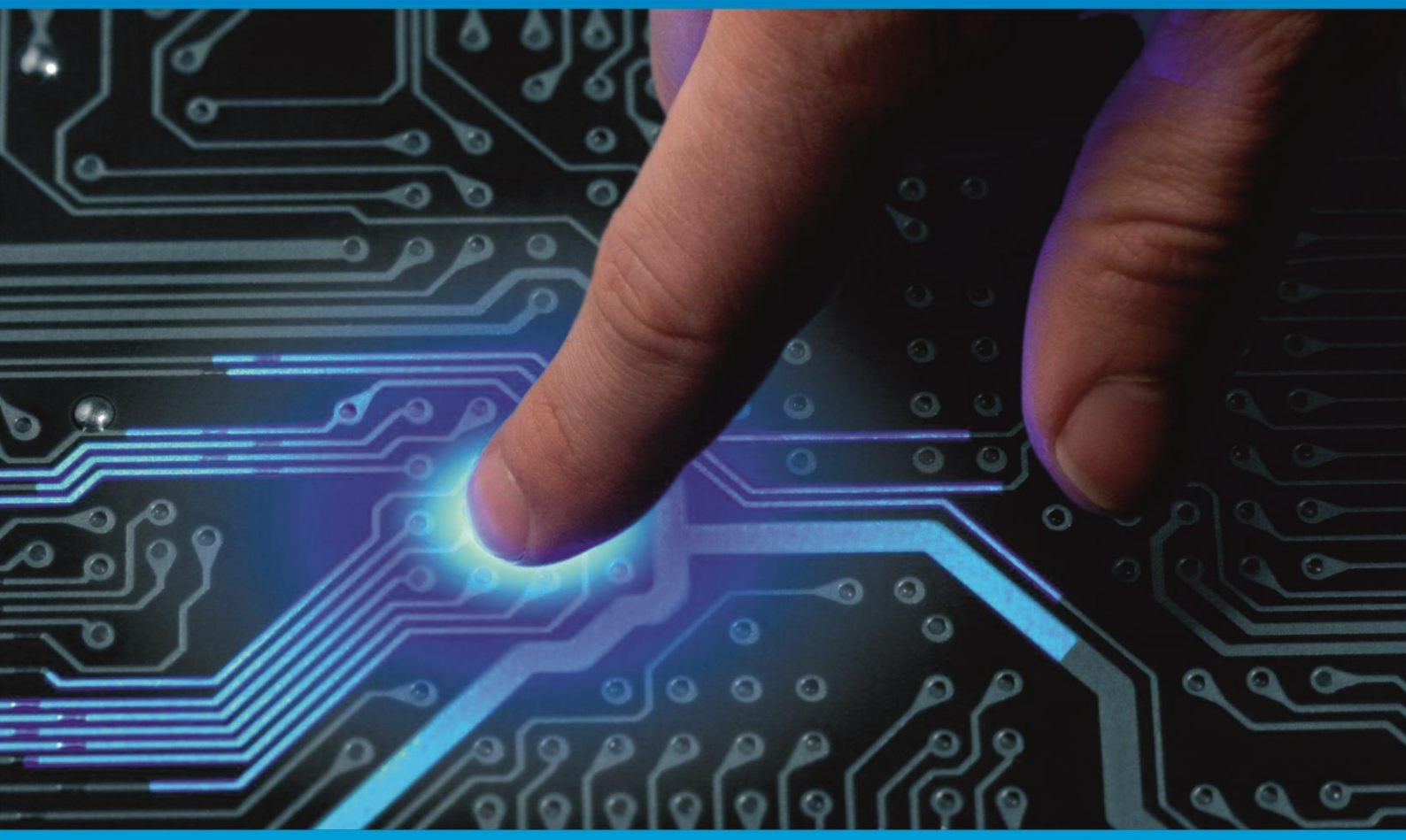
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 8, Issue 10, October 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

Phishing Detection and Blocking System (PDBS)

Jisha K J, Sini Thomas

Department of Computer Science, Christ College (Autonomous), Irinjalakuda, Kerala, India

Assistant Professor, Department of Computer Science, Christ College (Autonomous), Irinjalakuda, Kerala, India

ABSTRACT: Phishing website is one of the internet security problems that target the human vulnerabilities rather than software vulnerabilities. It can be described as the process of attracting online users to obtain their sensitive information such as usernames and passwords. In this paper, we offer an intelligent system for detecting and blocking phishing websites. The system acts as an additional functionality to an internet browser as an extension that automatically notifies the user when it detects a phishing website and blocking user to use phishing websites. The system is based on a deep learning method, particularly supervised learning. The system relies on the URL of the websites. It does not need to retrieve content of the target websites. We have selected the Bidirectional LSTM (Long Short Term Memory) due to its good performance in classification. Our focus is to pursue a higher performance classifier by studying the features of phishing website and choose the better combination of them to train the classifier.

KEYWORDS: Deep Learning, Bidirectional LSTM, URL, Phishing features, Phishing websites, Browser extension

I. INTRODUCTION

One of the most dangerous attacks in the today's internet trends are happening in the form of phishing sites. The major attacks are done to retrieve the personal information of the users from the banking sectors. Phishing is the act of acquiring electronic information such as Username, Password, and Credit Cards Information by masquerading as trustworthy authority. This information may be used to retrieve some information by logging into the system with these username and password or performing some transaction with the use of username, password and credit card information retrieved from this phishing. Phishing can be of many types but nowadays the very usual way of phishing is through the E-Mail or creating the websites of brands (like ICICI Bank, SBI Bank, www.faceboook.com,etc.) which looks very alike with their legitimate sites and asking users to enter their username password or any such personal information. Phishing sites are the major attacks by which most of internet users are being fooled by the phisher. The replicas of the legitimate sites are created and users are directed to that web siteby luring some offers to it. There are certain standards which are given by W3C (World Wide Web Consortium), based on these standards we are choosing some features which can easily describe the difference between legit site and phish site. Phisher is the community of hackers which creates the replicas of the legitimate web sites to retrieve user's personal information such as passwords, credit card number, and financial transaction information. As per the survey done by RSA fraud Surveyor, the Phishing attacks have been raised by 2\% since the last December 2012 to January 2013. The W3C has set some standards' that are followed by most of the legit sites but a phisher may not care to follow these standards as this site is intended to catch many fish in very small amount of time. There are certain characteristics f the URLs and source code of the Phishing site based on which we can guess the site is fake or not.

II. RELATED WORKS

There have been different studies conducted on phishing detection.
Jeeva and Rajsingh presented a system for prediction phishing URLs by generating rules of association rule mining. They used the Apriori algorithm to pick known information from frequent item set properties that were extracted from the dataset. Jeeva and Rajsingh also used another algorithm that performs on hidden data to obtain the accuracy of association rules, which is a predictive Apriori that engages the confidence and the support techniques that are measured in its accuracy, unlike a priori, which only mark rules that have the confidence technique. As a result, they presented significant features of the URL that distinguish if it is phishing or legitimate.

III. PROPOSED SYSTEM

It is an intelligent system for detecting and blocking phishing websites. The system acts as an additional functionality to an internet browser as an extension that automatically notifies the user when it detects a phishing website and blocking user to use phishing websites. The system is based on a deep learning method, particularly supervised learning. The system relies on the URL of the websites. It does not need to retrieve content of the target websites. Extract the features of phishing website URL through deep learning model. We develop a GUI, for detecting phishing websites and blocking user to use phishing websites and automatically display a pop up message when it detects a phishing websites. Do not let user, to enter into that websites. We have selected the Bidirectional LSTM (Long Short Term Memory) due to its good performance in classification. Our focus is to pursue a higher performance classifier by studying the features of phishing website and choose the better combination of them to train the classifier.

Working Diagram

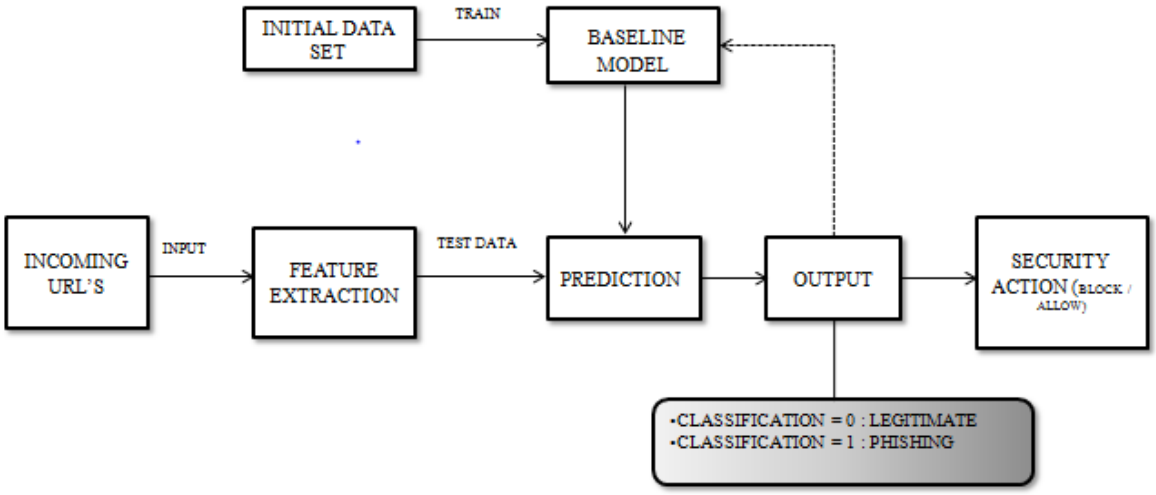


FIGURE 1: BLOCK DIAGRAM

IV. METHOD

We study all features to indicate the strongest, weakestand to remove the irrelevant features; the study is based on examining all possible combination of features. The main function of the system is to decide the state of the website if it is a phishing or legitimate site and then block that site from user. The datasets which we will be used for the initial entries of the training of the system is assumed to be the correct input for the system. The URL which is selected as the phishing URL must be the originally declared as phishing URL and vice-versa. The training dataset is taken from the online repositories like www.phishtank.com from where the known valid phish URL can be retrieved and some legitimate URL directly taken from Google search tool. The algorithm will be triggered whenever the user enters a new website; the role of the algorithm is to extract the features of the website using URL. Dots in URL, Slashes in URL, Length of URL, Suspicious characters in URL such features are extracted from dataset. Deep learning Bidirectional LSTM algorithm is used and classify phishing sites as 0 (zero) and legitimate site as 1. Suspicious URLs are blocked using GUI. Since this an phishing detection system, the user does not directly interact with the browser.

Rather, we develop a GUI, for detecting phishing websites and blocking user to use phishing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

Websites and automatically display a pop up message when it detects a phishing websites. The system acts as an additional functionality to an internet. It acts as an interface between the userand the browser. We use tkinter to develop GUI. It is the standard python interface to the tk guitool kit.

A. BIDIRECTIONAL LSTM

BiLSTM means bidirectional LSTM, which means the signal propagates backward as well asforward in time. Bidirectional LSTMs are an extension of traditional LSTMs that can improve model performance on sequence classification problems. In problems where all time steps of theinput sequence are available, Bidirectional LSTMs train two instead of one LSTMs on the input sequence. The first on the input sequence as-is and the second on a reversed copy of the inputsequence. This can provide additional context to the network and result in faster and even fullerlearning on the problem. Bidirectional LSTMs are supported in Keras via the Bidirectionallayer wrapper. This wrapper takes a recurrent layer (e.g. the first LSTM layer) as an argument. It also allows you to specify the merge mode, that is how the forward and backward outputsshould be combined before being passed on to the next layer. It learns the characteristics ofthe phishing website and then predicts new phishing characteristics. The model uses neuralnetworks as the underlying learning framework.

B. TKINTER

Python offers multiple options for developing GUI (Graphical User Interface). Out of all theGUI methods, tkinter is the most commonly used method. It is a standard Python interface tothe Tk GUI toolkit shipped with Python. Python with tkinter is the fastest and easiest wayto create the GUI applications. Creating a GUI using tkinter is an easy task.It creates GUI for entering a URL and displays a pop up message whether it's a phishing one or not.

V. EXPERIMENT

A. DATASET

We collect 16000 of phishing and legitimate URLs. Thephishing websites consist of 12000 phishing URLs that hasbeen collected from Phish Tank . In the other hand, thelegitimate websites consist of 4000 legitimate URLs that havebeen collected by a daily use from 10 chosen users. However,the final dataset after handling missing data and removing theduplicate is size of 6116.

B. FEATURES EXTRACTION

The phishing websites have certain characteristics and patterns that can be considered as features. In this sub section, we cover all phishing website features that have been used in the previous researches as possible. Furthermore, while we are studying the phishing characteristics and patterns we notice some new characteristics that can be considered as features.

C. BLOCKING

When a phishing website is detected, automatically notifies and blocking user to use phishing websites. Otherwise user will allowed entering into Google chrome or any other.
We develop a GUI, for detecting phishing websites and blocking user to use phishing websites and automatically display a pop up message. We use tkinter to develop GUI. It is the standard
python interface to the tk GUI toolkit.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

FIGURE 2: USER INTERFACE

FIGURE 3: USER INTERFACE

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488 |

|| Volume 8, Issue 10, October 2020 ||

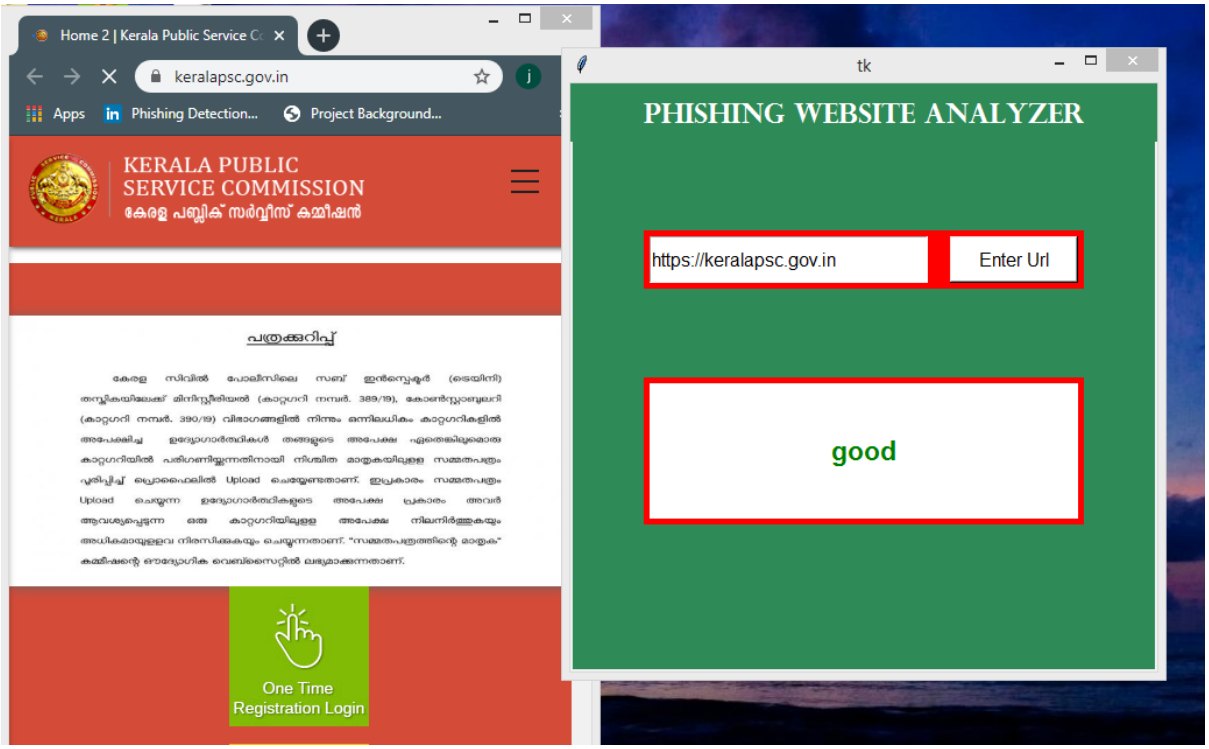


FIGURE 4: USER INTERFACE

VI. CONCLUSION

In our study, we design architecture to detect phishing webpage. Due to the BiLSTM, accuracy is enhanced in our system, which is less mentioned in others' study. The experimental results showed that compared with existing research, PDBS can detect the URL of the phishing website without relying on third-party data and search engines, with highest classification accuracy among other models. PDBS can also do blocking of the phishing sites i.e., when a phishing is detected an user can not entered into browser.In our experiments, the main problem was that the training time was too long, but the trainedPDBS model was far ahead of the existing research in terms of test time and accuracy. Thereare some other potential drawbacks to the classi_er. One obvious disadvantage is that whenthe phishing website URL itself does not have relevant semantics, PDBS will not be able toclassify correctly, and PDBS does not care whether the website corresponding to the URL isalive and if there is an error. Therefore, when applying PDBS to the actual detection scenario,it is necessary to verify the validity of the URL in advance.In addition, our system will be design as an extension to browser in thefuture.

REFERENCES

[1] AO Kaspersky lab. (2017). The Dangers of Phishing: Help employees avoid the lure of cybercrime. [Online] Available: https://go.kaspersky.com/Dangers-Phishing-Landing-Page- Soc.html [Oct 30, 2017].
[2] "Financial threats in 2016: Every Second Phishing Attack Aims toSteal Your Money" Internet: https://www.kaspersky.com/about/pressreleases/2017 financial-threats-in-2016. Feb 22, 2017 [Oct 30, 2017].
[3] Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: A Content-basedApproach to Detecting Phishing Web Sites," New York, NY, USA, 2007,pp. 639-648.
[4] M. Blasi, "Techniques for detecting zero day phishing websites." M.A.thesis, Iowa State University, USA, 2009.
[5] R. S. Rao and S. T. Ali, "PhishShield: A Desktop Application to DetectPhishing Webpages through Heuristic Approach," Procedia ComputerScience, vol. 54, no. Supplement C, pp. 147-156, 2015.
[6] "PhishTank — Join the fight against phishing." [Online]. Available:https://www.phishtank.com/. [Accessed: 29-Nov-2017].

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details