International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I10P108

Volume 69 Issue 10, 47-51, October 2021

Original Article

An Improved Web Fraud Detector and Preventive System with Cortical Algorithm

Laeticiannekaonyejegbu[1], Bunmi Deborah Millennial-Oriagbo[2]

Department of Computer Science, University of Port Harcourt and Federal Polytechnic of Oil and Gas Bonny, Rivers State, Nigeria.

Received Date: 15 September 2021
Revised Date: 20 October 2021
Accepted Date: 30 October 2021

Abstract - *Web fraud contents prevention and detection is now an important topic in the discipline of information security owing to the powerful emerging techniques often utilized by hackers to compromise a user computer system. Some or few web pages are benign midst the majority of web pages contain malicious web content. Most anti-virus packages in use are based on the use of signature-based access, and these are not able to reveal camouflaged malicious HTML codes. Hence, this project work proposed a malicious web page detective and preventive measure using the cortical algorithm approach of machine learning. This project examines the behavior of malicious web pages, compares the existing Naïve Bayes algorithm used in the detection of malicious web content with the proposed Cortical algorithm used for the design and implementation of this new system. Experimental results reveal that the new system is not just capable of detecting the malicious web content on webpages perfectly; it's also capable of removing iFrames and blocking Popups meant to cause distractions and to frustrate the user's efforts while working on the computer system. This research is the first of its kind to effectively block Popups and remove iFrames with the use of CORTICAL ALGORITHM.*

Keywords - *Web Fraud Detector, Cortical.*

I. INTRODUCTION

Fig. 1 Online fraud contents

A. Background to the Study

A malicious site page is a website page that contains a harmful substance that can destroy a customer-side PC framework. These malicious sites may attracts a weapon by cybercriminals to abuse different security dangers, for instance, phishing, drive-by download, data taking, and spamming.

As indicated by[1], malware is a type of PC program meant to taint a real client's PC and cause hurt on it in various ways. As the decent variety of malware is expanding, hostile to infection scanners can't fulfill the requirements of security, leading to a large number of hosts being assaulted.

According to [1], around 6 563 145 distinct hosts were assaulted, and 4 000 novel malware items were identified in 2015.

[2]Predicts the expense of information ruptures to increase to $2.1 trillion universally by 2019.

B. Statement of the Problem

The introduction of PC and the web has made the world a worldwide town. Hackers exploit this mechanical progression to participate in various types of malicious website pages and cybercrime, running from terrorism, web extortion, and malware to the arrival of complex infections, which is difficult to follow the culprits because of the absence of modern programming that can recover data of such exercises. This brought about the plan of a detection and prevention approach on these malicious website pages utilizing a cortical algorithm, which will help in following the activities of web clients and for the account of computerized proof of wrongdoing submitted in a PC framework or web.

C. Aim and Objectives of the Study

The study is aimed at developing a detective and preventive measure framework for fraudulent web content utilizing a cortical algorithm. The goals are:
To Design an efficient detection and prevention system that will detect and prevent fraudulent web content.
Implement the system utilizing JavaScript Programming Language and MySql database.

D. Importance of the Study

Malware and other conceivably unsafe programming have a genuine effect on clients' security, dependability, and protection. Attackers are increasingly persuaded by the monetary benefit to take confidential or individual data instead of simply vandalizing the customer machine, which makes clients experience moderate execution and dependability issues with their PC because of the quality of spyware and other malware activities.

This study is significant in that it raises new thoughts that give the premise to growing new ideas and their applications, to give a basic assessment, and to analyze the viability of the current practice (Stanish, 1999). This investigation hence draws out the issues characteristic in the conventional cybercrime model and augments the possibility of creating a detection and prevention approach for detection and prevention of web contents and malicious site page by not simply identifying the current vindictive substance on both PC framework and website pages, yet additionally distinguishing recently presented malicious content on the site pages and on the PC and evacuating them appropriately.

E. Scope of the Study

This research work covers the design and Implementation of Machine Learning dependent on the detection and prevention approach utilizing a cortical algorithm. This research covers both Local and wide area networks where individual and public organizations exchange have been carried out on the internet. It identifies however many site pages and system associations as could reasonably be expected. The new system can distinguish totally existing malicious web pages and recently presented malware substance on site pages and furthermore preventing their attacks on the PC.

F. Limitation of the Study

- The algorithm used is purely a new approach for solving the problem associated with the detection of malicious web content. The algorithm has never been used by anyone to tackle this problem.
- Lots of time and hard work were incurred in understanding the framework of the new algorithm utilized for the development of this newly-introduced system.
- A huge amount of money was involved in hosting the website domain name on the web.
- Suppose the user web browser does not have addon or plugin capabilities, e.g. (Mozilla). The program will not execute.
- If JavaScript is not installed on the user system browser, the code will not execute.

G. Review of Models for fraud detection and Prevention on Web Site

Web fraud detection and prevention of sites is one of the hot research themes in the field of security. The significance is assessed from these insights, which demonstrate that Google finds 9,500 new malignant Web locales daily. This is likewise significant on the grounds that it keeps the client from being victims of attacks, for instance, Phishing attacks, Drive-by–download attacks, spam attacks, Click-jacking, Plug-In, and Script-Enabled Attacks, Mal-advertising. Malicious web identification is seen as a way of recognizing those web URLs and website pages that can cause trade-offs in client security and influence the clients[3]. Without a doubt, malicious detection on the Internet (cybercrime) has expanded at a lot quicker rate and with high multifaceted nature. Cybercrime assaults include online frauds, breaking into the system, phishing assaults, DNS harming, malware assaults, information stealing, spamming, tricks, extorting. An ongoing case of cybercrime is the Sony Cyber Hack, through which the PC at the corporate network of Sony Pictures was ruptured and taken disconnected by a malware-based attack. Intrusive events to PC systems are expanding because of the loving and reception of internet and local area network computerized hacking instruments and techniques. PC,s are gaining the chance to be increasingly more exposed to attacks because of their widespread system network.

H. Roles of Malicious Detection System

The role of the detector and preventer is to find and dispense with unneeded data from the review trail. It then displays either manufactured views on the security-related moves created during normal utilization of the framework or the engineered views on the present security state of the framework. A choice is then made to assess the probability that these activities or this conduct can be considered as side effects of an intrusion or vulnerabilities. A preventive measure part would then be able to make the restorative move to either keep the activities from being executed or change the condition of the system back to a protected state.

Modules, for example, ActiveX, are notable for their powerlessness to pernicious dynamic substance. Moreover, the malicious dynamic substance can take a secret phrase or stick and later get to a site with classified data while causing it to seem like it was gotten to by the approved client. This can make it even hard to follow whether a malicious dynamic substance is used up in the attack.

I. The framework of Cortical on multi-class Malicious URLs Detection System

Cortical Algorithm (CAs) rose as a naturally roused methodology, displayed after the human visual cortex, which stores arrangements of examples in an invariant structure and reviews those examples auto associatively.

J. Cortical Algorithm Structure

The human brain is comprised of a six-layered structure comprising of an exceptionally enormous number of neurons emphatically associated by means of feedforward and feedback connections. A significant property of the neocortex is its basic and useful consistency: all units in the system appear to be comparable, and they play out a similar essential activity.

Precisely like brain architecture, CA design has minicolumns of shifting thickness [4]. A minicolumn is a gathering of neurons that offer the equivalent open field: neurons having a place with a minicolumn are related with the equivalent sensory input region. The minicolumn is the essential structure in a cortical system network, in contrast with neurons in an old-style ANN. A relationship of minicolumns is known as a hypercolumn or layer (in what pursues, the terms segment and minicolumn are utilized reciprocally). Associations in a CA system happen in two ways: horizontally, between segments in a similar layer, and vertically, between sections of sequential layers. In spite of the fact that associations between nonconsecutive layers are available in the human cortex, these associations are evacuated in CA for effortless purposes.

K. The idea of Anti-Malware Technologies

Malware and other possibly unsafe software strongly affect clients' security, dependability, and protection. Attackers are progressively aroused by the monetary benefit to take confidential or individual data and not simply vandalize the customer machine. In addition, clients can encounter genuine execution and security issues with their PC because of the quality of spyware and other malware.

An enemy of malware motor is fit for recognizing and evacuating malware as it endeavors to taint a PC. This motor performs three fundamental undertakings:

- Scanning
- Detection
- Removal

L. Challenges in the Malicious Web Content Detection and Prevention Techniques

Majority of the present or existing methods used in detecting and preventing malicious sites depend on their center procedures for a notable attack. Hence, the attackers continue creating amendments in the existing methodology and acquaints new systems with be implanted in the site page. The current technique depends upon the fixed arrangement of highlights, yet the aggressor continues making amendments in the current features and furthermore presents new features. Thus, the discovery and anticipation strategies are not equipped for distinguishing the new attacks. This requires that the detection methods be improved. The various strategies like signature-based, feature-based, and behavior-based methodologies utilized in detecting malicious sites and contents are confronting these confinements because of modern attacks. In accordance with the constraints, the various existing features are not adequate for the detection of malicious sites. For instance, existing methodologies are not fit for distinguishing malicious sites based on the area name because the attackers often change the space. Besides, none of the component gathering strategies can gather the rising highlights. The current detection techniques suffer false negativities. Thus, there should be a better way to solve these problems.

M. Iframe

The iframe is an html an html label utilized in including the outer site into your website (webpage).

N. Iframe infection

An Iframe virus is malicious code that infects pages on sites. These are considered a type of malware. A large portion of them uses iframe html code, causing harm by infusing iframe labels into the site. Code might be infused into html, php, records. The infection may make its essence is known by checking for landing page records, for example, index.php, index.html, or default.html, and infusing the iframe code in them. The iframe code is typically found close to the start on the website page. They may likewise contaminate through subjects or formats of substance the executive's frameworks. The virus will likewise alter .htaccess and hosts documents and make images.php records in registries named 'pictures'. The contamination isn't a server-wide endeavor, and it will just infect locales on the server that it has passwords. This ongoing flood in bargained web servers has produced exchanges in online discussions and web journals. Web malware diseases hurt organizations; google, firefox, web adventurer, and against infection organizations boycott contaminated destinations, organizations lose income, and locales endure harm to their image and notoriety. An iframe infection is a sort of badware. "badware makers are always growing new, innovative approaches to introducing badware onto your computer". Badware dispersion has been extended past conventional channels like email infections to harder-to-stay away from techniques like mechanized "drive-by downloads" that are propelled by traded off-site pages.

O) Iframe variations

Here and there, iframe variations come as JavaScript. iframe labels may not be found in plain content in the source since it is encoded. On the off chance that the encoded content code is decoded, it will contain code to conjure iframe by means of JavaScript.

P) Iframe Injection Website Attack and Tips to Clean the Infection

The most popular online attacks that happen all the time have to do with a huge amount of real sites being hacked with the Iframe code infusion assault bringing about Cross-Site Scripting (XSS) or quiet redirections to malevolent sites. This implies unsuspicious guests get contaminated with a malware infection when they visit those genuine sites. The result is that an "Iframe html code" is infused toward the end (for the most part) of index.php or index.html documents of the real sites. This iframe code inserts into the genuine site a malignant code that introduces an infection to the guest's PC or attempts to take delicate data. The <iframe> html tag is utilized to install content from another site into the present page. Typically an infused iframe code resembles the accompanying:

<iframe src="http://some-vindictive web-url" width=1 height=1 style=" display:none"></iframe>

II. FUNCTION OF MALWARE (MALICIOUS CONTENTS) ON WEBPAGES

- Recover all put-away usernames and secret words on the internet browser.
- Hinders site execution and the internet browser all in all.
- Overabundance popup on the page

A. Existing System Analysis

The current system was a Phishing URL detection system utilizing Naïve Bayes. This current methodology contrasts the suspicious website with the genuine one site by utilizing different highlights, and on the ground that the difference is more than the predefined limit esteem, it is proclaimed phishing if the threshold is exceeded.

The existing System catches images (screen capture) from suspicious URLs. It, at that point, contrasts the given images and the put-away images in the database utilizing ImgSeek[5] in the event that it will find the comparative pictures in the database. After picture correlation, the framework thinks about the area name. On the off chance that the space name is absent in the database, it implies that the info URL is extraordinary. By then, the framework pronounces the given URL phishing. In the event that there is no picture in the doubts site whose likeness is more prominent than the edge esteem, at that point, the proposed procedures return the outcome as obscure and register the picture in the database.

B. Algorithm of the Existing System

1) imgSeek(Imgsuspicious,Imglegitimate,Imgphish,Imgunknown)
2) if(Similarity(Imgsuspicious,Imglegitimate)>threshold)
3) if(domain(Imgsuspicious)=domain(Imglegitimate))then return(“Legitimate”)
4) else return(“Phishing”)
5) end if
6) else
7) storeDatabase({Imgsuspicious , domain(Imgsuspicious), Unknown})
8) end if

C. Analysis of the Proposed System

The proposed system is Web fraud pages detective and preventive system utilizing a cortical algorithm. The proposed system utilizes the cortical algorithm to prepare some datasets, and the limit was set in terms of malicious content and ordinary substance. The dataset is spoken to as vectors, and the estimations of the vectors as being set has 1.0 as would be expected characters while 5.0 is set as the vindictive threshold. The program show level of malicious attacks and how to cut off it has been influencing the PC system in quantities of popups. Every malware demonstrates diverse popups relying upon the record or report it has attacked.

D. Algorithm of the Proposed System

1) Initialization input training samples; then digitizing and normalizing the input data;
2) Reducing the dimension of the data on the browser to indicate two sets of vectors
3) Input vector with dimensionality reduction, network parameter to initialize the classifier;
4) Set the layer i=5;
5) Train the network layer by layer according to cortical learning rules, then save the result including the weights and biases;#
6) If i<=max layer, set i=i+1; when i > max layer, do the supervised learning
7) Input the dataset test samples into the trained classifier to detect malicious code and the normal code.

E. Advantages of the Proposed System

Coming up next are a portion of the advantages of the proposed system:

- Easy to use: The new proposed system is easy to use by internet users when compared to the existing system.
- It detects and prevents malicious content in the new and existing websites.
- It partitions the output into different rules, which aid in identifying the normal and the malicious data.
- User friendly and does not need an expert to operate.

The point of the structure stage is to design an answer to the issue determined by the prerequisite record. This stage is the initial phase in moving from the issue space to the arrangement area. Framework configuration is maybe the most basic factor influencing the nature of the product and majorly affects the later stages, especially testing and support. Each undertaking requires a structure so as to give the client mandates. In this venture, we utilized UML and calculation to make this program simple. Framework configuration is the way toward characterizing the engineering, segments, modules, interfaces, and information or a framework to fulfill indicated necessities.

III. DISCUSSION OF RESULT/COMPARISON OF THE EXISTING AND PROPOSED SYSTEM RESULT

The current methodologies contrast the suspicious site and the relating benign site by utilizing different features and on the ground that the similarity is more than the predefined threshold; at that point, it is considered phishing.

The existing System caches the images (screenshots) from suspicious URLs. It, at that point, contrasts the given images and then puts away images in the database utilizing ImgSeek, on the off chance that it will find the comparable pictures in the database. After images correlation, the system compares the domain name. If the domain name is not present in the database, it means that the input URL is different. At that point, the system declares the given URL phishing. If there is no image in the suspicions website whose similarity is greater than the threshold value, then the proposed techniques return the result as unknown and register the image in the database.

The proposed system is also a single algorithm for the detection and prevention approach. The system algorithm detection threshold is set to be 5%. The algorithm scans both the existing URL domain names and any new URL domain names for detection and prevention purposes. The new system has a high detection rate with other features capabilities not inherent in the existing system.

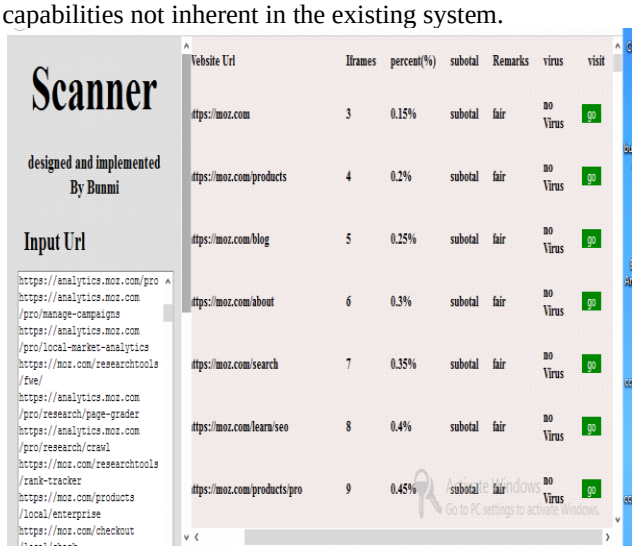

Fig. 2 Result of the Proposed System.

A. Discussion

a) Documentation and Implementation

Having confirmed that the proposed system meets the objectives of the project, the implementation phase begins. Implementation is the stage of a project during which theory is turned into practice. The proposed malicious web pages detection and prevention are implemented by installing a JavaScript integrated development environment. Also, the Xampp server was installed for the smooth running of the web page on the website.

b) Running the Application.

Lunch the JavaScript IDE on the computer, and it Click on the program and click on the run to execute the program. Then lunch the website on the computer system to link the website to the program and also copy the link of the site into the program to verify the content of the malicious code that has attacked the data and files in the web page. Then click on the browser to refer to each document and file on the website. The application will detect all the components of the file and document on the site and give results about the web page.

IV. SUMMARY, CONCLUSION, AND RECOMMENDATIONS

A. Summary

Malicious Web pages are increasingly spread while accessing the website on our different computers daily. However, this malicious content or code that attacks web pages on the web browser goes directly into the computer local disk and is stored in different files or document and also create a storage space, thereby affecting and damaging different data in the computer. These malicious contents that attack the web pages also cause different types of computer viruses that infect the computer system to malfunction. However, in spite of significant advances in processor power and bandwidth, the browsing experience on different computer devices is considerably different when the computer has been affected by these threats. The advancement in computers has made computer networks and other smart devices increase the number of services that are available on the Internet, with many people accessing the website through web applications. A number of these web applications provide convenient services to users, which include online commerce, communicating through social network applications and services, or surfing for information online.

B. Conclusion

A malicious web page or code is malware that contains harmful content that can destroy a client-side computer system. This type of attack is termed a web-based client-side attack. The attack is delivered as part of the web page itself and is designed to exploit client-side vulnerabilities such as flaws in the implementation of browser functionality, interpreters of active content within web pages, or scriptable client-side components such as HTML components. This new system is capable of preventing pop-ups meant to course unnecessary distractions when users operate the computer system and also prevents iFrames viruses that secretly penetrate and infect the user's system.

C. Recommendations

We recommend that this software should be packaged as an Addons and be made available on various web browser applications stores so that users can add it to their web browsers for detection and prevention of malicious contents purposes.

D. Contribution to Knowledge

We have been able to remove iFrame and pop-ups from webpages automatically with JavaScript, whether offline or online. This is the basis of webpages detection and prevention.

REFERENCES

[1] Kaspersky Labs., (2017) What is malware, and how to defend against it? www document. Available at: http://usa.kaspersky.com/internet-security center/internet-safety/what-is-malware-and-how-to-protect-against-it#.WJZS9xt942x. [Accessed 15 February 2017]

[2] Juniper Research. (2016). Cybercrime will cost businesses over $2 trillion by 2019. www document. Available at: https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion. [Accessed on 15 February 2017]

[3] Karl A. McCabe … US7607171B1 (en) *, 2002-01-17, 2009-10-20, Avinti, Inc … US20090271867A1

[4] Edelman, Gerald M., and Vernon B. Mountcastle., The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function. Cambridge, MA: Massachusetts Institute of Technology Press, (1978).

[5] Surendiran,R., Similarity Matrix Approach in Web Clustering. Journal of Applied Science and Computations (JASC), 5(1) (2018) 267-272. DOI:16.10089.JASC.2018.V5I1.140146.22858

[6] C. Jacobs, A. Finkelstein, and D. Salesin, Fast multi-resolution image querying, in Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95), (1995) 277–286. Los Angeles, Calif, USA.

[7] Ankit Kumar Jain and B. B. Gupta ., Phishing Detection: Analysis of Visual Similarity-Based Approaches National Institute of Technology, Kurukshetra, India, (2017).