-Fi: spoofing free Wi-Fi, attackers trick users into connecting to a malicious hotspot to perform man-in-the-middle exploits.

Definition:

The definition of phishing attacks is not consistent in the literature, which is due to the fact that the phishing problem is broad and incorporates varying scenarios. For example, according to PhishTank1:

"Phishing is a fraudulent attempt, usually made through email, to steal your personal information"

Phishing Motives:

According to Weider D. et. al, the primary motives behind phishing attacks, from an attacker's perspective, are:

• Financial gain: phishers can use stolen banking credentials to their financial benefits.

• Identity hiding: instead of using stolen identities directly,
phishers might sell the identities to others whom might be criminals seeking ways to hide their identities and activities (e.g. purchase of goods).

• Fame and notoriety: phishers might attack victims for the sake of peer recognition.

PHISHING DETECTION APPROACHES:

Phishing detection schemes which detect phishing on the server side are better than phishing prevention strategies and user training systems. These systems can be used either via a web browser on the client or through specific host-site software.

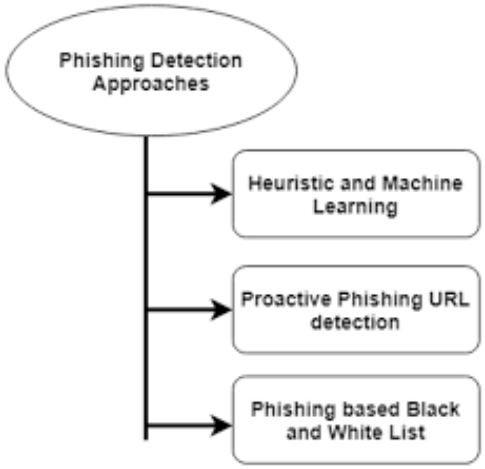

Fig:2 Anti phishing approaches

Heuristic and ML based approach is based on supervised and unsupervised learning techniques. It requires features or labels for learning an environment to make a prediction. Proactive phishing URL detection is similar to ML approach. However, URLs are processed and support a system to predict a URL as a legitimate or malicious. Blacklist and Whitelist approaches are the traditional methods to identify the phishing sites. The exponential growth of web domains reduces the performance of the traditional method

PHISHING REPORT:

The APWG Phishing Activity Trends Report analyzes phishing attacks and other identity theft techniques, as reported to the APWG by its member companies and its Global Research Partners.

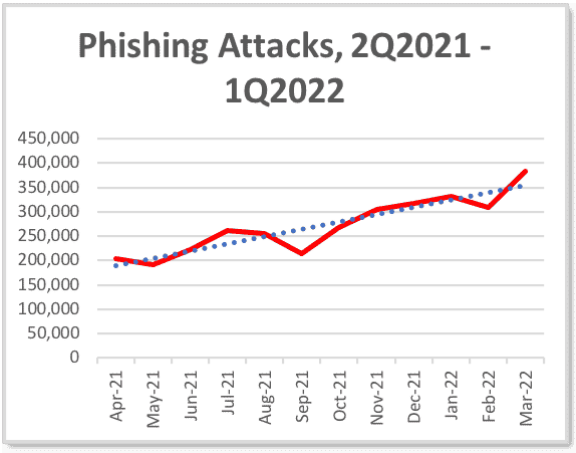

Fig:3 2Q2021-1Q2022 report

In the first quarter of 2022, APWG observed 1,025,968 total phishing attacks. This was the worst quarter for phishing that APWG has ever observed, and the first time that the quarterly total has exceeded one million.

The number of Unique Subjects has dipped as more submitted emails have had duplicative subject lines. The number of brands attacked each month has remained below the high of 715 observed in

September 2021, still reaching as high as 673 in March, 2022.

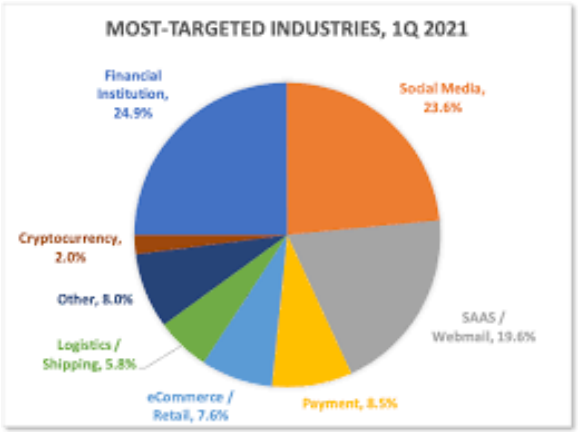

Fig:4 Most Targeted Industries

In the first quarter of 2022, APWG founding member OpSec Security found that phishing attacks against the financial sector, which includes banks, remained the largest set of attacks, accounting for 23.6 percent of all phishing. Attacks against webmail and software-as-a-service (SAAS) providers remained prevalent with attacks against retail/ecommerce sites falling from 17.3 to 14.6 percent after the holiday shopping season. Phishing against social media sets rose from 8.5 percent of all attacks in 4Q2021 to 12.5 percent in 1Q2022. Phishing against cryptocurrency target such as cryptocurrency exchanges and wallet providers—remained steady from late 2021, inching up from 6.5 to 6.6 percent in the latest quarter. OpSec Security offers world-class brand protection solutions.

METHODOLOGIES:

A. Protecting user against phishing using Antiphishing:

- AntiPhish is used to avoid users from using fraudulent web sites which in turn may lead to phishing attack. Here, AntiPhish traces the sensitive information to be filled by the user and alerts the user whenever he/she is attempting to share his/her information to a untrusted web site. The much effective elucidation for this is cultivating the users to approach only for trusted websites. However, this approach is unrealistic. Anyhow, the user may get tricked. Hence, it becomes mandatory for the associates to present such explanations to overcome the problem of phishing. Widely accepted alternatives are based on the creepy websites for the identification of "clones" and maintenance of records of phishing websites which are in hit list.

B. Learning to Detect Phishing Emails:

An alternative for detecting these attacks is a relevant process of reliability of machine on a trait intended for the reflection of the besieged deception of user by means of electronic communication. This approach can be used in the detection of phishing websites, or the text messages sent through emails that are used for trapping the victims. Approximately, 800 phishing mails and 7,000 non-phishing mails are traced till date and are detected accurately over 95% of them along with the categorization on the basis of 0.09% of the genuine emails. We can just wrap up with the methods for identifying the deception, along with the progressing nature of attacks.

C. Phishing detection system for using data mining: -

Phishing websites, mainly used for e-banking services, are very complex and dynamic to be identified and classified. Due

to the involvement of various ambiguities in the detection, certain crucial data mining techniques may prove an effective means in keeping the e-commerce websites safe since it deals with considering various quality factors rather than exact values. In this paper, an effective approach to overcome the "security issues" in the website phishing is used an intelligent resilient and effective model for detecting website phishing websites is put forth.

The objective of this project is to train machine learning models and deep neural nets on the dataset created to predict phishing websites. Both phishing and benign URLs of websites are gathered to form a dataset and from them required URL and website content-based features are extracted. The performance level of each model is measured and compared.

The following steps are followed in machine learning approach of the project,

1. Collection and analysis of dataset.
2. Visualization of data.
3. Preprocessing of data.
4. Train machine learning models using datasets.
5. Evaluation by comparison of models.

CONCLUSION:

In this project, is a mechanism to detect phishing websites. This methodology uses not just traditional URL based or content based rules but rather employs the machine learning technique to identify not so obvious patterns and relations in the data. We have used features from various domain spanning from URL to HTML tags of the webpage, from embedded URLs to favicon, and databases like WHOIS, Alexa, Pagerank, etc. to check the traffic and status of the website. Maximum efficiency can be obtained, thus classifying most websites correctly and proving the effectiveness of the machine learning based technique to attack the problem of phishing websites. We provided the output as a user-friendly web platform which can further be extended to a browser extension to provide safe and healthy online space to the users.