LITERATURE SURVEY

Team ID	**PNT2022TMID22501**
Project Name	**Web phishing detection**

LITERATURE SURVEY :

This article surveys the literature on the detection of phishing attacks. Phishing attacks target vulnerabilities that exist in systems due to the human factor. Many cyber attacks are spread via mechanisms that exploit weaknesses found in end-users, which makes users the weakest element in the security chain. The phishing problem is broad and no single silver-bullet solution exists to mitigate all the vulnerabilities effectively, thus multiple techniques are often implemented to mitigate specific attacks. This paper aims at surveying many of the recently proposed phishing mitigation techniques. A high-level overview of various categories of phishing mitigation techniques is also presented, such as: detection, offensive defense, correction, and prevention, which we belief is critical to present where the phishing detection techniques fit in the overall mitigation process.

INTRODUCTION :

Phishing is a type of **social engineering attack** often used to steal user data, including login credentials and credit card numbers. It occurs when an attacker, masquerading as a trusted entity, dupes a victim into opening an email, instant message, or text message. Internet constitutes a tremendous change in today's world because of its versatility. By utilising the sophisticated infrastructure of the internet, people can do transactions such as shopping, banking etc. whenever and wherever they want. The internet has many advantages, at the same time; it also has its own set of security and privacy problems. By using the anonymous and independent infrastructure of the internet, attackers create a prominent platform for cyber-attacks, such as phishing, malware distribution, privacy disclosure etc., which causes severe threats to the end-users of the internet.

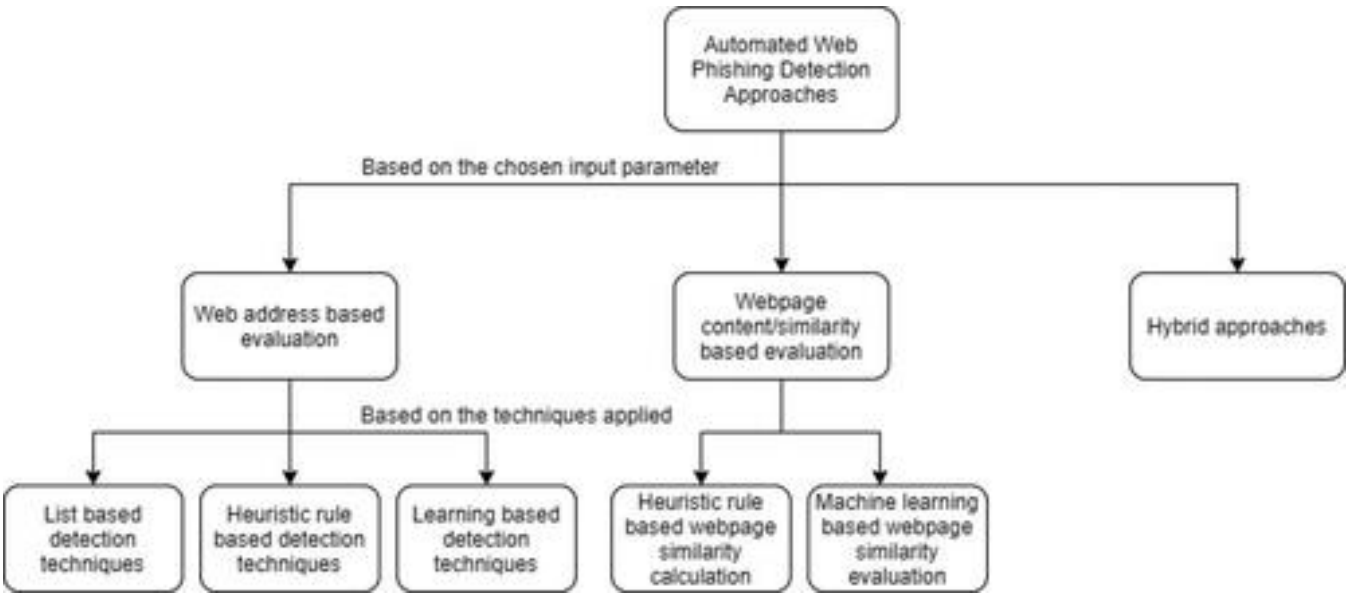
Phishing is the most widespread and pernicious cyberattack . which mostly targets the human rather than the computer by exploiting their vulnerabilities. According to Anti-Phishing Working Group (APWG), phishing is a criminal mechanism employing both social engineering and technical tricks to steal user's identity data and financial account credentials by disguising as a trusted one . To lure the end-users, attackers use malicious websites and e-mails by posing themselves as a trusted one. The supreme goal of phishing is to abduct confidential data such as user name, password, bank details, credit card details etc. Attackers perform phishing for many reasons – to gain benefits financially, to steal personal information, to ruin the reputation of the organisations and sometimes just to get fame .

The term 'phishing' is coined in the mid-1990s and is from the term 'fishing' because it involves trying to outwit someone into a trap . The history of phishing scams can be traced back to the beginning of the 1990s via America Online (AOL), which was carried out by generating random credit card numbers to make fake AOL accounts . Later, attackers turned it into a million-dollar growth business, by impersonating many organisations such as banks, credit card companies, online payment service providers (e.g. Instagram), and social media websites (e.g. Facebook). Even Internet giants such as Google and Facebook were scammed out of more than $100 million between 2013 and 2015 through email phishing . Recently in 2020, Texas school district lost $2.3 million in phishing raid and the Federal Bureau of Investigation is currently investigating on it. it is found that software as a service (SaaS) and webmail sites remained the biggest targets of phishing. Besides, 55% of phishing attacks detected in the second quarter of 2019 were using hypertext transfer protocol secure (HTTPS) encryption protocol and secure sockets layer (SSL) certificates to fool internet users.

TAXONOMY OF WEB PHISHING DETECTION:

There are two ways to protect the end-users against web phishing: user training and automated web phishing detection. During end-user training, the user is trained to confirm the originality of legitimate websites. However, end-user training is not sufficient enough to protect the users since the attackers promptly come up with alternate techniques to lure the users. Consequently, automated web phishing detection approaches were proposed by numerous researchers to protect the end-users from the web phishing

attack. In automated web phishing detection techniques, phishing detection is automated without any human intervention. The taxonomy of automated web phishing detection approaches presented is derived from the detailed study on the existing literature. All the web phishing detection approaches in the literature were classified based on the input parameters used in them. From the sub-class, the approaches were further classified based on the methodology applied to it.



Taxonomy of web phishing detection

The automated web phishing detection approaches are classified into the following three categories based on the input parameters.

- Web address-based evaluation
- Webpage content/similarity-based evaluation
- Hybrid approach

In web address-based evaluation schemes, the URL is analysed for detecting web phishing. In the case of a web page, content/similarity-based evaluation schemes the web page contents such as text features, hypertext mark-up language (HTML) features, cascading style sheets (CSS) features, and hyperlink features are evaluated to judge the

legitimacy of the website. Hybrid approaches are a combination of web address-based evaluation schemes and web page content/similarity based evaluation scheme.

PROBLEM IDENTIFICATION :

There are many users who purchase products through online platform and the payment is done through e-banking. There are some fake banking websites in which they collect the more sensitive information like username, password, credit card details etc , for illegal purpose. This type of websites are called phishing website. Here web phishing is one of the security threat to webservices on the internet

PROBLEM SOLUTION:

To overcome the problem of phishing website whenever we are clicking on one website it must show an alert box like it is a secure website or it is not a secure website. Then another way is that we can scan the website in order to prevent our system or mobile from the phishing attack. Even though technologies are there we as the user have to be aware of the websites whether it is secure or not. We should not click any unwanted websites .

CONCLUSION :

This paper aims to enhance detection method to detect phishing website using machine learning technology. Also , classifiers generated by machine learning algorithms identify legitimate phishing websites.The proposed technique can detect new temporary phishing sites and reduce the damage caused by phishing attacks. The performance of the proposed technique based on machine learning is more effective that previous phishing detection technologies. In the future, it will be useful to investigate the impact of feature selection using various algorithms. In this survey, a systematic review of current trends in web phishing detection is carried out and a taxonomy of web phishing detection is proposed

based on the input parameters chosen. The performance of the state-of-the-art web phishing detection approaches is evaluated and presented in detail. This paper also discussed the limitations of the existing web phishing detection techniques for future research direction. The analysis given in this paper will help the academia and industries to acknowledge the current status of web phishing detection and lead them to come up with new ideas to develop the best web anti-phishing technique.

REFERNCES:

1. *Anti-Phishing Working Group (APWG) Phishing activity trends report — second half2010*, December 2010, [online] Available: http://apwg.org/reports/apwg_report_h2_2010.pdf.
2. B. Schneier, "Lockheed Martin hack linked to RSA's SecurID breach", December 2011, [online] Available: http://www.schneier.com/blog/archives/2011/05/lockheed_martin.html.
3. B. Krebs, "HBGary Federal hacked by Anonymous", December 2011, [online] Available: http://krebsonsecurity.com/2011/02/hbgary-federal-hacked-by-anonymous/2011.
4. Higashino, M., et al. An Anti-phishing Training System for Security Awareness and Education Considering Prevention of Information Leakage. in 2019 5th International Conference on Information Management (ICIM). 2019.