WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Unmasking Malicious Web Links and Decoding Attack Types for Enhanced Security

Krishna Priya S[1] , Dr. Vaidehi V[2]

PG Student[1] ,Asst.Professor[2]
Email: skrishnapriya345@gmail.com
Email: vaidehi.mca@drmgrdu.ac.in
Department of Computer Applications,
Dr. M.G.R Educational and Research Institute Chennai-600095

ABSTRACT:

Malicious URLs, such as those used for spam, phishing, and malware, are typically employed in a variety of attack types. Detecting malicious URLs and determining the nature of the threat is crucial for preventing these attacks. By understanding the type of threat, the severity of the attack can be assessed, and appropriate measures can be implemented. Traditional methods typically display a single malicious URL. We suggest a method in this paper that employs machine learning to recognize malicious URLs for all known attack types and evaluate the nature of the attacks that these malicious URLs attempt to execute. Our method assesses several distinguishing elements, such as text structure, link structure, web content, DNS details, and network traffic. Several of these qualities are fresh and beneficial.

Our research on 40,000 reliable URLs and 32,000 unreliable URLs collected from the authentic Internet demonstrates that our method exhibits exceptional performance in identifying bad URLs with a success rate of over 98% and detecting attack types with a rate of 93%.

KEYWORDS: Malicious URLs, Cyberattack, DNS [Domin Name System] Information, Detecting, Enhanced Security.

INTRODUCTION:

In the contemporary digital era, the internet has become an indispensable component of daily life. Nevertheless, the extensive utilization of technology has also led to a rise in malicious activities and cyber-attacks. Web links that are designed to deceive and harm unsuspecting users are one of the most prevalent forms of these attacks. These harmful links often culminate in data breaches, identity theft, and other cybercrimes.[1]

URLs are the unique addresses of web pages and other resources on the internet. They are used to locate and access documents and other content on the World Wide Web. A comprises two main components: the Protocol Identifier (which indicates the protocol to be used) and Name of the Resource (which specifies the IP address or the name of the domain where the resource is located) that are separated by a colon and two forward slashes. The components of a URL, malicious URL, or malicious websites are common and serious threats to cybersecurity. They act as the gateway for unsolicited activities hosting a variety of luring content, such as spam and phishing. Innocent users visit such websites and subsequently become victims of various types of scams, including pecuniary loss, loss of personal data such as identity, credit cards, etc., and ransom ware installation on user devices, resulting in huge global losses.[2]

To combat this threat, it is crucial to have a deeper understanding of how these links work and how they can be identified. Unmasking malicious web links and decoding attack types is essential for enhanced security in the online world. It involves recognizing the various tactics used by hackers to disguise their harmful links and understanding the different types of attacks that can be carried out through these links.[3]

By unmasking these malicious web links, individuals and organizations can take proactive measures to protect themselves from potential threats. This may include implementing stricter cybersecurity protocols, educating employees or users on safe browsing practices, and using advanced tools to detect and block such links.[4]

In this article, we will delve into the world of malicious web links and explore techniques for identifying them as well as decoding different attack types. By gaining a better understanding of this topic, readers will be equipped with the necessary knowledge to bolster their online security measures. Let us embark on this journey towards enhanced security in an increasingly interconnected digital space.[5]

LITERATURE SURVEY

- **Shahreen Kasim,** et al., (2020) The objective of this study is to evaluate the effectiveness of machine learning in identifying and detecting malicious URLs. To achieve this, we utilized a bio-inspired algorithm to optimize URL features and select those that are most significant in detecting malicious URL applications. The static analysis technique was employed in conjunction with machine learning to detect these malicious applications. (6)

- **Ryuya Uda,** et al., (2011) There appears a new type of vicious spots that avoids check by malware checking spots. In this paper, existing protocols and styles related with the web are reanalysed in terms of protection from current attacks, and new protocol and system are indicated for the purpose of security of the web. (7)

- **Hesham Alshaikh,** et at., (2020) noted that ransomware can enter a system through various methods, similar as social engineering, malware advertising, spam emails, exploitation of vulnerabilities, drive- by downloads, or through open ports or back doors. Unlike traditional malware, the goods of ransomware are irrecoverable and delicate to mitigate without the assistance of its creator. This type of attack has a direct financial impact, driven by encryption technology and cryptocurrency. (8)

- **Yuefeng DuI,** et at., (2022) In light of this, we propose a new three- party paradigm PEBA with an intermediate third party disconnecting the direct commerce of users and personal blacklist merchandisers. To satisfy practical operation conditions, we express our design with trusted tackle, detailing how it can be leveraged to full fill the conditions of sequestration improvement and broader content at the same time. We also attack numerous perpetration challenges that surfaced from this deputy- grounded and tackle- enabled result. (9)

- **Harjinder Kaur,** et at., (2013) That system which detects the intrusion in the system is known as IDS (Intrusion discovery System). This conception has been around for two decades but lately seen a dramatic rise in the popularity and objectification into the overall information security structure. therefore, the intrusion can be set up substantially using two category ways Misuse or hand- based detection and the other is Anomaly detection. (10)

PROPOSED SYSTEM:

The enhanced Convolution Neural Network (CNN) model is proposed by the authors is one of such a candid system fettered with re-learning ability. In this vibrant system the authors have added a character-level embedding layer before the convolution layer which empowers the system to learn the intrinsic relationship between the characters of the query string.

In this system the authors have also modified the CNN filters which could extract the fine-grained features of the query string. The test conducted proves that the present model has lower False Positive Rare (FPR) while comparing with Random Forest (RF) and Support Vector Machine (SVM). Network plays a vital role in detecting several cyber security issues. This model is accurately classified the URL as safe or malicious based on the word to vector feature selection. The performances of the proposed method have the accuracy of 85%, analyses with different performance metrics viz., precision, recall, and F1 –score.

ARCHITECTURE DIAGRAM

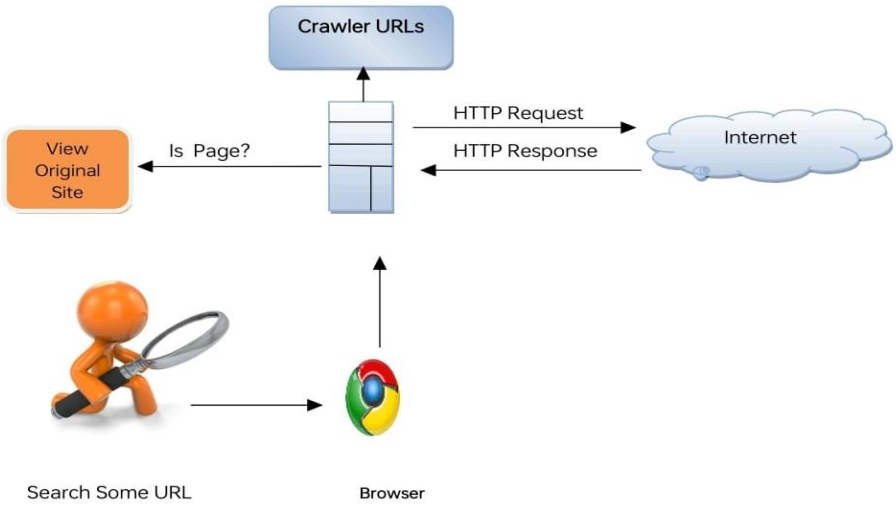

Fig: 1 Architecture of proposed system

EXPLANATION:

Fig 1. Demonstrates the framework diagram for the proposed system. The proposed web crawling system framework consists of factors similar as" Crawler URLs" for storing discovered URLs," Is Page?" for validating web pages," View Original Site" for displaying web page content," Search Some URL" for targeted querying, and" Browser" for user commerce. It also involves" HTTP Request" for transmitting requests and" HTTP Response" for server replies. The" Internet" facilitates communication between browser and garçon, essential for effective reclamation and exploration of web page content

MODULE DESCRIPTION

System development deals with the operations that are carried out to get desired affair from software product based on certain design specifications. This operation holds the following two main modules.

1. USER
2. ADMIN

USER:

Users interact with the application through modules such as "Search URL (Using Keyword)," where they input keywords to search for relevant URLs. Upon initiating a search, users can view the search results within the "View Search Result" module, enabling them to browse and select URLs of interest. In the event of encountering potentially malicious pages, users can utilize the "View Malicious Page" module to assess the content and take appropriate actions. Finally, the "Logout" module allows users to securely terminate their sessions, ensuring privacy and security. Each module serves a distinct purpose, empowering users to search, browse, and manage their interactions with URLs and web pages effectively within the system.

ADMIN:

The admin module serves as a comprehensive toolset for system administrators, offering functionalities to manage and oversee system operations efficiently. Administrators access the system securely through the "Login" module, gaining access to administrative features. They can add URLs to the system database using the "Add URLs" module, monitor existing URLs via the "View URLs" module, and perform real-time searches for specific URLs with the "Search Real-Time URLs" module. The "View Result" module provides administrators with insights into system actions and outcomes, while the "Download" module allows them to extract data or reports for analysis. Administrators can preprocess data using the "Pre-processing" module and classify URLs based on criteria using the "Classification" module. Finally, administrators can securely log out of their accounts using the "Logout" module. This module suite empowers administrators to effectively manage system resources, data, and operations.

IV. RESULT AND DISCUSSION

1. Home Page

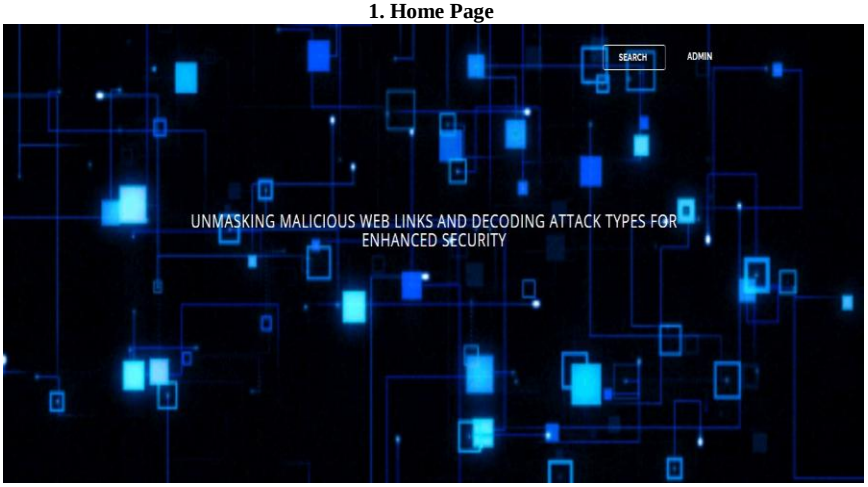

Figure 2: Home Page

In figure 2, it includes the admin module, user module and the registration page. After logging in, they access respective functionalities through the home page.

2. Admin Login Page

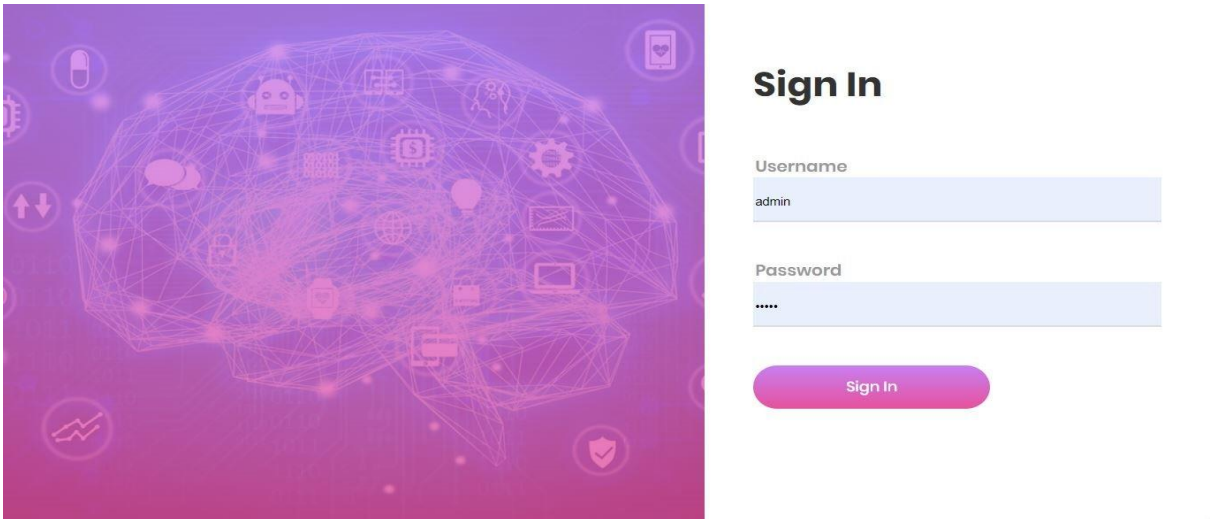

Figure 3: Admin Login Page

In figure 3, To access the crawler page, first the user should sign in the login page by entering the username and password in the given field.

3. Admin Home Page

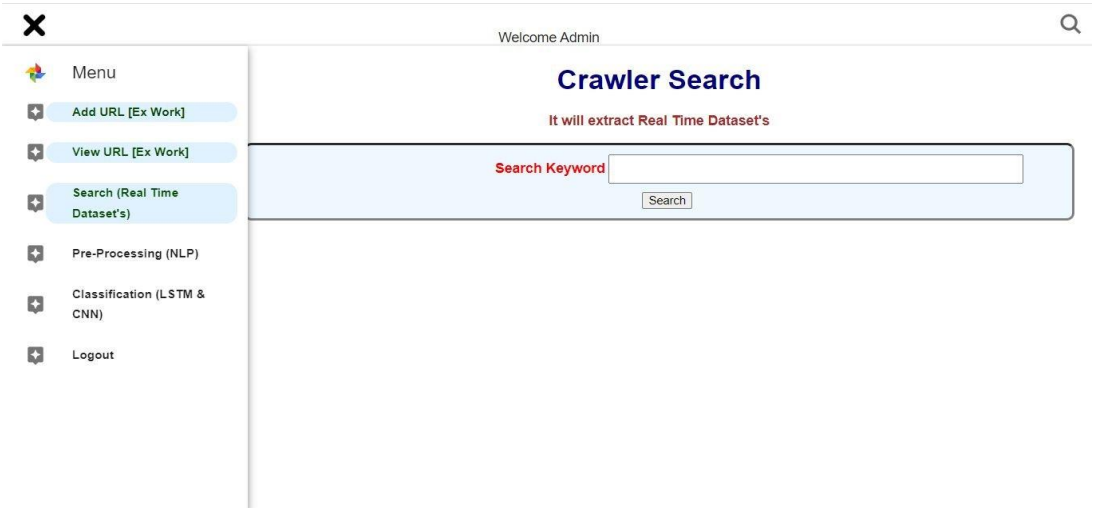

Figure 4: Admin Home Page

In figure 4, demonstrates on this page, the administrator has access to various features and functionalities related to managing the system. On the crawler page, there is a search feature that allows the administrator to input search criteria. The administrator to initiate the search process based on the entered keywords.

4. Add URL Page

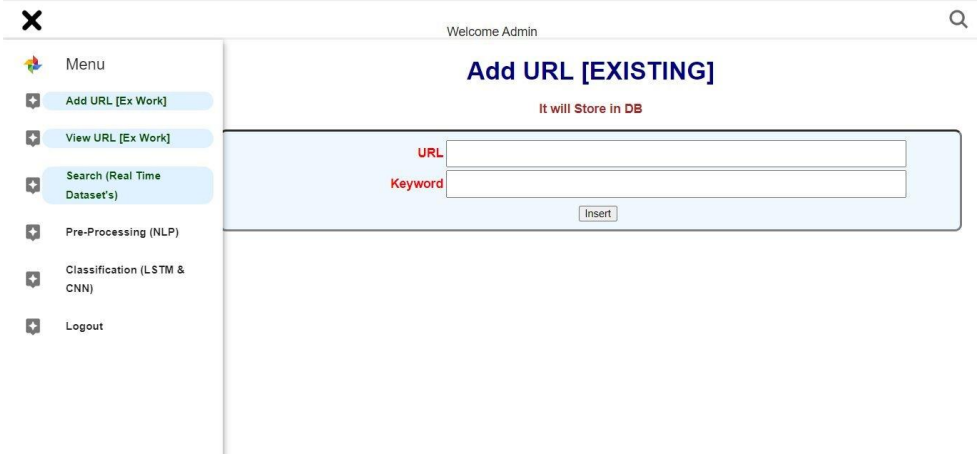

Figure 5: Add URL Page

In figure 5, demonstrates the "Add URL" page contains a form where the administrator can input a new URL. The insert button triggers the search process, and the results are displayed on the page for the administrator to review.

5. View Add URL Page

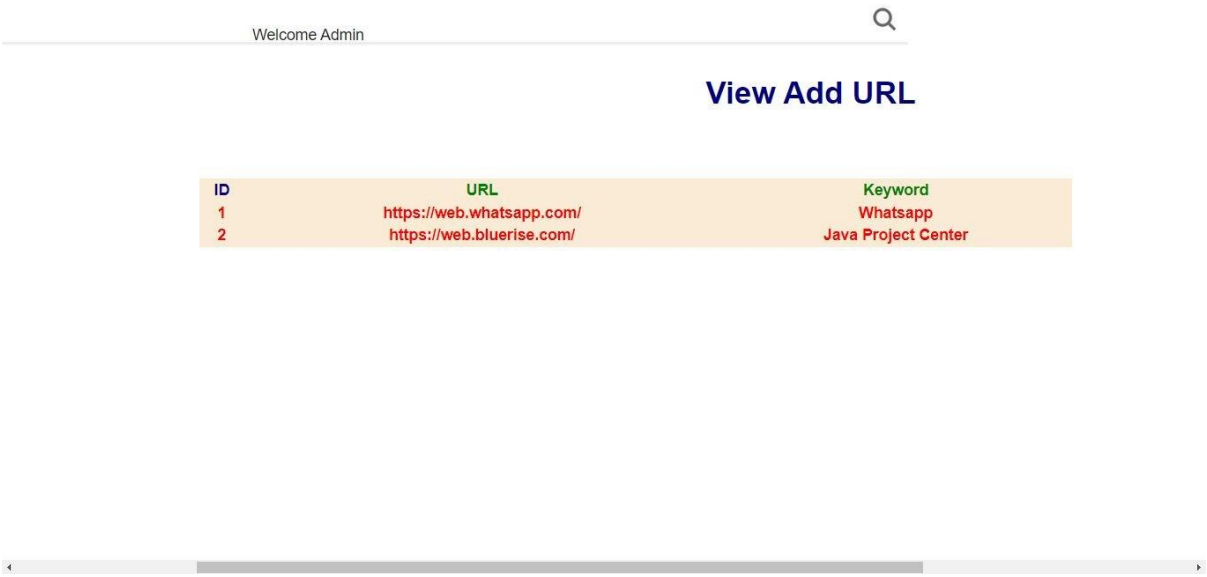

Figure 6: View Add URL Page

In figure 6, demonstrates on the "View Add URL" page, the values entered by the administrator on the "Add URL [Existing]" page are displayed. The URL value entered is displayed along with its associated keywords in list format.

6. Web Crawler Page

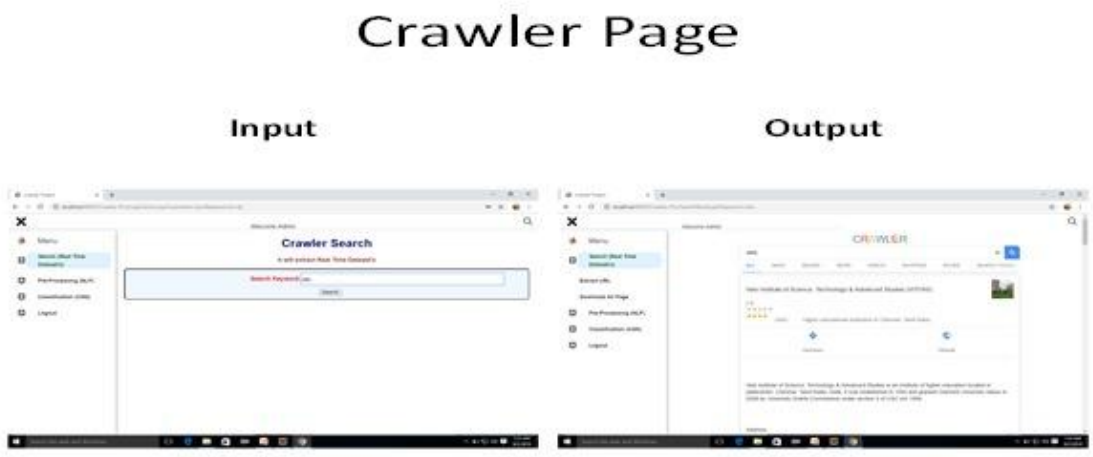

Figure 7: Web Crawler Page

In figure 7, demonstrates the Input as a keyword, it should be entered in the given search box Output will be shown as a list of links in the crawler page and it is the real time datasets.

7. Extract URL Page

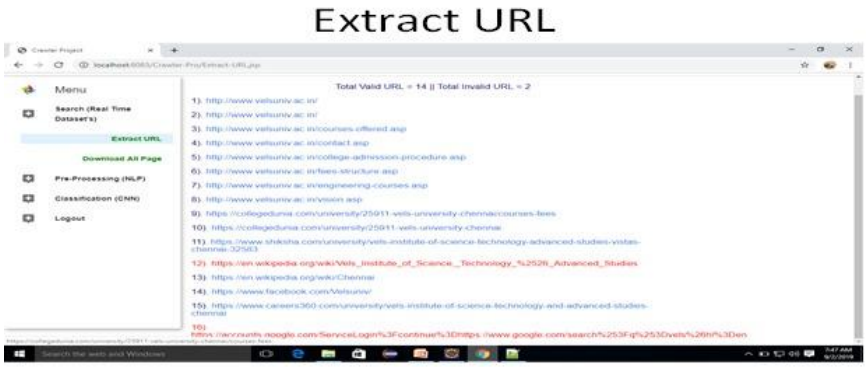

Figure: 8 Extract URL Page

After crawling, the results will be extracted and show whether it is valid URL and invalid URL.

8. Result

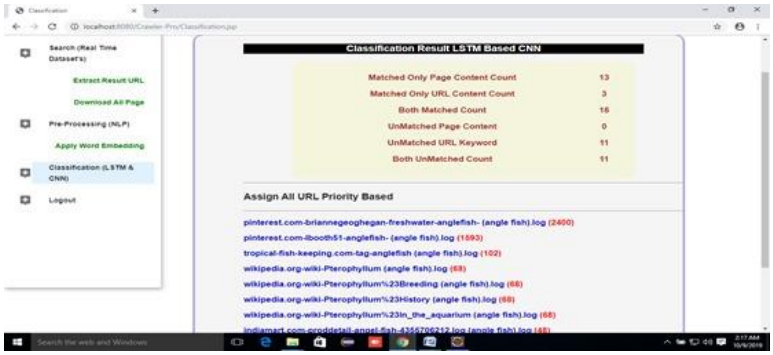

Figure: 9 Result

At last classification result, like total count of matched (relevant) content in a pages and URLs. And unmatched (irrelevant) content and URL keywords is shown for all the downloaded web pages. And Re-rank (assign) the URL by priority-based order. From larger too small.

CONCLUSION

The World Wide Web is a vast collection of millions of webpages and data spanning a diverse array of subjects. Despite its size, it lacks a centralized structure for organizing its content. Hence, searching for data related to a specific topic is difficult. With such challenging tasks, the role of web crawlers becomes more important, and design issues must be taken into consideration to produce more effective results.

This project presents a detailed study of web crawlers, extracts URL, downloads relevant web pages, preprocesses, and word embedding, and then classifies them using a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). Finally, the URL are re-ranked according to the content available on the webpage.

They are ranked based on the priority of URLs (the largest amount of matched content on the web page will be ranked top on the page). This will help the user search for more relevant webpages on the top of the list. It saves user's time.

V. REFERENCE :

[1] Sheetal Bairwa, Bhawna Mewara and Jyoti Gajrani "Vulnerability Scanners: A Proactive Approach to Assess web Application security" Vol.4, No.1, February 2014

[2] Moaiad Ahmad Khder "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application" Int. J. Advance Soft Compu. Appl, Vol. 13, No. 3, November 2021

[3] Firoz Khan, Jinesh Ahamed, Seifedine Kadry, Lakshmana Kumar Ramasamy "Detecting malicious URLs using binary classification through add boost algorithm" Vol.10, No.1, February2020

[4] Chanchala Joshi, Umesh Kumar Singh "Performance Evaluation of Web Application Security Scanners for More Effective Defense" Volume 6, Issue 6, June 2016

[5] Anjali B. Sayamber, Arati M. Dixit “Malicious URL Detection and Identification” International Journal of Computer Applications (0975 – 8887) Volume 99 – No.17, August 2012

[6] Ong Vienna Lee, Ahmad Heryanto, Mohd Faizal Ab Razak, Tole Sutikno “Machine Learning and Optimization Techniques in Malicious URL Detection System” Vol. 17, No. 3, March 2020

[7] Ryuya Uda "Protocol and Method for Preventing Attacks from the Web" International Journal of Computer and Information Engineering, 5(4), pp.380-385. Vol:5, No:4, 2011

[8] Alshaikh, H., Ramadan, N. and Ahmed, H., 2020. "Ransomware prevention and mitigation techniques" Int J Compute Appl, 177(40), pp.31-39. Volume 177 – No. 40, February 2020

[9] Du, Y., Duan, H., Xu, L., Cui, H., Wang, C. and Wang, Q., Peba. “Elevating Security in Web Browsing”. IEEE Transactions on Dependable and Secure Computing. Volume: 20 Issue: 5, 2022

[10] Harjinder Kaur, Gurpreet Singh, Jaspreet Minhas "A Review of Machine Learning based Anomaly Detection Techniques " International Journal of Computer Applications Technology and Research Volume 2– Issue 2, 185 - 187, 2013