International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

A Dynamic & Combined Framework for Predicting Phishing Attack

G. Ashwinraj, Sitaraa Krishna Kumar, Devansh Sharma, C. Ambhika

Abstract: ***With the fast improvement of web applications, and with the solace gave by these web applications, web clients' use this advantages, as it were, that they make practically the entirety of their everyday exercises, for example, newspaper reading, shopping, bill payment, ticket booking and entertainment with the assistance of the web. This wonder powers the clients of the web to get associated with the web for a drawn-out time and consequently it builds the odds of the clients to get trapped in the snare of phishing – an assault made by programmers to take delicate data by enticing the clients with rewarding offers at first and afterward diverting them to a fake website(which the client may not presume) where they can mislead the client by requesting that they present their credentials(usually clients present their credentials without realizing that these are phony offers made with a sole aim of taking delicate data). Notwithstanding the caution and mindfulness given by the web community right now, and more phishing craftsman prevail in their attack. Likewise, these phishing craftsmen create novel attacks, for example, tab grabbing, site mirroring and so forth that draws in increasingly more web clients to be trapped in the snare of phishing. Anyway, numerous tools and methodologies have been created to forestall phishing and to caution clients orally and outwardly. Still, the achievement paces of the phishing attack stay high and furthermore the methodologies identified with phishing detection endures high false negative and false positive proportion. In this proposed system numerous mechanisms used to prevent phishing have been analyzed and a proficient system has been proposed to forestall phishing.***

Keywords:Phishing, Mirroring. Tab grabbing.

I. INTRODUCTION

The vast development of the Web is because of the immensenumber of e-services such as informal communication, e-banking, e-business, forums, educational platforms, video sharing destinations, and entertainment. As of late, online business has become an indispensable part of our lives that have given a superior chance to put resources into numerous items through the Internet. The online business empowers clients and organizations to get or publicize data on items and products, and afterward, perform purchasing or selling exchanges through the Internet. Be that as it may, fraud sites might be made and utilized by fraudsters to take delicate data including login credentials and card details. This is termed as phishing attack, which is one of the fundamental security issues that has developed rapidly with an adverse effect on the web trade and businesses.

As of late, numerous tools and methods have been utilized to issue alerts about the anticipated phishing sites in the early phases of an attack. The most generally utilized procedures for phishingidentification use defined blacklist to perceive the phished sites. Inphishing attack, the attackers have numerous tricks to deceive web clients. Moreover, various newsites can be created by the attackers in a couple of seconds. Subsequently, the most well-known techniques are not productive in precisely identifying these new phishing websites, which might be delegated as legitimate websites. In this manner, powerful phishingsites detection requires intelligent and versatile solutions to be built up that can effectively anticipate the newly created phished websites.

The majority of the existing savvy approaches have proposed adopting diverse conventional machine learning classifiers all together to improve phishing website detection. Consequently, the inquiry on which machine learning classifier is the best and prescient in improving the performance of phishing websites detection stays as an immense subject of discussion. Hence, we are propelled to use the extraordinary capability of DNNs so as to precisely identify the new phishing websites. Besides, the regular feature selection methods utilized in the literature to discover the most significant website features are not ready to viably distinguish the most effective website features for all collected data sets. As a matter of fact, the performance of the classification can be improved further by choosing the most persuasive features in the training stage. Moreover, not all the selected features are similarly significant for classification. Subsequently, there is as yet a requirement for additional viable feature determination methods, which can contribute towards improving the accuracy of phishing website detection.

In contrast to the past works, this paper uses wisely the unrivaled performance of the fuzzy rough set algorithm in order to find the most compelling features and the ideal leverage of thesite features to be utilized for improving site phishing detection. Besides, the intense capacity of DNN in accomplishing higher classification precision over the traditional classifiers has been utilized and afterward upgraded in light of the website features chose and weighted by the fuzzy rough set theory to deliver an exceptional precision of the phished site detection.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
G. Ashwinraj*, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
Sitaraa Krishna Kumar, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
Devansh Sharma, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.
Mrs. C. Ambhika, Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India.

II. RELATED WORKS

A. Feature selection and fuzzy classification on diabetes dataset

Vaishali R and Dr. R Sasikala proposes that thedatasets deployed for the analysis have a record of every kind of website whether it's a phished website or a legit website.
In the analysis and selection criteria of the best features, fuzzy classification is chosen.It directly works on the principle of maximum classifier rate and minimum rules.As the Phishing recognization is the key component of this paper.It can be done by training the benchmarked datasets.With the help of a fuzzy rough set,the fetched features then can be passed to classifiers[4].

B. 3-D Feature selection to perform lip reading

Sunil SudamMorade andSuprava Patnaik proposes that there is a huge volume of data that needs to be processed.In order to maintain the time complexity and increase the simplicity of the working of the system,the Genetic algorithm reduces the complexity in data analysis and to improve the execution of the classifiers at a low cost of computation.Also both testing and training time for the classifier is reduced by compact feature size[1].

C. Android Malware Detection based on machine learning

Anam Fatima, ,RiteshMaurya, Malay Kishore Dutta, RadimBurget and Jan Masek proposes that in order to select appropriate features, a Genetic algorithm is used as a search method.Selected features from the genetic algorithm aretapped to train machine learning classifiers.The capability in the identification of malware or suspicious elements before and after selection is compared. Achieving detection accuracy is very high as 96 percent. The genetic algorithm gives optimized results and helps in reducing errors[2].

D. Application of GA Based on F-Ratio Rule

Ting An proposes that Feature selection during initial stages carries essential volume for the accuracy of the results.With F-Ratio, the Genetic algorithm which is used for feature selectionuses adaptive genetic operators to improve the algorithm.The experimental results of GA with the application of the F-Ratio rule can select the best and more accurate features.The F-Ratio rule here improves the overall efficiency of the recognition process[3].

E. Feature selection for effective intrusion detection system

In the paper proposed by Mr. Ketan Sanjay Desale and Ms. Roshani Ade, the Intrusion detection system is introduced which identifies malicious activity on the network.With this system,Only selected features are used to build models.The feature selection approach reduces the time complexity and overall efficiency of the system.There is an immense growth in overallaccuracy[5].

III. PROPOSED SYSTEM

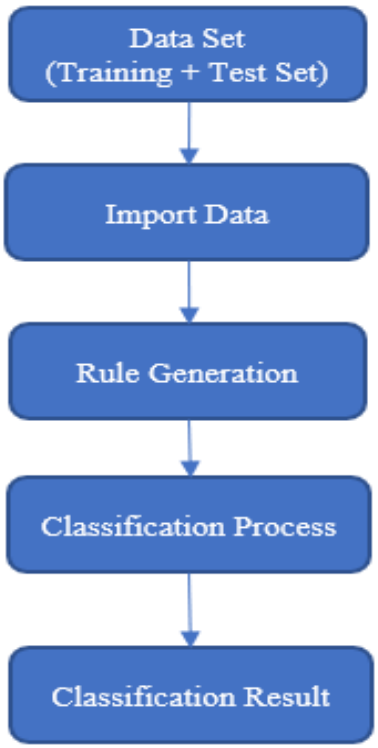

Fig. 1 FRS Classifier Workflow

The proposed framework applies Fuzzy Rough Set (FRS) hypothesis as an apparatus to choose the best features from three benchmarked data sets. The chosen features are nourished into three regularly utilized classifiers for phishing recognition. A multilayer perceptron, a random forest, and sequential minimal optimization SMO.
Also, this framework trains every classifier on a different out-of-test dataset to look at the capacity of the FRS feature selection in building up a generalizable phishing recognition. This preparation set which is made by arbitrary extraction out of sites from an online repository. All hyper-parameters are additionally set dependent on the values utilized in these past works. To assess the capacity for FRS feature selection to build up a generalizable phishing detector, we train every classifier on a different out-of-test dataset made by arbitrary extraction out of phishing and legit sites from an online repository.
The proposed system reduces the false detection of the phishing detector. The idea behind this approach is to use the web spider which will extract the name of the domain from Victim's URL. Then the page rank of the crawled domain name and the existing domain will be compared. If it results in a noticable rank difference, then the site will be reported as a phished site or if there is no noticable difference then it is reported as a legit site.This methodolgy dependent on the page rank can efficiently distinguish phished sites on the grounds that the phished sites just live for a short timeframe.

Advantages of the Proposed System

- Maximize the classification performance
- Higher classification sensitivity
- Increases the accuracy of the prediction
- Removes the irrelevant and redundant features
- High accuracy in complex application

IV. SYSTEM ARCHITECTURE

Fig. 1. System Architecture

A phishing attack is at first directed through browsing or getting a link inside an email, which seems, by all accounts to be from a real source, approaching clients for their credentials. By tapping the link,the clients will be redirected to someother site that could be a phished site or a legit one.

Firstly, the data sets that are classified as orginal phished and legit is collected and the raw data must be correctly trained with machine learning. Then the feature extraction aims to extract the well-known site features, which will help and add to the decision of the authenticity of the site, either phished or legitimate. A few late studies on phishing site detection has been completed to investigate and distinguish the pertinent and critical features that can be tapped to differentiate legitimate from phishing websites.

Web content analyzer gives a total examination that will help in measuring the quality of the web site. It basically analyses the URL and Meta descriptions. Also, it can view the hyperlinks that are embedded within the website and the quality of those links too. If at all there are any spams or any broken links, then those will be reported and will be logged.

Once the web content is analyzed based on the analysis, the page rank is calculated. Page rank is a link analyzing algorithm and it relegates a numerical weight to each element of the linked set of documents in order to measure its importance. For instance, if a page is connected to by numerous pages then the page rank associated with that website will be high. The page rank shows the significance of a specific page.

Lastly, after calculating the page rank, the calculated page rank of a specific website and page rank of the actual website will be compared in order for the purpose of classification. If there is a notable difference in terms of the rank between the collected website and the actual website, then the website is reported to be phished and if there are no significant rank differences between the websites, then it is reported as a legitimate website.

V. RESULT

The reliability of the proposed framework lies in preventing users from being victims of phishing. Clearly, we can't guarantee that each web clients are specialists in web security and henceforth a tool to detect phishing is required to protect the web clients. The rundown based and heuristic-based methodologies can detect phishing but there is a bogus caution rate associated with these methodologies which are unsuitable. Then again the Machine-learning and multi-level characterization approaches improve the authenticity of the prediction, yet it fails to detect the image-based phishing content. Likewise, when on one side, the web client endures privacy and financial misfortune, because of unintended exposure of delicate data, on the opposite side the phishing sites generally mirror the site of a well-reputed association or an organization and consequently they will lose important clients and also the economic status. Nevertheless, only some have taken the required steps to contribute to the detection of phishing. As of late, a progressed form of phishing based on flash content, which bypasses the techniques with which detection occurs. So to fix these we propose a framework that can identify the phished content that has embedded objects and helps in target disclosure. The proposed phishing detection framework will be designed in a way that it will have an ideal exchange off between time uncertainty and detection precision.

VI. CONCLUSION

This study suggested a dynamic and combined framework for the prediction of phishing attacksdependent on deep neural networks with fuzzy rough set based classifier for the objective of feature selection. In the proposed system, the most notable features were selected to increase the prediction accuracy of the phished sites. Appropriately, the features of the website are selected and tapped to train the deep neural networks in order to improve the prediction of phished sites. Therefore, the proposed Dynamic and combined framework for anticipating phishing attacks can be used as a solution to persuasively predict phished sites.

VII. FUTURE WORK

In spite of the fact that the proposed system can detect phishing attacks, still, there are some limitations associated with it. The data set for training purposes is fetched from an online repository which must be appropriately done in order to accomplish the detection of phishing attacks. So, in the future, we can implement a system so that it can also verify the authenticity of the data that is being retrieved from the online repository. This will help to accomplish outstanding detection of phishing attacks

REFERENCES

1. Sunil SudamMorade, Suprava Patnaik," A Genetic Algorithm-Based 3D Feature Selection for Lip Reading", 2015 International Conference on Pervasive Computing, Pune, India.
2. Anam Fatima, ,RiteshMaurya, Malay Kishore Dutta, RadimBurget and Jan Masek," Android Malware Detection Using Genetic Algorithm based Optimized Feature Selection and Machine Learning", 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary.
3. Ting An," Application of Genetic Algorithm Based on F-Ratio Rule in Signal Feature Selection", 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China.
4. Vaishali R, Dr. R Sasikala, "Genetic algorithm based feature selection and MOE fuzzy classification algorithm on Pima Indians Diabetes dataset", 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria.
5. Mr. Ketan Sanjay Desale, Ms. Roshani Ade,"Genetic Algorithm based Feature Selection Approach for Effective Intrusion Detection System", 2015 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India.
6. Anti-Phishing Working Group. Accessed: Sep. 2016. [Online]. Available http://www.antiphishing.org
7. Z. Dong, A. Kapadia, J. Blythe and L. J. Camp "Beyond the Lock Icon: Real-time Detection of Phishing Websites Using Public Key Certicates" , 2015 APWG Symposium on Electronic Crime Research (eCrime), Barcelona, 2015, pp. 1-12.
8. Naga Venkata Sunil and A. Sardana, "A PageRank based detection technique for phishing web sites," 2012 IEEE Symposium on Computers & Informatics (ISCI), Penang, 2012, pp. 58-63. doi: 10.1109/ISCI.2012.6222667
9. R. Dhamija and J.D. Tygar, "The Battle against Phishing: Dynamic SecuritySkins", Proc. Symp. Usable Privacy and Security, 2005,pp 77-88. Mobile Marketing Statistics. Accessed: Mar. 2017.
10. Chandrashekar, G., Sahin, F.: 'A survey on feature selection methods', Comput. Electr. Eng., 2014, 40, (1), pp. 16–28
11. Waleed Ali and Adel A. Ahmed," Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting", IET Information Security Vol:13 Issue:6

AUTHORS PROFILE

G Ashwinraj,Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, Research Interests on Cyber Security, Cloud Computing and Network Security.

SitaraaKrishnakumar, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India,Research Interests on Artificial intelligence, Cyber Security and Network Security.

Devansh Sharma, Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, Research Interests on Data Analytics, Cyber Security and Data Science.

C Ambhika, Faculty, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India, Research Interests on Computer Networks and Machine Learning.

.