International Journal of Engineering Research and Advanced Technology (IJERAT)

E-ISSN : 2454-6135

DOI: http://doi.org/10.31695/IJERAT.2018.3285

Volume.4, Issue 7

July -2018

An Efficient Analysis Technique to Detect Suspicious Web Pages in Real-Time

Mr. Bhaskara L[1] and Mrs. Ankitha K[2]

Asst. Professor[1-2]

Sahyadri College of Engineering & Management

Affiliated to VTU, Belagavi

Mangaluru - 575007

India

ABSTRACT

Web has turned into a valuable piece of our general everyday life as we do the greater part of our social and monetary exercises on the web. Today every people are vigorously relying upon web and online exercises, for example, web-based shopping, web-based Banking, internet booking, online Recharge and some more. Phishing is a type of web risk, phisher make the copy of unique site and wrongfully endeavor to get Victim's own data like client name, secret key, Mastercard points of interest, SSN number and utilize it for possess advantage. A Non-customary client can't recognize whether site is phished or authentic. There is no any single answer for stop this fake movement. System can use the static features of a webpage to determine the genuineness of the website. The features such as number of iframe present in the page, number of redirections after clicking the URL, availability of JavaScript and embedded JS, special and extra characters included in the URL address etc. We have extracted such feature from testing site and those features are analyzed with the features of site present in the dataset. System has processed the structure of about One lakh site and trained to determine the suspicious webpage. System is attached to the browser as its extension to verify the URL entered by the user and it determines whether that site is suspicious or not.

Keywords: *Malicious, Web pages, Real-time analysis.*

1. INTRODUCTION

Today Internet has turned into an integral part of one's life. Web gives vital framework that assumes a significant part in different space for instance: communication, finance, business, E-administration and training. Some of the prominent application of web is Hybrid cloud, Big Data, E-shopping. Sadly, on the opposite side, violations over the Internet (cybercrime) have expanded at significantly quicker rate and with high multifaceted nature. Cybercrime assaults incorporates online cheats, cracking into the system, phishing assaults, DNS harming, malware assaults, information burglary, spamming, scams, blackmailing.

A "suspicious webpage" refers to a page that contains noxious substance that can misuse a client– side PC framework. Noxious site might be utilized as a weapon by cybercriminal to abuse different security dangers, for example, phishing, drive-by download what's more, spamming. Malignant Web destinations are jump in transit of Internet security. To deal with there is having to build up a programmed framework to perceived malignant site.

Phishing attack in one type of attack where attacker will send an email or message urging the end user to specify their credentials by providing a URL from inside the email. Now a day, there are different methods to do phishing and most of the methods will uses common features to get the advantage of trickery. This helps to detect the malicious site and it also helps to differentiate the

malicious pages from the genuine web page. This is achieved by fetching the features of visited web sites with the signature of pages that are already declared as suspicious web pages.

Phishing is ordinarily performed by email parodying or texting and it for the most part gives risk message or data refreshing notice which will diverts casualty to the page to give their classified points of interest. The phishy site looks fundamentally the same as the authentic site as normal client can't recognize it until he/she is general guest of that site. It is exceptionally typical for the non-IT proficient individual to overlook the risk message and move to the phishing site. Common individual can't distinguish the phishy site with first view. Indeed, even IT proficient people are observed to be effectively assaulted.

2. OBJECTIVE

- Objective is to outline and build up a system to distinguish web particular malignant pages progressively.
- The system extricates static highlights from a site page and make forecasts about its potential noxiousness.
- The system initially examines the list of capabilities utilized as a part of proposed component took after by the gathering procedure of the dataset.
- Lastly, creating browser expansion that will examine website pages continuously and give input to the client.

3. SYSTEM ARCHITECTURE

A system architecture is the applied model that characterizes the structure, conduct, and more perspectives of a framework. An engineering depiction is a formal description and representation of a framework, sorted out in a way that support thinking about the structures and practices of the framework.

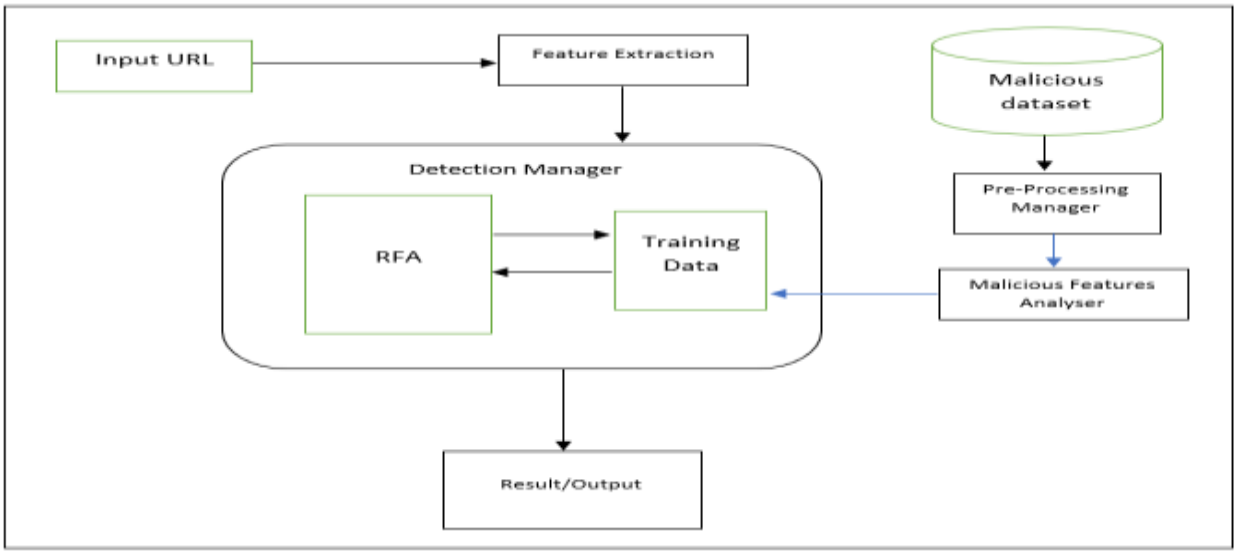

Fig: 1 System Architecture of suspicious webpage detection

The system architecture of the suspicious webpage detection is shown in Fig 4.1. The system will accept the URL entered by the user and extract the features from the website. And then system will give the extracted features to the detection manager. The detection manager contains the training data and Random Forest Algorithm(RFA) modules which in turn takes the input from the malicious feature analyzer. The training data contains malicious features which are extracted from the malicious dataset. The pre-processing manager takes data from the stored malicious dataset and processes the data. The detection manager extracts the features from site and compares features that are present in the dataset. Feature extracted from dataset are maintained in a file and it is useful in the test phase. This system is added as extension to the chrome browser so that it works as browser Add-ons. When the user enters any URL in the address bar, extension will check the URL and determines and provides the result as either phishing or normal page. System operation includes following steps.

3.1 Step A: Input

Dataset of URL is encouraged to the model at the underlying stage. Which contains 200 Legitimate and in addition phishing sites URLs, which will be utilized to Train the Machine Learning Algorithm and Test its execution.

DOI : 10.31695/IJERAT.2018.3285

3.2 Step B: Feature Extraction

It performs Extraction of highlights from input URL utilizing Script. For example, Lexical highlights, Host highlights portrayed. The consequence of each element is extricated utilizing guidelines and it will be utilized as a part of later stage. The removed aftereffect of the highlights is then used to give as a contribution to the following stage.

3.3 Step C: Classification

This stage utilizes the classifier to at long last get result. Classifier is only the machine learning calculation which is prepared to anticipate the outcome and perform arrangement. Since no single classifier is flawless and exact. The classifiers are picked for the most part since they have been utilized as a part of the issues like our own, for example, in distinguishing of spam, phishing messages, phishing sites and vindictive URLs and so on framework just needs to utilize it for conclusive forecast and grouping task.

4. PSEUDO CODE

Pseudo code is the partial code that portrays the activity to be performed. It is composed in a basic language and henceforth it turns out to be simple for the programmer to comprehend and dissect before the real implementation.

Pseudo code for Malicious Webpage Detection Process

URL – MWPDA(URL)

//Input : The URL of the website

//Output : Detecting URL as genuine or Malicious.

MT ← " ◯ " // MT are set of Malicious features

Initialize decision variable S=Genuine

1. For each URL, do the following
2. Check the URL if already marked as malicious

   If it is malicious

   Set S =" Malicious" and Go to step **5**
3. Extract webpage content
4. Compare and analyze with training data using RFA

   If match found, then Set S=" Malicious"
5. If S==" Malicious"

   Display "URL is Malicious"

   Display the set of malicious features in MT

   Block the URL and Alert the user

   Else

   Display URL is Genuine.

   Allow the URL to proceed.

   End for

End MWPDA

ABBREVIATIONS AND ACRONYMS

MWPDA- Malicious Web Page Detection Algorithm

URL- Uniform Resource Locator

RFA- Random Forest Algorithm

DNS- Domain Name System

5. CONCLUSION AND FUTURE WORK

Implemented Malicious recognition procedure uses different features such as IP address, URL length, Domain registration length, abnormal URL etc. Moreover, analyzation produced through Random-Forest algorithm recognizes the true genuine and malicious sites. System has used Random-Forest which indicated precision of 96% and low false-positive rate. The proposed model can diminish harm caused by malicious assaults since it can recognize new and malicious web pages.

In future work, we can consider new and more features to use and try to improve more accuracy and reduce false positive estimation of the system. Also, it is also possible to anticipate new features with high effect to identify Malicious page. Likewise, browser extension can be enhanced which will caution user in better way about malicious site and reduce harm cause with it however much as could be expected.

REFERENCES

[1] Dr. Jitendra Agrawal, Dr. Shikha Agrawal, Anurag Awathe,(et.al), "Malicious Web Page Detection through Classification Technique: A Survey", *IJCST Vol. 8, Issue 1*, Jan - March 2017.

[2] Jayakanthan.N, A.V.Ramani, "Malicious web page Detection: a state of art survey", *International Journal of Computer Science And Technology, ISSN : 2394-1537 , IJCST Vol. 5, Issue 3*, Jan - March 2016

[3] Ebubekir Buber, Önder Demir, Ozgur Koray Sahingoz, "Feature Selections for the Machine Learning based Detection of Phishing Websites" , *IEEE -978-1-5386-1880-6/17/* ©2017

[4] Chaitrali Amrutkar, Young Seuk Kim, and Patrick Traynor, "Detecting Mobile Malicious Webpages in Real Time" , DOI 10.1109/TMC.2016.2575828, IEEE

[5] N.Jayakanthan, A.V.Ramani andM.Ravichandran, "Two phase Classification Model to Detect Malicious URLs" , *International Journal of Applied Engineering Research ISSN 0973-4562* Volume 12, Number 9 (2017), pp. 1893-1898

[6] Jaydeep Solanki, Rupesh G. Vaishnav, "Website Phishing Detection using Heuristic Based Approach", *International Research Journal of Engineering and Technology (IRJET),* p-ISSN: 2395-0072 , Volume: 03 Issue: 05 | May-2016

[7] Rami M. Mohammad, Fadi Thabtah, Lee McCluskey, "Intelligent Rule based Phishing Websites Classification" *IEEE*

[8] Bhumika P Patel, Ghanshyam I Prajapati, "Phishing Attacks and its Detection", *International Journal of Research and Scientific Innovation (IJRSI) | Volume IV, Issue VIS*, June 2017 | ISSN 2321–2705

[9] Hyunsang Choi, Bin B. Zhu, Heejo Lee, "Detecting Malicious Web Links and Identifying Their Attack Types", binzhu@ieee.org.

[10] Himgouri P. Barge,Pankaj R. Chandre, "Malware Detection In Mobile Through Analysis of Application Network Behavior By Web Application", *International Journal of Engineering Development and Research (www.ijedr.org) 2016 IJEDR* | Volume 4, Issue 2 | ISSN: 2321-9939

[11) N. Manimozhi, S. Ramesh & Dr. T. Senthil Prakash, "A Security System to remove malicious webpages on mobile using kayo", *International Journal of Advanced Trends in Engineering and Technology (IJATET),* 5.665, ISSN (Online): 2456 – 4664, Volume 1, Issue 2, 2016

[12] Sadia Afroz, Rachel Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them", *IEEE*

[13] Sneha .J.Tripathi, Prof. V.S Gangwani, "Design the Framework for Detecting Malicious Mobile Webpages in Real Time ",*International Journal of Engineering Science and Computing*, May 2017*ISSN 2017 IJESC,* Volume 7 Issue No.5, 2017.

DOI : 10.31695/IJERAT.2018.3285

[14] Abubakr Sirageldin, Baharum B. Baharudin, and Low Tang Jung, "Malicious Web Page Detection: A Machine Learning Approach" *Springer-Verlag Berlin Heidelberg 2014*, DOI: 10.1007/978-3-642-41674-3_32.

[15] Yue Zhang, Jason Hong, Lorrie Cranor, "CANTINA: A Content-Based Approach toDetecting Phishing Web Sites" *International World Wide Web Conference Committee (IW3C2)*. May 8–12, 2007.

[16] Malware Domains List. Online available at: http://mirror1.malwaredomains.com/files/domains.txt.

[17] Phishtank Online available at: http://www.phishtank.com/.

[18] VirusTotal Online available at: https://www.virustotal.com/en/.