|| Volume 6 || Issue 7 || July 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

ILLEGITIMATE WEBSITES DETECTION USING DEEP LEARNING FRAMEWORK

Miss. Rani Shinde[1], Dr. Gitanjali Shinde[2]

Department of Computer Engineering
Student, Smt. Kashibai Navale College of Engineering, Vadgaon, Pune[1]
Professor, Smt. Kashibai Navale College of Engineering, Vadgaon, Pune[2]

Abstract: Phishing is a crime involving robbery of confidential user data. The phishing websites are aimed at individuals, businesses, and cloud storage and government websites. Hardware- based anti-phishing methods are generally used, but software- based approaches are favored because of costs and operational factors. There is no solution to the problem such as zero-day phishing attacks from current phishing detection approaches. A three-phase attack detection called the Phishing Attack Detector based on Web Crawler was proposed to resolve these problems and precisely detect phishing incidences using recurrent neural network. It includes the input features Web traffic, web content and Uniform Resource Locator (URL) based on the classification of phishing and non-phishing pages.

Keywords: —*Recurrent Neural Network, Deep Learning, illegitimate URLs, cyberattacks*

I INTRODUCTION

Phishing is a cyber-crime where a person who poses as a legitimate agency contacts a victim or target via email, phone or text message to attract the person to supply information, information about personal identity, banking and credit card information and passwords. Phishing is a crime. The new term 'fishing' refers to the attacker's invitation to visit a counterfeit site by creating a website look, and to get personal information from users such as username, password, financial information, account details, national security identifier, etc.. Phishing is a new term that was developed using 'fishing.' The information collected is used for potential target ads or even identity robberies, attacks (for example, money transfer from one's account). The attack method that is widely used is to send e-mails, messages that can lead to data theft or personal information. Social networking account Passwords, credit cards or attackers provide upgrades to their websites, encourage you to comply with your personal information and change it via fake website, are mis-entered daily. If you are entering your personal data, the attackers will collect it successfully on your server side, and will be able to carry out the next move with your information and to use it for their malicious purposes.

Phishing is described as a reverberation of a website of a remarkable business that snaps private data of consumers, for example usernames, passwords and structured savings numbers. Mail spammers can be categorized with their target in mind. Some telemarketers are spammers who send a few hundred/a large number of e-mail customer's spontaneous messages. Spammers have the following classification, which continues to randomly send messages, but are near zero enthusiastic. Often they spam or promote materials with irrelevant topics. Some of the cases are sees, knowledgeable news, or statements about meetings. Phishing is itself a new idea, but the criminals, i.e. the phishers, have more and more used it in recent years to steal your personal data and carry out business and social crimes. The number of phishing attacks has significantly risen in four to five years. Phishing is widely used and is easy to carry out on your destination. Phishing usually uses social engineering to attract a victim by submitting a spoofed link to a fake website. The spoofed connection can be found on common web pages or sent to the victim via email. Similar to the legitimate website the fake website is made. So it is directed to the attacker site instead of guiding the victim request to the true web server.

A. Motivation

The fundamental principle behind the development of such a system is to ensure that financial information for a customer is safe, and so banks and other financial institutions provide various security measures to minimize the risk of unauthorized access to their online account. Online banking has been completely relayed on online transactions through various applications nowadays, so it is most important that this online banking activity is secured.

B. Contribution

•Aim is to develop application for peoples who make our nation more digital and scam free through an online banking.

•The objective of the proposed system is to provide best possible security mechanism to provide confidence to the people make most of transaction online.

•The objective behind this system is to invent a system widely acceptable for providing vital role in security concern for banking era.

•We have to provide perfect approach for online banking with the help of anomaly based detection and prevention of phishing attacks.

II.REVIEW OF LITERATURE

In this paper [1], we did a comprehensive study on the security vulnerabilities caused by mobile phishing attacks, including the web page phishing attacks. Author propose MobiFish, a novel automated lightweight anti-phishing scheme for mobile platforms. MobiFish verifies the validity of web pages, applications, and persistent accounts by comparing the actual Identity to the claimed identity.Existing schemes designed for web phishing attacks on PCs cannot effectively address the various phishing attacks on mobile devices.

To [2] an online user into elicit personal Information. The prime objective of this review is to do literature survey on social engineering attack: Phishing attack and techniques to detect attack. The paper discusses various types of Phishing attacks such as Tab-napping, spoofing emails, Trojan horse, hacking and how to prevent them. Every organization has security issues that have been of great concern to users, site developers, and specialists, in order to defend the confidential data from this type of social engineering attack.

Commercial and retail account [3] holders at financial institutions of all sizes are under attack 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 activity. Implementing anomaly detection will not only meet FFIEC Expectations, it will decrease the total cost of fraud, and will increase customer loyalty and trust.

This paper [4] 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 I will explain the concepts and technology behind phishing, show how the threat is much more than just a nuisance or passing trend, and discuss how gangs of criminals are Using these scams to make a great deal of money. Unfortunately, a growing number of cyber-thieves are using these same systems to manipulate us and steal our private information.

Author suggest in this paper[6] a technique called optimum RT-PFL for classifying malicious URLs detected on the websites from non-malicious URLs. In order to generate feature components, the data set should both be encoded as lexical and host functions for the URL. The function extraction method extracts those features. Optimum URL

Functions are chosen according to the proposed selection process, namely the Rough Set Theory algorithm based on Gray Wolf Optimizer. This proposed algorithm will define a minimal reduction in the attributes from the highly effective data collection, which in turn enhances the efficiency of classification systems. In order to decide whether the approved URL is good or malicious, the URL should be inserted into the classifier. The classification of URLs depends on the newly formulated fuzzy logical approach to particle filtering. The following categories are strengthened with the detection of a large number of suspicious URLs from malicious pages.

This paper [7] provides a detailed empirical analysis on 1529,433 malicious URLs in the last two years. Author evaluate tactical actions of attackers with respect to URLs and extract common capabilities. Author then divide it into three usable pools, so that the compromise levels of unknown URLs are calculated. Author use a similarity matching technique to leverage detection speeds. Author assume that the attackers' normal URL manipulation behaviors will classify new URLs. This method covers a wide range of malicious URLs with limited function sets. The exactness of the proposed method is rational (up to 70 percent) and the approach requires only analysis of the attributes of URLs. During preprocessing this model can be used to assess if input URLs are friendly or to estimate if an input URL is malicious as a web filter or a risk scaler.

This paper's [8] objective is twice. First, author will talk in depth about the history of phishing attacks and the motivation of attackers. Then, the different forms of phishing attacks are taxonomied. Second, to protect users from phishing based on the attacks found in our fiscalonomics, our services will provide taxonomies of many solutions suggested in the literature. In addition, we addressed the effects of Internet of Things phishing attacks (IoTs). We conclude our paper on several still existing literary issues and challenges that are relevant for the fight against phishing threats.

In this paper [9], author suggest a new method for protecting against phishing attacks by automatically updating the white list of legit sites visited by the user. Our solution proposed has high detection and short access time. The browser warns users not to reveal personal details when they attempt to open a page that is not available in the white list. In addition, we verify the validity of a website with hyperlinks. This is done by extracting hyperlinks from your website source code and using the proposed phishing detection algorithm. Our experimental results show the proposed solution to phishing as it has a true positive rate of 86.02%, whereas a false negative level of less than 1.48% is very successful.

III.PROPOSED METHODOLOGY

The basic concept behind develop such system is to providing security to a customer's financial information is vital and therefore banks and other financial institutes offer different security mechanisms to reduce the risk of unauthorized access to their online customer accounts.

Now days online banking era has been fully relay on online transaction through different application gateway, So its most needed that the provide security to these online banking activities.

A. Architecture

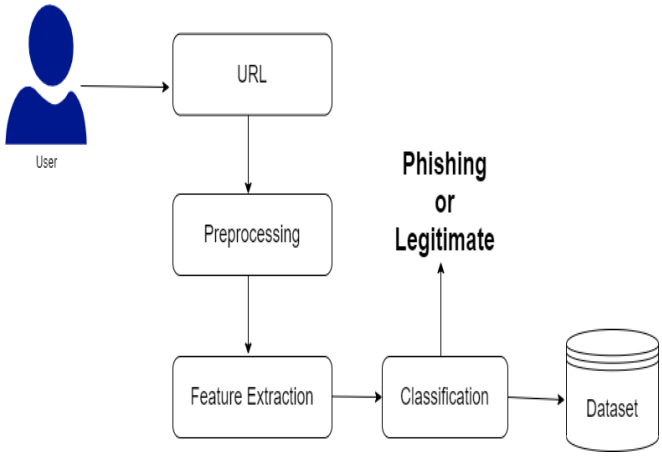

Fig. 1. Proposed System Architecture

A. Algorithm

Recurrent Neural Network(RNN)

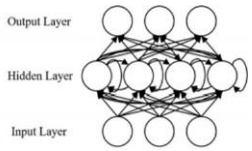

Fig. 2. Recurrent Neural Network

As shown in Fig. , for a RNN, let our input x be a sequence whose length is T, x = x1 , x2 ,...,xT , and each item xt is a feature vector. At time step t, given the previous hidden layer state ht1 , the current hidden layer state ht and the output layer state yt can be calculated by,

$$h_t = o_h(w_h x_t + U_h h_t - 1 + b_h)$$

$$y_t = o_y(w_y h_t + b_y)$$

where Wh and Wy denote the input-to-hidden and hidden-to-

output weight matrices, respectively, Uh is the matrix of the recurrent weights between the hidden layer and itself at two adjacent time steps, bh and by are the biases, and h and y denote the activation functions.

At each time step, the input is propagated in a standard feed forward fashion, and then, a learning rule is applied. The back connections lead to the result that the context units always maintain a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). Thus, the network can maintain a state, allowing it to perform such tasks as sequence prediction that are beyond the power of standard multilayer perception.

Formula for calculating current state:

$$ht = \int (h_t - 1, x_t)$$

where,

ht=current state

ht-1=Previous state xt= Input state

Formula for applying Activation function:

$$h_t = activation(w_h h h_t - 1 + w_x h x_t)$$

where,

whh= Weight at recurrent neuron wxh= Weight at input neuron

Formula for calculating output:

$$y_t = w_h y h_t$$

where,

yt=Output

why=Weight at output layer

IV.RESULTS AND DISCUSSION

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

Let TP be the number of correctly classified phishing pages, TN be the number of correctly classified legitimate pages, FP be the number of wrongly classified legitimate pages and FN be the number of wrongly classified phishing pages. The performance metrics used in our approach are:

Accuracy = TP + TN/FP + TN + FN + TP

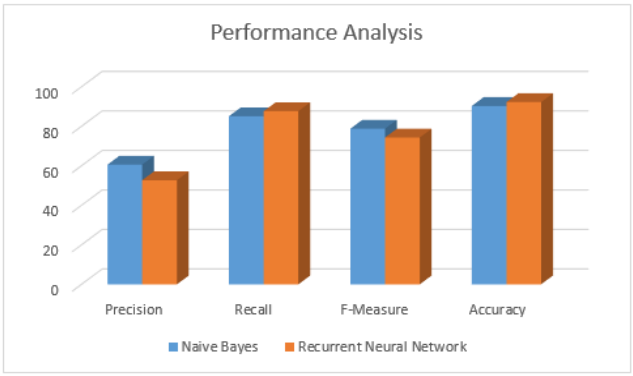

Fig. 3. Classification Results

	Naive Bayes	Recurrent Neural Network
Precision	60.6	52.70
Recall	85.1	87.64
F-Measure	78.8	74.31
Accuracy	90.29	92.26

Table 1: Comparative Result

V.CONCLUSION

Phishing is one of the most damaging web security threats. We have created a prediction model for the detection of Phish- ing websites by analysing the attributes of the attack according to our study. The deep-seated learning model of the Deep recurrent neural Network overcomes other machine learning models via prediction and achieves the highest precision.

REFERENCES

[1]Surbhi Gupta et al., "A Literature Survey on Social Engineering Attacks: Phishing Attack," in International Conference on Computing, Communi- cation and Automation (ICCCA2016), 2017, pp. 537-540.

[2]Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, "Phishing- Alarm: Robust and Efficient Phishing Detection via Page Component Similarity".

[3]Zou Futai, Gang Yuxiang, Pei Bei, Pan Li, Li Linsen, "Web Phishing De- tection Based on Graph Mining", Guardian Analytics,"A Practical Guide to Anomaly Detection Implications of meeting new FFIEC minimum expectations for layered security". Accessed: 08 Jan 2018.

[4]Ibrahim Waziri Jr., "Website Forgery: Understanding Phishing Attacks Nontechnical Countermeasures," in IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015,IEEE.

[5]LongfeiWu et al,"Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms," IEEE 2016, pp. 6678-6691.

[6]K. Rajitha and D. Vijayalakshmi, "Suspicious urls filtering using optimal rt-pfl: A novel feature selection based web url detection," in Smart Computing and Informatics, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer Singapore, 2018, pp. 227–235.

[7]S. Kim, J. Kim, and B. B. Kang, "Malicious url protection based on attackers' habitual behavioral analysis," Computers Security, vol. 77, pp. 790 – 806, 2018.

[8]B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," Telecommunication Systems, vol. 67, no. 2, pp. 247–267, Feb 2018.

[9]A. K. Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list," EURASIP Journal on Information Security, vol. 2016, no. 1, p. 9, May 2016.