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:

Title: Web Phishing Detection using Machine Learning

Year: 2022

Authors: N Kumaran, Purandhar Sri Sai, Lokesh Manikanta

Description:

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. Because of its nonlinear system, it has a high level of fashion ability in terms of web security, particularly for the detection of anomalies on internet spots. The features retrieved utilizing machine literacy approaches are compared to extracting features through URLs, primary law, or third-party services. A process of machine trust ability on a particularity meant for the reflection of the besieged deceit of stoners through electronic communication is a relevant approach for detecting these attacks.This method can be used to find phishing websites or textbook dispatches sent over email to confuse the victims. This method was presented by S. Marchal et al. to distinguish Malicious URLs based on the assessment of legitimate point garçon record data. By the off operation or the detection of a malicious site. Open source demonstrates several remarkable characteristics, including high proximity, total autonomy, excellent linguistic flexibility, quickness in choosing, inflexibility towards active phishing, and inflexibility towards development in phishing methods. Mustafa Aydin et al. presented the bracket method to fraudulent site detection that involves rooted websites ‘URL properties and evaluating subset-grounded Point selection approaches. For the detection of phishing websites, it uses point birth and selecting styles. Fadi Thabtah et al. evaluated vast numbers of ML methods to actual malware datasets and according to many parameters. The goal of this comparison is to highlight the benefits and drawbacks of ML predictive models, as well as their real performance in phishing attempts. Covering approach models are more appropriate as anti-phishing results, according to the experimental results. Muhemmet Baykara et al. developed the Anti Phishing Simulator, which gives data on the phishing discovery challenge as well as how to detect phishing emails. Only utilize the textbook of the e-mail as just a term to execute complicated word processing, according to the study's recommendations.

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.

Moreover, attackers can use technical vulnerabilities to construct socially engineered messages (i.e., use of legitimate, but spoofed, domain names can be far more persuasive than using different domain names), which makes phishing attacks a severe problem. Effective mitigation would require addressing issues at the technical and human layers. Since phishing attacks aim at exploiting weaknesses found in humans (i.e., system end-users), it is difficult to mitigate them. For example, as evaluated, end-users failed to detect 29% of phishing attacks, even when trained with the best performing user awareness program . On the other hand, software phishing detection techniques are evaluated against phishing attacks, which makes their performance practically unknown with targeted forms of phishing attacks. These limitations in phishing mitigation techniques have almost resulted in security breaches against several organizations, including leading information security providers .

More specifically, we highlight the main contributions as follows:

 The novel extension of the Google Chrome web-browser is based on Blacklisting and semantic analysis methods that will be integrated successfully to efficiently identify and prevent the phishing attack.

 IP URLs and redirection of the user's information are checked successfully using phishing detection. User's redirection algorithms are proposed.