LITERATURE SURVEY

WEB PHISHING DETECTION

ABSTRACT :

Phishing is a crime in which a perpetrator sends the fake e-mail, which appears to come from popular and trusted brand or organization, asking to input personal credential like bank password, username, phone number, address, credit card details.

A web service is one of the most important Internet communications software services. Using fraudulent methods to get personal information is increasingly widespread these days. Though, it makes our lives easier, it leads to numerous security vulnerabilities to the Internet's private structure. Web phishing is one of the many security risks that web services face. Phishing messages propagate over e-mail, SMS, instant messengers, social networking sites, VoIP, and so forth, but e-mail is the popular way to perform this attack and 65% of the total phishing attack is achieved by visiting the hyperlink attached to the e-mail. Phishing assaults are usually detected by experienced users. However, security is a primary concern for system users who are unaware of such situations. Phishing is the act of portraying malicious web runners as genuine web runners to obtain sensitive information from the end-user. Phishing is currently regarded as one of the most dangerous threats to web security. Vicious Web sites significantly encourage Internet criminal activity and inhibit the growth of Web services. As a result, there has been a tremendous push to build a comprehensive solution to prevent users from accessing such websites. We suggest a literacy-based strategy to categorize Web sites into three categories: benign, spam, and malicious. Our technology merely examines the Uniform Resource Locator (URL) itself, not the content of Web pages. As a result, it removes run-time stillness and the risk of drug users being exposed to cyber surfer-based vulnerabilities. When compared to a blacklisting service, our approach performs better on generality and content since it uses learning techniques.

Keywords: Security; Web Services; URL; Vulnerabilities

LITERATURE SURVEY:

The current circumstance is that the population's maturity has been wisecracked, causing them to unknowingly give their private information to hackers. Several banned websites have already been established to seem like that of an actual point of contact through obtaining stoners' private information. Passcode, savings account, and shipping information are just a few examples. Late in 2016, the amount of hacking activities was at an all-time high since the company started monitoring this in 2004. The overall identified phishing attacks in 2016 were 1,609. This represents a 65 percent increase over 2015. Within the final quarter of 2004, there would be scamming attempts each month. Machine Learning was used to find the phishing website. The use of machine literacy to surround the supplied features is the basis of Grounded Malware Monitoring Systems. Features are generated by assembling items in a specific order, such as URLs, sphere names, website features, and website content.

Phishing attacks aim to steal confidential information using sophisticated methods, techniques, and tools such as phishing through content injection, social engineering, online social networks, and mobile applications. To avoid and mitigate the risks of these attacks, several phishing detection approaches were developed, among which deep learning algorithms provided promising results. However, the results and the corresponding lessons learned are fragmented over many different studies and there is a lack of a systematic overview of the use of deep learning algorithms in phishing detection. Hence, we performed a systematic literature review (SLR) to identify, assess, and synthesize the results on deep learning approaches for phishing detection as reported by the selected scientific publications. We address nine research questions and provide an overview of how deep learning algorithms have been used for phishing detection from several aspects. In total, 43 journal articles were selected from electronic databases to derive the answers for the defined research questions. Our SLR study shows that except for one study, all the provided models applied supervised deep learning algorithms. The widely used data sources were URL-related data, third party information on the website, website content-related data, and email. The most used deep learning algorithms were deep neural networks (DNN), convolutional neural networks, and recurrent neural networks/long short-term memory networks. DNN

and hybrid deep learning algorithms provided the best performance among other deep learning-based algorithms. 72% of the studies did not apply any feature selection algorithm to build the prediction model. PhishTank was the most used dataset among other datasets. While Keras and Tensorflow were the most preferred deep learning frameworks, 46% of the articles did not mention any framework.

REFERENCES:

1. Detecting Phishing Websites Using Machine Learning by Sagar Patil,Yogesh Shetye, Nilesh Shendage published in the year 2020.
2. Phishing Website Classification and Detection Using Machine Learning by Jitendra Kumar, A. Santhanavijayan, B. Janet, Balaji Rajendran, B.S. Bindhumadhava was published in the year 2020.
3. Adebowale MA, Lwin KT, Hossain MA (2020) Intelligent phishing detection scheme using deep learning algorithms. J Enterp Inf Manag. https://doi.org/10.1108/JEIM-01-2020-0036
4. Alom MZ, Taha TM (2017) Network intrusion detection for cyber security using unsupervised deep learning approaches. In: 2017 IEEE national aerospace and electronics conference (NAECON). IEEE, pp 63–69
5. Benavides E, Fuertes W, Sanchez S, Sanchez M (2020) Classification of phishing attack solutions by employing deep learning techniques: a systematic literature review. In: Developments and advances in defense and security. Springer, Singapore, pp 51–64