LITERATURE SURVEY

WEB PHISHING DETECTION

Team ID	PNT2022TMID00904

INTRODUCTION:

Phishing is a social engineering attack that aims at exploiting the weakness found in system processes as caused by system users. For example, a system can be technically secure enough against password theft, however unaware end users may leak their passwords if an attacker asked them to update their passwords via a given Hypertext Transfer Protocol (HTTP) link, which ultimately threatens the overall security of the system.

Moreover, technical vulnerabilities (e.g. Domain Name System (DNS) cache poisoning) can be used by attackers to construct far more persuading socially-engineered messages (i.e. use of legitimate, but spoofed, domain names can be far more persuading than using different domain names). This makes phishing attacks a layered problem, and an effective mitigation would require addressing issues at the technical and human layers.

LITERATURE SURVEY:

[1] Social Phishing
AUTHORS: Rami Mustafa

Phishing is described as the art of emulating a website of a creditable firm intending to grab user's private information such as usernames, passwords and social security number. Phishing websites comprise a variety of cues within its content-parts as well as browser-based security indicators. Several solutions have been proposed to tackle phishing.

[2] New Rule-Based Phishing Detection Method
AUTHORS: Moghimi M. Varjani A.

A new rule-based method to detect phishing attacks in internet banking. Our rule-based method used two novel feature sets, which have been proposed to determine the webpage identity. Our proposed feature sets include four features to evaluate the page resources identity, and four features to identify the access protocol of page resource elements. We used approximate string matching algorithms to determine the relationship between the content and the URL of a page in our first proposed feature set.

[3] A Framework for Auto-Detection of Phishing Websites
AUTHORS: Hossein Shirazi, Kyle Haefnar, Indrakshi Ray.

For phishing websites, machine-learning data can be created using this framework. In this, they have used reduced features set and using python for building query. They build a large labelled dataset and analyse several machine-learning classifiers against this dataset.

[4] Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions
AUTHORS: Andrew H. Sung

Phishing has become an important cybersecurity problem. The centralized blacklist approach used by most web browsers usually fails to detect zero-day attacks, leaving the ordinary users vulnerable to new phishing schemes; therefore, learning machine based approaches have been implemented for phishing detection. Many existing techniques in phishing website detection seem to include as many features as can be conceived, while identifying a relevant and representative subset of features to construct an accurate classifier remains an interesting issue in this particular application of machine learning.

[5] Two Level Filtering Mechanism To Detect Phishing Sites Using Lightweight Visual Similarity Approach
AUTHORS: Rao R.S. Pais A.R.

The visual similarity-based techniques detect the phishing sites based on the similarity between the suspicious site and the existing database of resources such as screenshots, styles, logos, favicons etc. These techniques fail to detect phishing sites which target non-whitelisted legitimate domain or when phishing site with manipulated whitelisted legitimate content is encountered. Also, these techniques are not well adaptable at the client-side due to their computation and space complexity. Thus there is a need for light weight visual.

[6] A Phish Detector Using Lightweight Search Features
AUTHORS: Varshney G. Misra M

Web phishing is a well-known cyber-attack which is used by attackers to obtain vital information such as username, password, credit card number, social security number, and/or other credentials from Internet users via deception. A number of web phishing detection solutions have been proposed and implemented in the recent years. These solutions include the use of phishing black list, search engine, heuristics and machine learning, visual similarity techniques, DNS, access list and proactive phishing.