Web Phishing Detection

Literature Survey

Survey 1:

Title: Phish-Defence: Phishing Detection Using Deep Recurrent Neural Networks

Abstract: In the growing world of the internet, the number of ways to obtain crucial data such as passwords and login credentials, as well as sensitive personal information has expanded. Page impersonation, often known as phishing, is one method of obtaining such valuable information. Phishing is one of the most straightforward forms of cyber-attack for hackers, as well as one of the simplest for victims to fall for. It can also provide hackers with everything they need to access their targets' personal and corporate accounts. Such websites do not provide a service but instead gather personal information from users. In this paper, we achieved state-of-the-art accuracy in detecting malicious URLs using recurrent neural networks. Unlike previous studies, which looked at online content, URLs, and traffic numbers, we merely look at the text in the URL, which makes it quicker and catches zero-day assaults. The network has been optimized so that it may be utilized on tiny devices like Mobiles, Raspberry Pi without sacrificing the inference time.

Algorithm Used: Long Short-Term Memory, Gated Recurrent unit

Survey 2:

Title: Phishing website detection using machine learning and deep learning techniques

Abstract:

Phishing has become more damaging nowadays because of the rapid growth of internet users. The phishing attack is now a big threat to people's daily life and to the internet environment. In these attacks, the attacker impersonates a trusted entity intending to steal sensitive information or the digital identity of the user, e.g., account credentials, credit card numbers and other user details. A phishing website is a website which is similar in name and appearance to an official website otherwise known as a spoofed website which is created to fool an individual and steal their personal credentials. So, to identify the websites which are fraud, this paper will discuss the machine learning and deep learning algorithms and apply all these algorithms on our dataset and the best algorithm having the best precision and accuracy is selected for the phishing website detection. This work can provide more effective defenses for phishing attacks of the future.

Algorithm Used: Logistic Regression, K Nearest Neighbour, Decision Tree, Random Forest, XG Boost, Ada Boost

Survey 3:

Title: Intelligent Phishing Detection Scheme Algorithms Using Deep Learning

Abstract:

Phishing attacks have evolved in recent years due to high-tech-enabled economic growth worldwide. The rise in all types of fraud loss in 2019 has been attributed to the increase in deception scams and impersonation, as well as to sophisticated online attacks such as phishing. The global impact of phishing attacks will continue to intensify and thus a more efficient phishing detection method is required to protect online user activities. To address this need, this study focused on the design and development of a deep learning-based phishing detection solution that leveraged the universal resource locator and website content such as images, text and frames.

Algorithm Used: Convolutional Neural Network, Long Short Term Memory, Intelligent Phishing Detection System

Survey 4:

Ttitle: Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning

Abstract: Phishing has become one of the biggest and most effective cyber threats, causing hundreds of millions of dollars in losses and millions of data breaches every year. Currently, anti-phishing techniques require experts to extract phishing sites features and use third-party services to detect phishing sites. These techniques have some limitations, one of which is that extracting phishing features requires expertise and is time-consuming. Second, the use of third-party services delays the detection of phishing sites. Hence, this paper proposes an integrated phishing website detection method based on convolutional neural networks (CNN) and random forest (RF). The method can predict the legitimacy of URLs without accessing the web content or using third-party services. The proposed technique uses character embedding techniques to convert URLs into fixed-size matrices, extract features at different levels using CNN models, classify multi-level features using multiple RF classifiers, and, finally, output prediction results using a winner-take-all approach. On our dataset, a 99.35% accuracy rate was achieved using the proposed model. An accuracy rate of 99.26% was achieved on the benchmark data, much higher than that of the existing extreme model

Algorithm Used: Random Forest, Convolutional Neural Network

References:

1. Aman Rangapur, Tarun Kanakam and Dhanvanthini P, “Phish-Defence: Phishing Detection Using Deep Recurrent Neural Networks” -September 2022
2. M Selvakumari et al, “Phishing website detection using machine learning and deep learning techniques” – 2021
3. M.A.Adebowale, K.T.Lwin, M.A.Hosaain, “Intelligent Phishing Detection Scheme Algorithms Using Deep Learning”
4. Rundong Yang, Kangfeng Zheng, Bin Wu, Chunhua Wu and Xiujuan Wang, “Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning”