WEB PHISHING DETECTION

LITERATURE SURVEY

Literature Survey 1:

TITLE:

Phishing Web Page Detection Methods: URL and HTML Features Detection

AUTHORS: Humam, Faris, Setiadi, Yazid.

ABSTRACT:

Phishing is a form of online fraud where fraudulent web pages replicate real web pages and attempt to entice users into giving private and confidential information to the phisher. According to the numbers provided by APWG and Phistank, there will likely be a rise in phishing sites between 2015 and 2020. Numerous studies have been conducted to find solutions to this issue by utilising a variety of techniques. Sadly, the use of many approaches was concluded to be ineffective since design and assessment are solely concerned with achieving accuracy in detection and not with application in the actual world. A security detecting device should, however, be deployable, effective, and have good performance. In this study, the authors assessed a number of approaches and put forth rule-based software programs that are more effective in phishing detection.

Literature Survey 2:

TITLE:

Phishing Detection in Websites using Parse Tree Validation

AUTHORS: C. Emilin Shyni, Anesh D Sundar, G.S.Edwin Ebby.

ABSTRACT:

Phishing is a method of deceiving people into divulging personal information, such as usernames and passwords, credit card information, sensitive bank information, etc., through the use of spoof emails, instant messages, or phoney websites with a realistic-looking design. This study proposes a method for determining whether a webpage is real or phishing, known as parse tree validation. By capturing every hyperlink on a page using the Google API and building a parse tree out of the hyperlinks, it is a new way to identify phishing websites. Thousand phishing pages and thousand genuine pages are used to test this strategy. A false negative rate of 7.3% and a false positive rate of 5.2% was achieved.

Literature Survey 3:

TITLE:

WC-PAD: Web Crawling based Phishing Attack Detection

AUTHORS: Nathezhtha, Sangeetha, Vaidehi.

ABSTRACT:

Phishing is a crime that involves the stealing of users' private information. Phishing websites target people, businesses, hosting services for cloud storage, and official websites. Although software-based approaches to anti-phishing are preferable owing to cost and operational considerations, hardware-based alternatives are still often deployed. The existing methods for phishing detection are unable to address issues like zero-day phishing website attacks. A three-phase attack detection system called the Web Crawler based Phishing Attack Detector (WC-PAD) has been presented to address these problems and accurately detect the existence of phishing. It classifies phishing and non-phishing websites based on input factors such as web traffic, web content, and Uniform Resource Locators (URLs). With datasets gathered from actual phishing situations, the proposed WC-experimental PAD's study is carried out. According to the testing results, the suggested WCPAD provides 98.9% accuracy in phishing attack detection, including zero-day phishing attacks.

Literature Survey 4:

TITLE:

Phishing Detection from URLs Using Deep Learning Approach

AUTHORS: Shweta Singh, M.P. Singh, Ramprakash Pandey

ABSTRACT:

The Internet is now accessible everywhere. People enjoy using an e-commerce platform to buy or sell their goods all over the world. As a result, cyberattackers now gravitate toward crimes in cyberspace. Phishing is one such strategy where attackers/criminals have used the unidentified structure of the Internet with the intention of misleading people with the use of the fictitious website and emails in order to gain their credentials (like account numbers, passwords, and PINs). Due to this semantic framework, it might be difficult to tell whether a web page is real or phishing. In order to stop such attempts, a phishing detection system is created in this work using deep learning methods. Convolutional neural networks (CNNs) are used by the system to analyse URLs in order to identify phishing websites. Our proposed system's accuracy was better than the prior model in paper [19], which had a 97.98% accuracy rating, at 98.00%. As the CNN automatically extracts features from the URLs through its hidden layers, this system doesn't

require any feature engineering. This is another benefit of the proposed approach above that which was previously disclosed in [19], as feature engineering takes a lot of effort.

Literature Survey 5:

TITLE:

A Deep Learning-Based Framework for Phishing Website Detection
AUTHORS: LIZHEN TANG, QUSAY H. MAHMOUD

ABSTRACT:

Phishing attackers spread phishing links through e-mail, text messages, and social media platforms. They use social engineering skills to trick users into visiting phishing websites and entering crucial personal information. In the end, the stolen personal information is used to defraud the trust of regular websites or financial institutions to obtain illegal benefits. With the development and applications of machine learning technology, many machine learning-based solutions for detecting phishing have been proposed. Some solutions are based on the features extracted by rules, and some of the features need to rely on thirdparty services, which will cause instability and time-consuming issues in the prediction service. In this paper, a deep learning-based framework for phishing website detection is proposed. When a user visits

a website, the framework has been implemented as a browser plug-in that can alert them if there is a phishing danger in real time. The real-time prediction service integrates a number of techniques, such as whitelist filtering, blacklist interception, and machine learning (ML) prediction, to increase accuracy, lower false alarm rates, and shorten computation times. We compared various machine learning models utilising various datasets in the ML prediction module. The RNN-GRU model has the highest accuracy of 99.18% according to the trial findings, proving the viability of the suggested approach.