IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

# Phishing websites Detection Using Deep Learning

**Mr.Harish Kadam [1], Prof. Sneha Tirth [2]**

P.G. Student, Department of Computer Engineering, Trinity College of Engineering and Research Pune, India [1]

Associate Professor, Department of Computer Engineering, Trinity College of Engineering and Research Pune, India [2]

**ABSTRACT**: Phishing is a crime that involves the theft of confidential user information. Those targeted by phishing websites include individuals, small businesses, cloud storage providers, and government organisations and websites. The majority of phishing prevention techniques involve hardware-based solutions, although software-based options are preferred due of cost and operational considerations. There is no answer to the problem of zero-day phishing assaults from the present phishing detection approaches since there is no solution to the problem. The Phishing Attack Detector based on Web Crawler, a three-phase attack detection system, was designed to handle these issues and accurately detect phishing incidences using a recurrent neural network in order to resolve these issues. It covers the input aspects of web traffic, web content, and Uniform Resource Locator (URL) based on the classification of phishing and non-phishing pages, as well as the output features of phishing and non-phishing pages.

**KEYWORDS**: Deep Learning, URL dataset, Phishing URLs, security, cyberattacks Text detection.

## I. INTRODUCTION

Phishing is a type of cybercrime in which a person impersonating a legitimate agency contacts a victim or target via email, phone, or text message in order to entice the person to provide information, such as personal identification information, banking and credit card information, and passwords, among other things. Phishing is a serious offence. When an attacker creates the appearance of a legitimate website, he or she is inviting visitors to visit the site in order to obtain personal information from them, such as their login, password, financial information, account details, national security identifier, and so on. Phishing is a new term that was coined by combining the words 'fishing' and 'phishing.' In some cases, the information gathered is used for possible target advertisements or even identity robberies and attacks (such as money transfers from one's bank account). The most often utilised attack method is to send e-mails, messages that might lead to data theft or the theft of personal information from the recipient. Account on a social networking site Daily, people make mistakes while entering passwords, credit cards, or other sensitive information onto websites. Attackers use bogus websites to provide upgrades to their websites, entice you to comply with your personal information, and then change your information. As long as you are submitting your personal information, the attackers will be able to successfully collect it on your server side, and they will be able to carry out the following move with your information and utilise it to further their nefarious objectives.

Phishing is defined as the replication of a legitimate business website in order to steal private information from customers, such as usernames, passwords, and structured savings numbers. Phishing is defined as the theft of private information from customers. Mail spammers can be classified based on who they are trying to reach. The majority of telemarketers are spammers, who send a few hundred or a big number of e-mail messages to customers who have sent them spontaneously. Spammers are divided into the following categories, each of which continues to send messages at random intervals, albeit with little enthusiasm. Frequently, they spam or push items that are about completely unrelated themes. Some of the examples include sightings, authoritative reports, and comments regarding meetings. Phishing is not a new concept, but fraudsters, also known as phishers, have been increasingly adept at using it in recent years to steal your personal information and commit economic and social crimes in the process. In the last four to five years, there has been a dramatic increase in the frequency of phishing assaults. Phishing is a common practise that is simple to carry out once you get at your location. Phishing is a method of attracting a victim by submitting a faked link to a bogus website, which is known as social engineering. It is possible to find the faked connection on often visited online pages or to send it to the victim via email. A false website is created in the same manner as the actual website. The victim's request is thereby routed directly to the attacker's website rather than to the legitimate web server.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

## II. REVIEW OF LITERATURE

In this paper [1], authors 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

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

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.

A. Gap Analysis

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 minimise the risk of unauthorised 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.

## 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

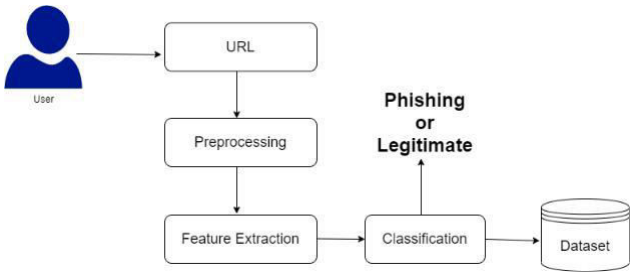


Fig. 1. Proposed System Architecture

When Costumer first time login to System that time all details IP address, current device, OS, time, location stored user log files at bank server for future anomaly detection.
Next time when user want to login that time these details compared with users current details if matched then access granted otherwise security Questions asked to user if ok with this then n only then access granted.
User login to the system, after login user gets authen-tication permission to access or view system. System is responsible for giving access to user by simply match-ing anomalies if matched then ok otherwise check for identity.
We have to provide perfect approach for online banking with the help of anomaly based detection and prevention of phishing attacks.

B. Algorithm

Recurrent Neural Network(RNN)

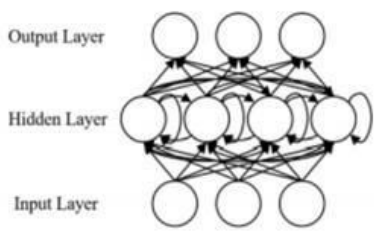


Fig. 2. Recurrent Neural Network

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

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,

$$ht = oh(whxt + Uhht \quad 1 + bh)$$

$$yt = oy(wyht + by)$$

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:

$$ht = \quad Z \quad (ht \quad 1; xt)$$

where,

ht=current state

ht-1=Previous state

xt= Input state

Formula for applying Activation function:

$$ht = activation(whhht \quad 1 + wxhxt)$$

where,

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

Formula for calculating output:

$$yt = whyht$$

where,

yt=Output

why=Weight at output layer

Exemplar based Inpainting technique is used for inpainting of text regions, which takes structure synthesis and texture

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

## 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:

TPR is the rate of correctly classified phishing pages. It is calculated as:

$$TPR = TP=TP + FN$$

TNR is the rate of correctly classified legitimate pages. It is calculated as:

$$TNR = TNR=FP + TN$$

FPR is the rate of wrongly classified phishing pages. It is calculated as:

$$FPR = FP=FP + TN$$

FNR is the rate of wrongly classified legitimate pages. It is calculated as:

$$FNR = FN=FN + TP$$

Accuracy is the rate of correctly classified web pages. It is calculated as:

$$accuracy = T\,P + T\,N=F\,P + T\,N + F\,N + T\,P$$

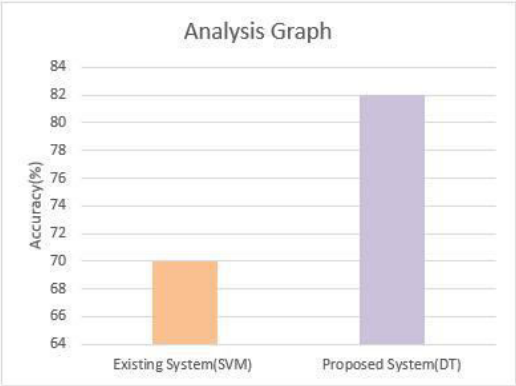


Fig. 3. Graph

| Sr. No. | Existing Sys-<br>Tem | Proposed<br>System |
|---|---|---|
| 1 | 70% | 82% |

Table 1:Comparative Result

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106090|

## 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.
[10] F. D. Abdi and L. Wenjuan, "Malicious url detection using convolutional neural network," Journal International Journal of Computer Science, Engineering and Information Technology, vol. 7, no. 6, pp. 1–8, 2017.
[11] E. Buber, B. Dırı, and O. K. Sahingoz, "Detecting phishing attacks from url by using nlp techniques," in 2017 International Conference on Computer Science and Engineering (UBMK), Oct 2017, pp. 337–342.
[12] T. Shibahara, K. Yamanishi, Y. Takata, D. Chiba, M. Akiyama, T. Yagi, Y. Ohsita, and M. Murata, "Malicious url sequence detection using event de-noising convolutional neural network," in 2017 IEEE International Conference on Communications (ICC). IEEE, 2017, pp. 1–7.
[13] M. Babagoli, M. P. Aghababa, and V. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," Soft Computing, Feb 2018.
[14] T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in 2018 IEEE 12th International Conference on Semantic Computing (ICSC), Jan 2018, pp. 300–301.
[15] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," Expert Systems with Applications, vol. 117, pp. 345–357, 2019.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



Scan to save the contact details