Literature Survey

Paper 1:

Title: Phishing Website Detection Using ML

Year: 2021

Authors: Nikhil K*, Dr. Rajesh D S, Dhanush Raghavan

Description:

In emerging technology, industry, which deeply influence today's security problems, has given a headache to many employers and home users. Occurrences that exploit human vulnerabilities have been on the upsurge in recent years. In these new times there are many security systems being enabled to ensure security is given the outmost priority and prevention to be taken from being hacked by those who are involved in cyber-offenses and essential prevention is taken as high importance in organization to ensure network security is not being compromised. Cyber security employee are currently searching for trustworthy and steady detection techniques for phishing websites detection. Due to wide usage of internet to perform various activities such as online bill payment, banking transaction, online shopping, etc. Customer face numerous security threats like cybercrime. Many cybercrime is being casually executed for example spam, fraud, identity theft cyber terrorisms and phishing. Among this phishing is known as the most common cybercrime today. Phishing has become one amongst the top three most current methods of law breaking in line with recent reports, and both frequency of events and user weakness has increased in recent years, more combination of all these methods result in greater danger of economic damage. Phishing is a social engineering attack that targets and exploiting the weakness found in the system at the user's end. This paper proposes the Agile Unified Process (AUP) to detect duplicate websites that can potentially collect sensitive information about the user. The system checks the blacklisted sites in dataset and learns the patterns followed by the phishing websites and applies it to further given inputs. The system sends a pop-up and an e-mail notification to the user, if the user clicks on a phishing link and redirects to the site if it is a safe website. This system does not support real time detection of phishing sites; user has to supply the website link to the system developed with Microsoft Visual Studio 2010 Ultimate and MySQL stocks up data and to implement database in this system.

Paper - 2:

"A Practical Guide to Anomaly Detection Implications of meeting new FFIEC minimum expectations for layered security".

Commercial and retail account holders at financial institutions of all sizes are under attacks by sophisticated, Organized, Well-funded cyber criminals. Anomaly detection solutions are readily available, are deployed quickly and immediately and automatically protect all account holders against all types of fraud attack with minimal Disruption to legitimate online banking activityAnomaly detection solutions are readily available, are deployed quickly and immediately and automatically protect all account holders against all types of fraud attack with minimal Disruption to legitimate online banking activity

---

Paper - 3:

Title: Detection of Phishing Websites using Machine Learning

Year: 2020

Authors: Abdul Razaque, Fathi Amsaad , Mohamed Ben Haj Frej

Description:

With the widespread usage of the Internet for online banking and trade, phishing attacks and forms of identity theft-based scams are becoming extremely popular among the hacker communities. In 2004 alone, more than 50 million phishing emails were sent. Their result was 10 billion dollars of damage to banks and financial institutions . Most of the recent phishing attacks are carried out as a three-step process. In the first step, the phishers send emails to their victims from social engineering attacks, webpages, and forums. Large volumes of phishing emails with legal banking domains are sent out using anonymous servers or compromised machines. These emails contain hyperlinks with an appearance similar to the legitimate website. The fake webpage contains input forms requesting personal critical information such as credit card, social security numbers, mother's maiden name, etc. Although existing spam filtering techniques can be employed to combat phishing emails, these measures are not entirely scalable. Several readily available tools can bypass both the statistical and rule-based spam filters. As these mechanisms are not uniquely tuned for the detection of phishing emails despite their existence, the threats caused by phishing emails are prevalent. Furthermore, unlike spamming, which impacts bandwidth, phishing attacks directly affect their victims by inflicting a hefty loss due to monetary damage.

Paper 4:

M. Amaad et al. presented a hybrid model for classification of phishing website. In this paper, proposed model carried out in two phase. In phase 1,they individually perform classification techniques, and select the best three models based on high accuracy and other performance criteria. While in phase 2, they further combined each individual model with best three model and makes hybrid model that gives better accuracy than individual model. They achieved 97.75% accuracy on testing dataset. There is limitation of this model that it requires more time to build hybrid model.

Paper 5:

SANS Institute, "Phishing : An Analysis of a Growing Problem",2007.

This paper gives an in depth analysis of phishing : what it is, the technologies and security Weaknesses it takes advantage of the dangers it poses to end users.
In this analysis author explain the concepts and technology behind
phishing, show how the threat is much more then just a nuisance or passing trend, and discuss how gangs of criminals are Using Unfortunately, a growing number of cyber-thieves Are using these same systems to manipulate us and steal our private information.

Paper 6:

"**Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning**" ,

This research was funded by the National Key R & D Program of China Grant Numbers 2017YFB0802800 and Beijing Natural Science Foundation (4202002)

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 A 99.35% correct classification rate of phishing websites was obtained on the dataset.Experiments were conducted on the test set and training set, and the experimental results proved that the proposed method has good generalization ability and is useful in practical applications. It takes longer to train. However, the trained model is better than the others in terms of accuracy of phishing website detection. Another disadvantage is that the model cannot determine whether the URL is active or not, so it is necessary to test whether the

URL is active or not before detection to ensure the effectiveness of detection. In addition, some attackers use URLs that are not imitations of using CNN models, classify multi-le vel features using multiple RF classi fiers, and, finally, output prediction results using a winner-take-all approach. other websites, and such URLs will not be detected