International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 2, pp 10-14, November-2017

A Framework for Detecting Phishing Websites using EDA algorithm and URL based Website Classification

Dr Jayaprakash Chinnadurai

Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering for Women, Hyderabad -500100, Telangana, India

Abstract: In the modern era of information technology being connected on a social media platform and mail services have been an abundant process in the life of human being; along with the virtues of instant connectivity and exchange of information via an internet platform, some common social engineering attacks are been carried out by the evil-minded people namely called as hackers. Web attacks are the major part of cybercrime in which criminal uses internet services or URLs of related or similar identity as a mediator for resembling the legitimate website with a motive to steal some personal information about user entity that is not been publicly available and use them this information for personal benefit to gain access to the social media accounts or to access the bank account for laundering the money or to gain profit in any term. This website could be extremely dangerous for both the user end and the service provider. Thus, to detain the user from getting fraud and detect various phishing websites a proper proactive data analysis is abundant so that by using the analyzed data, the internet services can be more secure and reliable to transact with.

Keywords: Phishing attacks, fraud, scam, detention, hackers, EDA algorithm- Booster, intrusion detection, Address bar.

1. INTRODUCTION

Phishing nowadays can be so powerful that it can mislead a human to perform a cybercrime too. Humans are been habitual of Searching for information or purchasing a product from the online market because it is a very easy task nowadays. Also, many people are been addicted to surfing on the internet so it has been a boon in the life of human beings, but along with this some bad intentions people make use of their evil mind and try to exploit the users of the internet for their benefit and make the internet a curse. The Attackers use many of such cybercrime attacks each day and is often successful in this attack due to lack of information by the USER using the resources. The Research paper is structured in the below format; in which section 2, we have summarized information about phishing and try to elaborate some questions like, what is phishing? What is the technical perspective of phishing? How it is held? We have also discussed some of the related work and existing projects for the detection of phishing URLs. Section 3 consists of all the information about the proposed system including its detailed description, the purpose of the system, its architecture, Feature extraction parameters, Implementation of the resultant output filesystem. The resultant output of our experiment is analyzed in section 4. The conclusion is described in section 5.

Social-Engineering attack (Phishing)

Social engineering attack mainly deals with the human psychology and susceptibility to manipulate and unleash a sensitive data from a victim or break security measures that allow an attacker to access the network. Phishing is one of the most preferred and used methodologies to trap the victim. In these Phishing attacks, the attacker delivers a phony transmission that seems to be designated through a legitimate source. The purpose against performing such an attack is to gain secret information including the credit card details, password of an account or to install an untrusted source file to the victim's computer as a means to gain access to the system. This attack is a prevalent kind of cyber-attack where each user of the internet must learn to bulwark themselves. Fig(1) describes the step-by-step method used by the attacker. In which, the attack begins with an unauthentically bogus email or some form of transmission that is intended to attract a victim. The message seems to come from a reliable source in this form of attack. If the victim is attracted by the appealed offer, the victim is more likely to lead some sensitive details on a scammed platform. A malware file or a trojan file of the virus is often attached with such emails and the if victim downloads such file to the device the virus starts to destroy the system.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 2, pp 10-14, November-2017

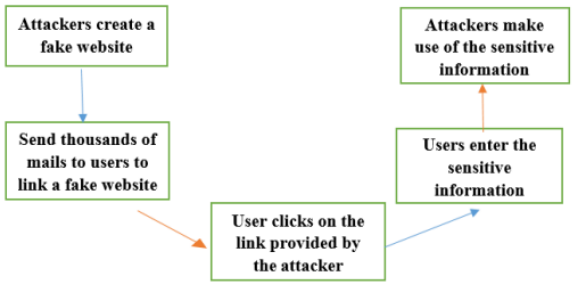

Fig (1) Method used by an attacker for Phishing

Assailants get to benefit financially from obtaining the credit card details or other personal information of their victims [5]. Phishing emails are also sent to obtain authentic information or other data about staff to use them in a sophisticated assault on a specific organization. Phishing is a popular starting point for cyberattacks such as Advanced Perpetual Threats and ransomware. In a phishing attack, information gathering is a very crucial step also called reconnaissance in which the attacker collects context knowledge of targets' personal and job backgrounds, intrigues, or social activities by using social information tools, like gregarious networks including LinkedIn, Facebook However Twitter. Hackers may use this information to classify prospective victims' names, work titles, and email addresses, as well as the knowledge about denominations of important workers at the workplace. This data will then be used to compose a trustworthy e-mail or text message. Criminals can also be an advanced sedulous threat group, for example, typically begin with a malicious connection or attachment in an e-mail. The most common susceptibility or clickable phishing environments in this form of attack have been described as Facebook victuals. If phishing assailants are created, they are often used to spread false information. A victim is usually given a however if the high seems to come from Kennedy individual, the agency [17]. Malevolent file injection, which involves phishing malware, or connections to malevolent websites is used to carry out the attack.fig (2) In any situation, the aim is to guide the target to a malevolent site where they can download malevolent software or be duped into sharing personal and financial information, and also in some cases the device used by the victim person is compromised and damaged.

2. ABOUT THE SYSTEM

In this project, Artificial intelligence models are used for automation in detecting Phishing sites. Working of these models depends on the Pre-requisite function that is the Feature Extraction Module where discrimination of data into legitimate sites and Phishing sites is derived from a collection of websites in a CSV (comma-separated values) file. The overall perspective of this project mainly focuses on increasing the average accuracy rate of ML and proactive monitoring with warning of a webpage with help of embedding integration of web plug-ins that detects the site to be legitimate or fraudulent before redirection of URL and defies the user from forgery. In machine re-training alone it is not enough to overcome new attacks, new features and strategies are needed to stop the attack from deceptive detection systems. Thus continuous proactive monitoring should be done on the activities which can be achieved by using a web extension that can detect the URL nature.

Phishing can be very dangerous. Its severity could be so high e.g., intrusion of the backdoor file in a system to gain access to user documents and also to use the system for personal gain. One might not be aware that such criminal activities are also being conducted by evil people; the lack of knowledge is the main reason for the attack to be successful. Thus, creating awareness about the attack and its importance is not suitable for all users. Hence an automated system is thus crucial for checking whether the recipient's website is Legitimate or it is a kind of Bait received for catching a Whale.No single solution has been introduced to give the total immunity from the attack held by phishing.

3. SYSTEM ARCHITECTURE

The system architecture of a proposed system for the detection of Phishing URLs focuses on the correct selection of Features based on which the ML will perform its operation. The system design includes the prevention of URL redirection and pop-up a window to indicate the results to the user.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 1, Issue 2, pp 10-14, November-2017

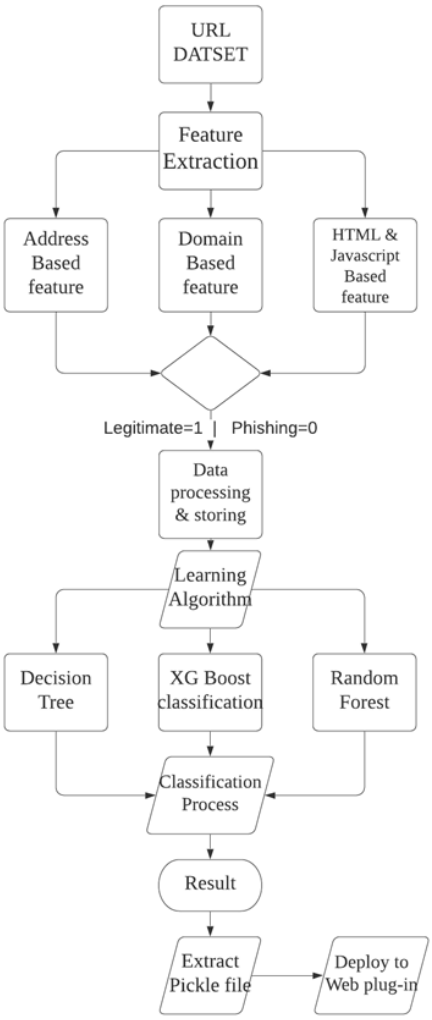

Fig (2) System Block Diagram

The system is based on an analysis of the feature extraction performed on the open-source and user-entered database. Extraction of Key factor includes:

Address Bar Based Features

a) The domain of URL- Extraction of Domain present in the URL by matching with the URL-parse after "WWW."

b) IP Address in URL: -If an IP address is present there might be a possibility that an attacker is trying to exploit it. As a result, if an IP address is used instead of a domain name, it should indicate spoofing.

c) The existence of the '@' sign in the URL is verified. When you use the "@" symbol in a URL, the browser ignores anything before the "@" symbol, and the actual address always comes after the "@" symbol.

d) URL Length- Calculating the URL's length. Phishers will mask the suspicious aspect of a URL in the address bar by using a long URL. If the URL is longer than or equivalent to 54 characters, it is considered phishing. Otherwise, it is considered valid.

e) Depth of URL- Computes the depth of the URL. This feature calculates the number of sub-pages in the given URL based on the '/'. The value of the feature is numerical based on the URL.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 2, pp 10-14, November-2017

f) URL redirection "/": - The presence of the character "//" in the URL indicates that the user will be forwarded to another website. The "//" in the URL's position is calculated. We discovered that if the URL begins with "HTTP," the "//" should be placed in the sixth row, however, if the URL, uses "HTTPS," the "//" should appear in the seventh point.

g) "HTTP/HTTPS" in Domain Name-Checks for the presence of "HTTP/HTTPS. To deceive users, phishers can add the "HTTPS" token to the domain part of a URL.

h) Shortening URL: - It is a method of reducing the length of a URL while still directing it to the desired webpage on the "World Wide Web." This is achieved by using an "HTTP Redirect" on a short domain name that points to a long URL website.

i) Prefix or Suffix "-" in Domain-Checking the presence of '-' in the domain part of URL. In legitimate URLs, the dash symbol is seldom used. Phishers also apply prefixes or

4. RESULT

Because of the increasing proliferation of phishing attacks, various methods of avoidance have evolved. Existing server and anti-spam filtering clients are used to detect different aspects of spam messages, but since the firewall is just a wall of security with tiny gaps or vulnerabilities, the attacker might be able to get through it. In this experiment, Several useful tricks have been developed in which a model decides whether a dataset containing websites is true or fake, thanks to the integration of Machine Learning. To get the best results selection of feature and classifier is very crucial.

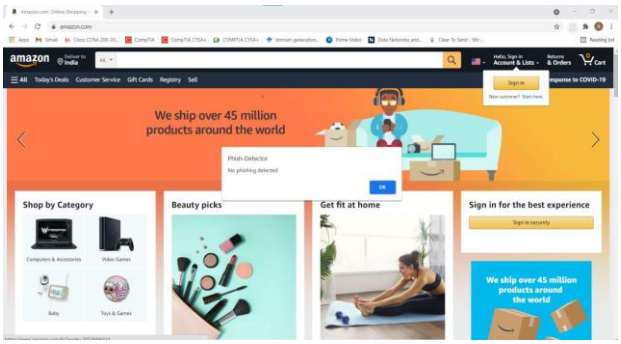

Fig (3): The pop-up indication of Legitimate Website

5. CONCLUSION

This project aims to solve the issue of unawareness of the nature of websites and identify it by using datasets created specifically for this purpose and to train machine learning models to detect phishing websites. The dataset, which contains both phishing and benevolent URLs of websites, is used to generate the necessary URL and website content-based functionality. The performance standard of each model is estimated and compared. The Implementation of this proposed system will avoid the phishing site to compromise a user by identifying the attack and notifying the same. Henceforth, to avoid such an attack the preventive measure is taken into account as a troublesome job within the system security domain. A good identification system is now able to detect phishing attacks with a limited number of false positives. Data analysis and heuristics, machine learning, and deep learning algorithms are some of the guiding techniques discussed in this article. While heuristic and data analysis methods have low False Positive rates and high computational costs, they are better at identifying phishing attacks. As opposed to other approaches, ML procedures provide the most straightforward outcomes. Some machine learning algorithms can detect TP up to 91 % of the time. As malicious URLs are generated daily, attackers employ tactics to dupe users and change URLs to assail. The web Plug-in is an astonishing method that indicates a warning message to the user and avoids the attack.

REFERENCES

[1] Abdulghani Ali Ahmed, Nik Quosthoni Sunaidi "Malicious Website Detection: A Review" Faculty of Computer Systems & Software Engineering University, Pahang, Malaysia (February 01, 2018)

[2] Moruf akin Adebowale, Khin T.Lwin, M.A.Hossain "Intelligent phishing detection scheme using the deep learning" (convolutional neural network (CNN) and the long short-term memory (LSTM)). (04 Aug 2018)

ISSN (online): 2581-3048
Volume 1, Issue 2, pp 10-14, November-2017

[3] Priya Saravanana, Selvakumar Subramanian. "A Framework for Detecting Phishing Websites using GA-based Feature Selection and ARTMAP based Website Classification".

[4] Ayam el Assael, Shahryar Baki, Avishai Das, Rakesh M Verma "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs."

[5] M. Noushad Rahim and K.P. Mohamed Basheer "A survey on anti-phishing techniques: From conventional methods to machine learning."

[6] Muhammet Baykara, Zahit Ziya Gure, NC "Detection of Phishing Attacks" 2018 6th International Symposium on Digital Forensic and Security (ISDFS) Firat University, Elazig, Turkey.

[7] Athulya A.A, Praveen K. "Towards the Detection of Phishing Attacks" 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)

[8] More akin Adebowale, Khin T.Lwin, M.A.Hossain "Intelligent phishing detection scheme using deep learning" (convolutional neural network (CNN) and the long short-term memory (LSTM)). (04 Aug 2018)

[9] Cassandra cross, Rosalie Gillet "Exploiting Trust for Financial gain: an overview" on Business email Compromise fraud. (22 April 2020)

[10] M somewhat "Efficient Deep Learning Technique for the Detection of Phishing Website"(27 June 2020)

[11] Anu Yadav and Jatin Gemini "The Security threat in Cyber World – cybercrime as PHISHING" p- ISSN: 2393-9907; e-ISSN: 2393-9915 (April-June, 2017).

[12] Phirashisha, Syiemlieh, Golden, Mary Khongsit1, Usha Mary Sharma, Bobby Sharma "Phishing-An Analysis on the Types, Causes, Preventive Measures And Case Studies in the Current Situation" e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 01-08 IOSR Journal of Computer Engineering (IOSR-JCE).

[13] Ram Basnet, Srinivas Mukkamala, and Andrew H. Sung"Detection of Phishing Attacks: A Machine Learning Approach" New Mexico Tech, New Mexico 87801, USA.

****

