IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A SECURE BROWSER FOR DETECTING PHISHING WEBSITES

[1] ENEEYA SRI.D.S, [2] DHARSHINI.M, [3] LATHA JOTHI.V

[1] Student, [2] Student, [3] Professor

[1] Computer Science and Engineering

[1] Velalar College of Engineering and Technology, Thindal, Erode

***ABSTRACT*:** The Internet has widely spread all over the world covering every field of wok. So that the users who depend on the Internet for businesses are also increasing considerably. This research paper is about to detect the phishing websites where user lost their personal information's. We have designed The Anti-Phish browser that check for the secured protocol, if it is not the case it will pop-up an alert message as it is not secured. Whenever the user enter his/her password in the login page of a faked website, then it will halt the navigation of the link and alert the user that you are going to submit your credentials to particular(local host) server. When the user needs to get the original link for the navigated fake link then the user can click the anti-phish button to get the relevant original link. The risk level for each page is indicated in colors (green, yellow, red) within the progress bar that helps the user to be aware of the phishing websites.

KEYWORDS: Internet, Phishing, Navigation, Risk level, link, Anti-Phish.

I. INTRODUCTION

Spoofing or Phishing attacks often take the shape of an email that pretends to be a trusted entity. The e-mail states that the user has to provide information, like master card numbers, identity information, or login credentials, often to correct some alleged problem found with an account. Some number of users falls for these attacks by providing the requested information, which may result in fraudulent charges against credit cards, withdrawals from bank accounts, or other undesirable effects [4]. The primary attempts at applying learning to those problems took the shape of browser toolbars, like the Spoof guard and Net craft toolbars. At some cases toolbars prompts on the user window which may mislead the users with unexpected window. These warnings or dialogs interrupt the normal flow. The user has to wait until they clicks on a link and goes to an internet site to deal with the matter.

The proposed system has allowed the user to be distracted by illegitimate sites, and get notifies when they were misinterpreted or misleaded. The main objective of this paper is to develop a way which will be easily employed by every-one to detect non-legitimate websites in real time. Anti-spoofing services were provided by the Internet service providers, consists of mail servers and clients, web browsers, and available as toolbars for application. However, these services and tools don't effectively protect against all spoofing attacks, as attackers and tool developers are engaged in a continuous race. The proposed system uses methods like HTTPS verification alert, password check, number of dots in URL format, anchor check, warning rate, reverse domain name services and scoring is calculated using associative classification algorithm in data mining for detecting phishing sites.

II. LITERATURE REVIEW

Number of techniques has been proposed by several researches to prevent and detect phishing websites. Here are the existing approaches for anti-phishing,

With the growth of Internet usage in the past years, attackers have the advantage of this to hack the client devices. This can be overcome by the detection of malicious URL using machine learning. The algorithm used in this paper is Random Forest Algorithm in Machine Learning. They took the sample dataset that contains some malicious URL and non-malicious URL. With the help of the machine learning algorithm the entered URL can be predicted as malicious or not. This work is done with the help of a supervised data using Random Forest algorithm that uses labeled data to learn how to classify unlabeled data. The cross-validation carried over the dataset. The method was repeated 3 times and hence the accuracy of the trained model is achieved. The main disadvantage in this paper is the method is repeated 3 times to achieve maximum accuracy, hence it consumes more resources, time and it detects only the URL is malicious or not [11].

Phishing is one of the cybercrime methods which seems to be increasing steadily targeting suspecting users. The one of the solution to this problem is a novel association classification algorithm and it is applied to the well-known phishing websites dataset in the UCI repository. The algorithm used here is Association and Classification in Data Mining. The proposed PWCAC algorithm essentially works in three phases, those are as follows:

- Extracting frequent item-sets
- Classification rules creation
- Predict new unseen examples

The performance of the PWCAC has been compared against several association classification algorithms and rule induction algorithms. The main disadvantage is that this system is suggested only for the well-known phishing websites present in the UCI repository [7].

Phishing is the fraudulent attempt to obtain sensitive information such as usernames, passwords etc., the technique used here is Block-Chain. It classifies the main types and schemes of phishing attacks on the block chain, suggests methods of protection against them. All phishing attacks can be divided into two types: social engineering schemes and technical schemes. Social engineering schemes are based on the deception and subsequent independent wrong actions of the victim, while the technical schemes use vulnerabilities and imperfections of software and infrastructure. The main disadvantage is, it doesn't contain any protocol implementation and it is difficult to classify the types of attack as there are many subtypes under each category [1].

With the growing use of internet across the world, the threats posed by it are numerous. Malicious websites play a pivotal role in affecting your system. A model is developed to forecast a URL is malicious or benign, based on the application layer and network characteristics. Machine learning algorithm for Classification is used. This algorithm is used to develop a classifier using the targeted dataset. The dataset is divided by two categories as the training set and the validation set. These sets are used to train and validate the model. The disadvantage is that confusion matrix is difficult to design and more number of mathematical calculations was carried out to achieve high accuracy [12].

Our proposed system overcomes all those drawbacks discussed in above systems by providing the browser with the indication of risk level in progressing bar which makes the user to be aware of the phishing websites.

III. BROWSER DESIGN

The primary purpose of an online browser is to bring information resources to the user. This process begins when the user inputs a same Resource Identifier, as an example **"**http://en.wikipedia.org/**"**, in browser. The prefix of the URI determines how the URI are going to be interpreted. The foremost commonly used reasonably URI starts with http: and identifies a resource to be retrieved over the Hypertext Transfer Protocol (HTTP).Spoofing or Phishing attacks often take the shape of an email which purports to be from a trusted entity, as like eBay or PayPal.

Newly architected browser introduces special kind of segment to platform spoofing detection operations in recent-times. The browser indicates the user with the alert message, when they are redirected to the undefined websites, it will display the relevant page of the redirected page. We have prototyped this method to make sure maximum security, better accuracy within the identification of phishing websites. Web-pages consists of several properties, those properties can be used to distinguish a legitimate webpage with the phishing one.

IV. METHODOLOGY

This method consists of the following modules that checks for the entered websites,

HTTPS VERIFICATION ALERT

- It is used to check whether the login page of the website has "HTTPS" protocol or not.
- HTTPS is considered as the extension of HTTP protocol.
- HTTPS is encrypted in order to increase security of data transfer.
- This is particularly important when user transmits sensitive data.

SUSPICIOUS LINK CHECK

- It is used to check the address after the @ symbol.
- Every time the page is loaded with the website which is placed after @
  ex: http://google@gmail.com, here the page is loaded with the Gmail.

DOTS IN URL

- It is used to check the dots present in the URL.
- The phishing link is identified with the number of dots present in the websites.
- If there are more than 5 dots then it is considered as there is a risk in that website.

SCORING

- We combine the test results into a total spoof score, TSS, using a standard aggregation function.
- The associative algorithm analyzes all the links present in the loaded website with the criteria set in the modules, and the scoring is calculated.

- The classification algorithm produces the target based on the score calculated.
- Based on the score, the color of the progress bar is decided as green or yellow or red.

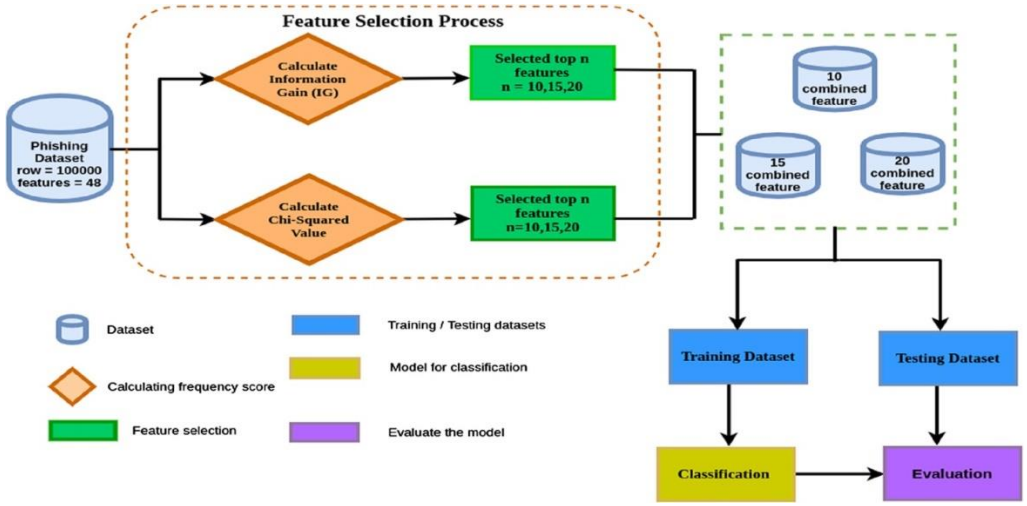

Fig.1 scoring process

V. FLOW DIAGRAM

Fig.2 flow diagram

When the URL is entered the criteria of the modules are verified. The webpage started loading, based on the scoring calculated the user can see the risk level in the progress bar at the top of the browser window as green, yellow and red. Green

indicates that the page is not phishing one. Yellow indicates that phishing risk is tolerable. Red indicates that the page load is a phishing one.

VI. SYSTEM IMPLEMENTATION

Implementation is the stage in the project where the theoretical design is developed into a working system. The most critical stage is achieving a successful system and giving confidence on the new system for the users that it'll work efficiently. It involves careful planning, investing of the present system, and its constraints on its implementation, design of methods to realize the change over, and evaluation of the changeover methods. The implementation process begins with preparing a idea for the implementation of the system. According to the plan, the activities should be applied; discussion has been made regarding the equipment, resources and the way to test the activities.

The coding step converts a design representation as a program design language realization. The coding must have the following characteristics:

- Simple design to code translation
- Code Efficiency
- Memory Efficiency
- Maintainability

VII. COMPARISON CHART

The comparison made against other techniques in terms of efficiency, accuracy, execution time and false positives.

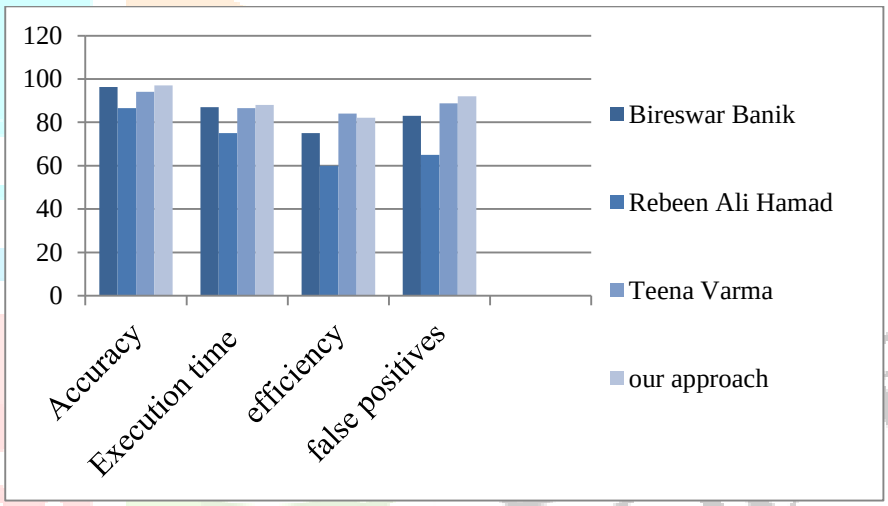

Fig.3 comparison chart

VIII. ADVANTAGES OF PROPOSED SYSTEM

- User is intimated whenever the login page is redirected.
- The user can check the risk level in progress bar for each page they visit with the color indication of green, yellow and red.
- The user can visit the original page of the re-directed websites by clicking on the phishing button on top of the browser.
- The browser is user friendly.

IX. CONCLUSION

In this developed system, it has been proved that it is possible to detect phishing sites with high accuracy by using specialized criteria, such as checking the password, anchor check and HTTPS verification, suspicious link detection, dots present in the URL. A significant number of industrial and academic anti-phishing solutions have been proposed to mitigate phishing attacks. But, all of these solutions have some shortcomings. But the developed system has used several techniques and associative classification algorithm to reduce false positives. The proposed system can catch around 97% phishing sites with about 6% false positives. After combining with some heuristics the system is able to catch about 90% of phishing websites with only 1% of false positives. Hence the developed system is unique and rich in the information regarding and helps the user to trust the sites they visit.

REFERENCES

[1]. Andryukhin, A.A, (2019), "Phishing Attacks Preventions in Block-Chain, International Conference on Engineering Technologies and Computer Science (EnT), Russia".

[2]. Biresware Banik; Abhijit Sarma, 2018, "Phishing URL detection system based on URL features using SVM in ML, Gauhati University".

[3]. Christopher Had Nagy; Michele Fincher, 2015, "Phishing Dark Waters: The Offensive and Defensive Sides of Malicious E-mails".

[4]. Chuan Yue; Haining Wang, 2008, "Anti-Phishing in Offense and defense, Annual Computer Security Applications Conference, IEEE, USA".

[5]. Juan Chen; Chuanxiong Guo, 2006, "Online Detection and Prevention of Phishing Attacks", IEEE (23).

[6]. Lissa Pollacia; Yan Zonh Ding; Seung Yang 2015, "Why Phishing Works: Project for an Information Security Capstone Course, ISEDJ", Vol.13.

[7]. Mohammed Alqahtani, 2019, "Phishing websites Classification with Association Classification, International Conference on Computer and Information Science".

[8]. Pantech Solutions, 2019, "Fake News Detection using Machine Learning".

[9]. Rebeen Ali Hamad; Longzhi Yang; Xuan He; Hao Wang; Bin Gao and Wai Lok Woo, 2019, "A Deep-Learning-Driven Light-Weight Phishing Detection Sensor, Northumbria University, UK".

[10]. Steve Morgan, 2021, "Comprehensive Anti-Phishing Guide [New E-Book Guide]".

[11]. Teena Varma; Pratik Zinjad; Shreeniket Vast; Idris Vohra; Hannan Sunsara, A; 2020, "Malicious URL Detection using ML, IRJET".

[12]. Varaprasada Rao, P; Govinda Rao, S; Chandrasekhar Reddy, P; Anil Kumar, 2019, "Detection of Malicious Uniform Resource Locator, International Journal of Recent Technology and Engineering" Vol.08.

[13]. Weili Han; Jiping Zhou, 2007, "Anti-phishing by Smart Mobile device, IFIT International Conference on Network and Parallel Computing Workshops, IEEE, China".