WEBSITE PHISHING DETECTION

Team Members:

RAMANIDHARAN V-2019105031

PRADEEP T – 2019105555

VIGNESHKUMAR S – 2019105066

GOWTHAM T - 2019105533

ABSTRACT:

Phishing attack is a simplest way to obtain sensitive information from innocent users. Aim of the phishers is to acquire critical information like username, password and bank account details. Cyber security persons are now looking for trustworthy and steady detection techniques for phishing websites detection. This paper deals with machine learning technology for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. Decision Tree, random forest and Support vector machine algorithms are used to detect phishing websites. Aim of the paper is to detect phishing URLs as well as narrow down to best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm.

LITERATURE SURVEY

S.NO	TITLE	AUTHORS	ABSTRACT
1.	Detecting Phishing Websites Using Machine Learning	Amani Alswailem, Bashayr Alabdullah, Norah Alrumayh, Dr.Aram Alsedrani.	Phishing website is one of the internet security problems that target the human vulnerabilities rather than software vulnerabilities. It can be described as the process of attracting online users to obtain their sensitive information such as usernames and passwords. In this paper, we offer an intelligent system for detecting phishing websites. The system acts as an additional functionality to an internet browser as an extension that automatically notifies the user when it detects a phishing website. The system is based on a machine learning method, particularly supervised learning. We have selected the Random Forest technique due to its good performance in classification. Our focus is to pursue a higher performance classifier by studying the features of phishing website and choose the better combination of them to train the classifier. As a result, we conclude our paper with accuracy of 98.8% and combination of 26 features.

2.	Detection of Phishing Websites by Using Machine Learning-Based URL Analysis	Mehmet Korkmaz, Ozgur Koray Sahingoz, Banu Diri.	In recent years, with the increasing use of mobile devices, there is a growing trend to move almost all real-world operations to the cyberworld. Although this makes easy our daily lives, it also brings many security breaches due to the anonymous structure of the Internet. Used antivirus programs and firewall systems can prevent most of the attacks. However, experienced attackers target on the weakness of the computer users by trying to phish them with bogus webpages. These pages imitate some popular banking, social media, e-commerce, etc. sites to steal some sensitive information such as, user-ids, passwords, bank account, credit card numbers, etc. Phishing detection is a challenging problem, and many different solutions are proposed in the market as a blacklist, rule-based detection, anomaly-based detection, etc. In the literature, it is seen that current works tend on the use of machine learning-based anomaly detection due to its dynamic structure, especially for catching the "zero-day" attacks.
3.	Machine LearningTechniquesfor Detection of Website Phishing: A Review forPromises and Challenges	Ammar Odeh, Eman Abdelfattah, Ismail Keshta.	Attackers fool the users by presenting the masked webpage as legitimate or trustworthy to retrieve their essential data. Several solutions to phishing websites attacks have been proposed such as heuristics, blacklist or whitelist, and Machine Learning (ML) based techniques. This paper presents the state of art techniques for phishing website detection using the ML techniques. This research identifies solutions to the website's phishing problem based on the ML techniques. The majority of the examined approaches are focused on traditional ML techniques. Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Ada Boosting are the powerful ML techniques examined in the literature. This survey paper also identifies deep learning-based techniques with better performance for detecting phishing websites compared to the conventional ML techniques. Challenges to ML techniques identified in this work include overfitting, low accuracy, and ML techniques' ineffectiveness in case of unavailability of enough training data.
4.	Detection of Phishing Websites from URLs by using Classification Techniques on WEKA	Buket Geyik, Kiiubra Erensoy, Emre Kocyirit.	The Internet is getting stronger day by day and it makes our lives easier with many applications that are executed on cyberworld. However, with the development of the internet, cyber-attacks have increased gradually and identity thefts have emerged. It is a type of fraud committed by intruders by using fake web pages to access people's private information such as userid, password, credit card number and bank account numbers, etc. These scammers can also send e-mail from many important institutions and organizations by using phishing attacks which imitate these web pages and acts as if they are original. Traditional security mechanisms can not prevent these attacks because they directly target the weakest part of connection : end-users. Machine learning technology has been used to detect and prevent this type of intrusions. The antiphishing method

			has been developed by detecting the attacks made with the technologies used. In this paper, we combined the websites used by phishing attacks into a dataset
5.	Phishing Website Detection on Machine Learning Algorithm	Weiheng Bai	Phishing websites are a means to deceive users personal information by using various means to impersonate the URL address and page content of a real website.This paper analyses the structural ,extracts features of the URL of the phishing website,extracts 12 kinds of features and uses four machine learning algorithms for training. Then, use the best performing algorithm as our model to identity unknown URL's.After the Reconition is completed,a snapshot of the web page is extracted and compared with the regular web pages snapshot to implement the recommendation of the original regular web page of the phishing web page.