International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 9, pp 37-42, September-2023

https://doi.org/10.47001/IRJIET/2023.709004

WebGuardian: Holistic Approach to Address Dynamic Web Application Threat Landscape

[1]Aththanayaka P.A.G.P.B., [2]Ranasinghe M.H., [3]Ranaweera H.N.K, [4]Rathnayake S.D., [5]Amila Senarathne, [6]Kanishka Yapa

[1,2,3,4]Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka
[5,6]Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka

***Abstract* - Web applications have become an integral part of our daily lives, transforming various industries, and enabling smooth online interactions. The increasing number of web applications has also led to significant security challenges. People with malicious intent continuously exploit weaknesses in these web applications, posing a risk to the confidentiality, integrity, and availability of web applications. The primary objective of this research is to develop a comprehensive system that automates the identification and mitigation of vulnerabilities, prevention of threats, assessment of risks, and management of user access in web applications. This system uses advanced technologies like machine learning, runtime application self-protection (RASP), and risk calculation algorithms to take a well-rounded approach to web application security.**

This research project presents a comprehensive system that automates web application security, addressing the challenges posed by evolving threats. By utilizing advanced technologies and combining various security elements, the system offers a strong and effective solution to improve the security of web applications. This ensures their ability to withstand and maintain their integrity in today's interconnected digital world.

Keywords: Authentication, Threat, Risk, RASP, Vulnerabilities.

I. INTRODUCTION

This research project focuses on developing an advanced solution that automates and strengthens the security of web applications [1]. It is a combination of outcomes from four different research areas. By leveraging cutting-edge technologies such as machine learning [2], runtime application self-protection (RASP) [3], and risk calculation algorithms, this system aims to provide a comprehensive and efficient approach to mitigating security risks in web applications.

This research reveals a notable shortage of vulnerability information sources, especially concerning the severity of identified vulnerabilities. Addition to that this research introduces a novel method for calculating risks, incorporating both the impact and financial aspects [4], which increase the accuracy of risk assessments. Ultimately this research paper outlines the importance of creating a comprehensive yet user-friendly security solution for web applications that remains affordable for any organization.

II. LITERATURE REVIEW

In past, Researchers evaluate the need for a web application security product by analyzing potential threats, studying the market landscape, gathering user feedback through surveys and interviews, assessing risks of security breaches, examining case studies, considering industry standards, conducting comparative analyses of existing solutions, and staying updated on emerging trends and threats.[5] Through this comprehensive approach, they determine the necessity of a new security product based on technical factors, user needs, market gaps, and potential risks [6].

In the early 2000s, the vulnerability detection and remediation process was primarily manual, with limited vulnerabilities to address. However, the increasing rate of vulnerability discovery led to the need for a more structured approach. Vulnerability Management (VM) emerged as a cyclical practice to identify, prioritize, and mitigate software vulnerabilities. Traditional vulnerability scanners had limitations, such as network disruption and large, overwhelming reports. In response, Nodeware, a novel vulnerability management platform, was developed by IGI in 2016 [7].

Nodeware revolutionized vulnerability management by adopting continuous scanning, individualized treatment of assets, and smart threat intelligence updates. Its approach optimized network utilization, maintained an up-to-date asset inventory, and focused on high-risk devices [7]. The platform offered real-time visibility into network risks, simplified remediation verification, and provided on-demand reports. Over the years, Nodeware has proven effective in offering partners and customers access to the latest vulnerability intelligence, enhancing network security.

ISSN (online): 2581-3048
Volume 7, Issue 9, pp 37-42, September-2023
https://doi.org/10.47001/IRJIET/2023.709004

The evolution of machine learning usage in security products has been remarkable. Initially, security solutions relied on signature-based detection to identify known threats [8]. However, as cyber threats became more complex, machine learning introduced advanced techniques like anomaly detection, behavioral analysis, and predictive modeling. These approaches enable security products to adapt and respond to new and emerging threats effectively [9].

The integration of deep learning and neural networks further enhanced security capabilities, enabling tasks like facial recognition and automated response. Nevertheless, the rise of machine learning also brought about challenges such as adversarial attacks, where attackers exploit vulnerabilities in the models themselves. Overall, machine learning has revolutionized the field of cybersecurity, offering proactive and adaptive defense mechanisms against the constantly evolving threat landscape [10].

The Center for Internet Security (CIS) benchmarks were created to provide organizations with security configuration guidelines [11]. Developed by the nonprofit Center for Internet Security, these benchmarks offer recommendations for enhancing the security of different technology platforms. They're the result of collaborative efforts involving security experts and industry stakeholders, covering areas like operating systems, databases, network devices, and web applications.

Over the past two decades, the journey of multi-factor authentication (MFA) and its predecessor, two-factor authentication (2FA), reflects a transformation from niche tools to user-friendly solutions. While 2FA's origins date back to the 1990s, it only gained traction in the mid-2000s due to consumer resistance to its inconvenience and the prevailing belief that passwords sufficed. However, the advent of smartphones and their integration into business processes led to the adoption of more accessible 2FA methods, such as SMS-based authentication. The 2000s and 2010s saw a shift driven by alarming data breaches that targeted private industry, individuals, and government entities. High-profile breaches like Sony Pictures and the U.S. OPM prompted President Obama's call for stronger protection, giving rise to initiatives like the #Turnon2FA campaign and the integration of biometric authentication techniques. MFA, employing three authentication factors, emerged as a more robust solution, though it's not impervious to emerging threats like data breaches and SIM swapping. This underscores the need for continued evolution in MFA to safeguard businesses effectively [12].

III. METHODOLOGY

The purpose of this research project is to develop an automated security solution for web applications. The aim is to streamline the process of securing web applications by integrating various components that collectively enhance the application's security posture. This solution aims to tackle the challenges faced by web application owners in effectively managing and preserving the security of their applications. Additionally, it will provide valuable information about the overall risk environment. The solution comprises four key components that work together to automate web application security.

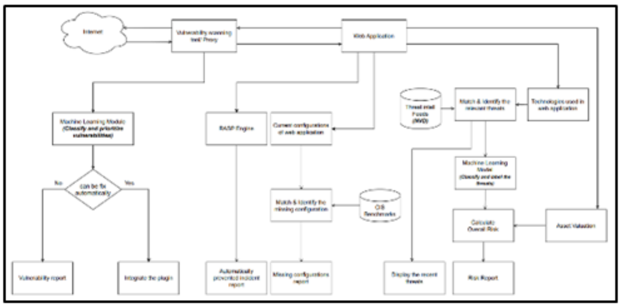

Internet
Vulnerability scanning tool/ Proxy
Web Application
Machine Learning Module (Classify and prioritize vulnerabilities)
RASP Engine
Current configurations of web application
Threat intel Feeds (NVD)
Match & identify the relevant threats
Technologies used in web application
Machine Learning Model (Classify and label the threats)
No
can be fix automatically
Yes
Match & identify the missing configuration
CIS Benchmarks
Calculate Overall Risk
Asset Valuation
Vulnerability report
Integrate the plugin
Automatically prevented incident report
Missing configurations report
Display the recent threats
Risk Report

Figure 1: System High Level Diagram (Component 1,2,3)

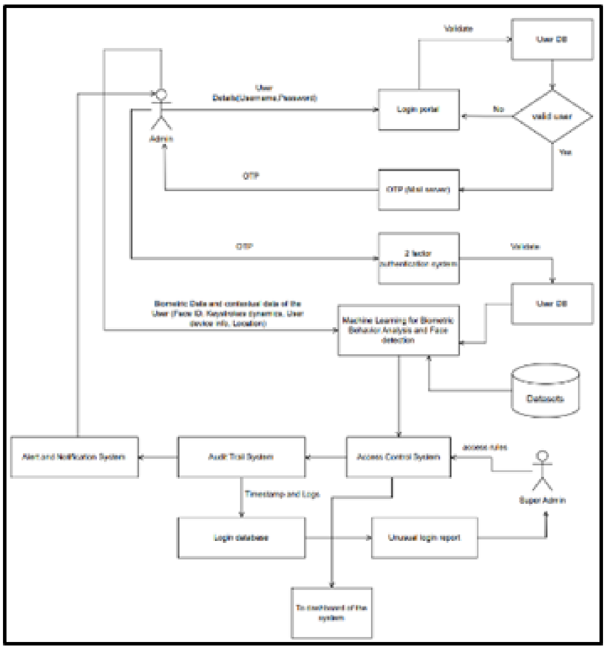

Validate
User DB
User Details(Username,Password)
Login portal
No
valid user
Admin
Yes
OTP
OTP (Mail server)
OTP
2 factor authentication system
Validate
User DB
Biometric Data and contextual data of the User (Face ID, Keystrokes dynamics, User device info, Location)
Machine Learning for Biometric Behavior Analysis and Face detection
Datasets
Alert and Notification System
Audit Trail System
Access Control System
access rules
Super Admin
Timestamp and Logs
Login database
Unusual login report
To dashboard of the system

Figure 2: System High Level Diagram (Component 4)

In the vulnerability identification component, web applications are on boarded to our system by providing details such as IP, web app category, financial value, and technologies used. Subsequently, the web application undergoes scanning with the assistance of a third-party scanner, and the scan results are integrated into the system. Using data from the National Vulnerability Database (NVD)

ISSN (online): 2581-3048

Volume 7, Issue 9, pp 37-42, September-2023

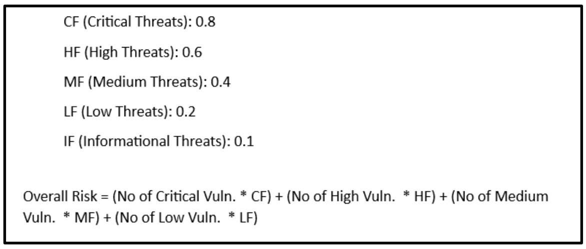
https://doi.org/10.47001/IRJIET/2023.709004

spanning from 1999, a machine learning model powered by natural language processing is trained to classify vulnerabilities.[13] These categorized vulnerabilities are then divided into two groups: those that can be automatically fixed and those that require manual intervention. For the automatically fixable vulnerabilities, we offer a plug in that can be seamlessly integrated into the web application by its owners. On the other hand, a report containing the details of the manually fixable vulnerabilities is sent to the web application owners for further action.

In the second component, Configurations of the web applications are captured and compared against the benchmarks set by the Center for Internet Security (CIS) to identify any configurations that are absent or incomplete. Any missing configurations are then communicated to the owners of the web application, who can take appropriate measures to address them. Addition to that, Runtime Application Self-Protection (RASP) engine will be given as an additional security feature since we cannot directly integrate with the customer's web application as we do not have the access to the source code. This will help web application owners to identify threats that specifically target the web application, including DOS attacks and DDOS attacks.[14] Web application owners can integrate it to their web application to enhance the security.

The third component revolves around calculating the risks associated with web applications. By utilizing the NVD API, real-time threat data is collected, and relevant threats are identified based on the web application's technology stack, which is obtained during the on boarding phase. An ML model trained using the NVD database is employed to classify these identified threats. Then these classified threats are used in the calculation process, which considers the severity and financial value of the web application, using a specific formula. The resulting list of threats and the corresponding risk value are presented on a monitoring dashboard for easy observation and tracking.

For the mentioned risk calculations, we invented a accurate formula which considers both the financial value of the web application and risk levels as below.



CF (Critical Threats): 0.8

HF (High Threats): 0.6

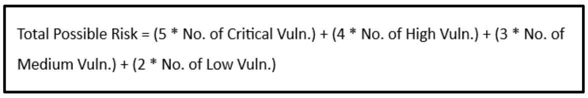
MF (Medium Threats): 0.4

LF (Low Threats): 0.2

IF (Informational Threats): 0.1

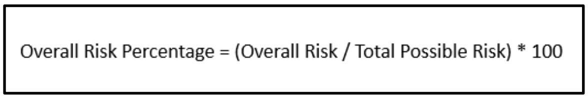
Overall Risk = (No of Critical Vuln. * CF) + (No of High Vuln. * HF) + (No of Medium Vuln. * MF) + (No of Low Vuln. * LF)

Figure 3: Overall Risk Formula



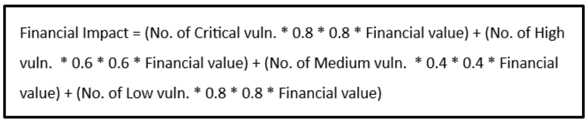
Total Possible Risk = (5 * No. of Critical Vuln.) + (4 * No. of High Vuln.) + (3 * No. of Medium Vuln.) + (2 * No. of Low Vuln.)

Figure 4: Total Possible Risk Formula



Overall Risk Percentage = (Overall Risk / Total Possible Risk) * 100

Figure 5: Overall Risk Percentage Formula



Financial Impact = (No. of Critical vuln. * 0.8 * 0.8 * Financial value) + (No. of High vuln. * 0.6 * 0.6 * Financial value) + (No. of Medium vuln. * 0.4 * 0.4 * Financial value) + (No. of Low vuln. * 0.8 * 0.8 * Financial value)

Figure 6: Financial Impact Formula

The authentication component focuses on granting access to administrators. Administrators register by providing their username, password, and additional contextual information as location, device usage, and work shift. When logging in, administrators are prompted to enter their username and password. Then, a two-factor authentication process is initiated, involving the generation and verification of a one-time password (OTP). In case the OTP authentication fails, further validation is conducted using factors such as face ID and contextual details. The super admin has the authority to manage access control lists (ACLs) for administrators. Login details are stored in the database for auditing purposes, and the super admin can monitor failed login attempts.[15] Once successfully logged in, administrators receive a message confirming their successful login.

IV. DISCUSSION

4.1 Comparison with Previous Research

The proposed approach addresses several limitations found in existing approaches to web application security. Existing solutionslike DarkTrace, VectraAI, Palo Alto Networks, and others often lack integration with security tools, transparency, complexity, and visibility. In contrast, the proposed solution offers seamless integration with a wide range of security tools, providing organizations with the ability to leverage their existing infrastructure effectively. Moreover, it offers a wide range of customization options that allow organizations to customize the solution to meet their specific needs. The proposed approach simplifies complexity and enhances visibility, providing organizations with a clearer understanding of their web application security posture. These advancements differentiate the proposed approach, making it a more comprehensive, adaptable, and user-friendly solution for web application security.

ISSN (online): 2581-3048
Volume 7, Issue 9, pp 37-42, September-2023
https://doi.org/10.47001/IRJIET/2023.709004

4.2 Addressing Research Questions

The proposed solution effectively addresses the research questions related to each component. For the first component, the research question revolves around automating vulnerability identification and classification. The proposed solution uses a machine learning model trained on NVD data to automatically label vulnerabilities, addressing this research question. The second component focuses on enhancing real-time threat detection and response. By providing a RASP engine as an added service, the proposed solution addresses the research question and provides an innovative approach to web application security. The third component aims to improve web application risk assessment. By utilizing machine learning, real-time data analysis, and a holistic approach to risk calculation, the proposed solution effectively addresses the research question. The fourth component focuses on secure access and authentication. By implementing a robust login system with two-factor authentication and capturing login details for auditing, the proposed solution effectively addresses the research question.

4.3 Methodological Considerations and Limitations

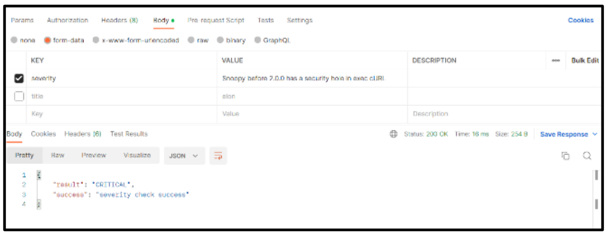
The effectiveness of the machine learning models used in the solution heavily relies on the quality and representativeness of the training data, which is sourced from the NVD database. If the NVD data does not adequately cover the variety of vulnerabilities and threats, it may affect the accuracy of vulnerability labelling and threat identification. Secondly, the proposed solution relies on third-party vulnerability scanners, which may have their own limitations in terms of coverage, accuracy, and timeliness. It is crucial to select reliable and up-to-date scanners to ensure the effectiveness of vulnerability identification. Lastly, the risk calculation component relies on a special formula that considers severities and financial value. The accuracy of the risk calculation depends on the relevance and reliability of these factors. It is important to ensure that the formula adequately captures the potential impact and likelihood of risks to provide meaningful risk assessments.

4.4 Novelty and Contribution

The proposed solution brings novelty and contributes to the field of web application security in several ways. It introduces an integrated approach that combines vulnerability identification, real-time threat detection, risk assessment, and secure access systems into a one solution. The integration of machine learning, and threat intelligence feeds enhances the accuracy and effectiveness of the solution. The customization options and emphasis on context-aware security set the proposed solution apart from existing approaches. Overall, the novelty and contribution of the proposed solution lie in its comprehensive, adaptable, and user-friendly nature, addressing the limitations of existing approaches and providing organizations with a more effective web application security solution.

4.5 Results

In the result section of this research paper, we emphasize how our Machine Learning (ML) module, particularly using Natural Language Processing (NLP) methods, identifies the severity of vulnerabilities and threats. There was a big challenge caused by not having enough past data for certain severity categories. This lack of data, even in the National Vulnerability Database (NVD), which is the official source for vulnerability and threat info, demonstrate how complex is the process of building an ML module for vulnerability identification product using open sources.



	KEY	VALUE	DESCRIPTION
☑	severity	Snoopy before 2.0.0 has a security hole in exec cURL	
□	title	elon	
	Key	Value	Description

Status: 200 OK Time: 16 ms Size: 254 B

```
{
    "result": "CRITICAL",
    "success": "severity check success"
}
```

Figure 7: Severity Prediction Results

Despite the data limits, this severity prediction ML module shows good accuracy in its predictions. This success shows how strong our approach is and how our underlying algorithms support it. By using the available data effectively, our model shows high proficiency in correctly categorizing severity levels according to the testing.

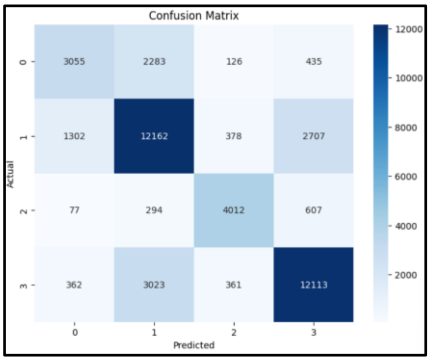
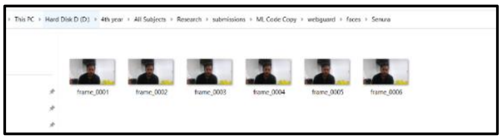


Figure 8: Severity Prediction Confusion Matrix

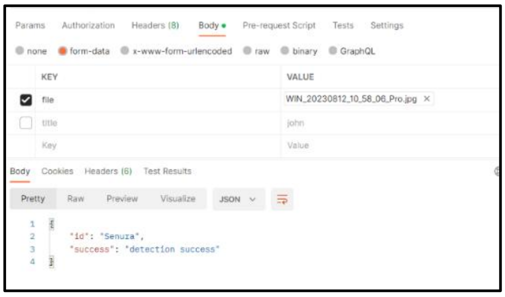
And this system accurately captures the CVE information using an API for the given technologies using them as keywords.

Also, our research introduces a special face detection module that uses a special facial recognition library. This module is good at identifying users, making our system more secure.



This PC › Hard Disk D (D:) › 4th year › All Subjects › Research › submissions › ML Code Copy › webguard › faces › Senura

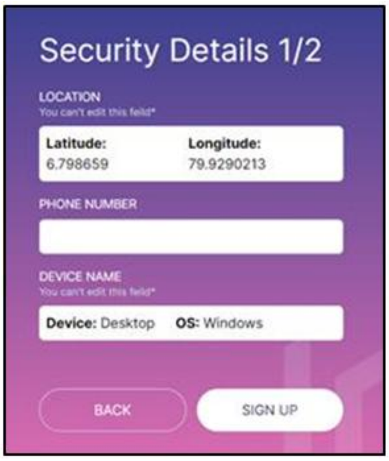
Figure 9: Face Prediction Frame Splitting



	KEY	VALUE
☑	file	WIN_20230812_10_58_06_Pro.jpg
□	title	john
	Key	Value

```
{
    "id": "Senura",
    "success": "detection success"
}
```

Figure 10: Face Prediction Results



Security Details 1/2

LOCATION
You can't edit this field*

Latitude: 6.798659
Longitude: 79.9290213

PHONE NUMBER

DEVICE NAME
You can't edit this field*

Device: Desktop
OS: Windows

Figure 11: Context Data Capture Results

In terms of matching context data, our system is showing good results at finding where users are and getting detailed info about their devices.

V. CONCLUSION

In conclusion, this research paper introduces the product named WebGuardian, designed to enhance the security of web applications. WebGuardian consists of four main parts which are the use of machine learning to find and classify vulnerabilities, automatically determine the correct settings for web applications using CIS benchmarks and a RASP as on demand security tool, provide real-time tracking of threats tailored to the specific technology stack of web applications, and offer a robust user authentication system that employs multi-factor authentication.

The paper also addresses the challenge of acquiring sufficient data to train the machine learning component for identifying web application vulnerabilities. Despite this challenge, we have diligently worked to automate the solution to the best of our abilities. We acknowledge that there are certain aspects that we haven't fully automated due to current limitations. However, as technology advances, there is potential for these limitations to be addressed, ultimately leading to the complete automation of web application security.

In the dynamic realm of online threats and security, WebGuardian represents a significant advancement in powering the safety of web applications. This paper not only outlines the functionalities of WebGuardian but also gives an idea to future in which technological progress could enable us to achieve full automation of web application security.

REFERENCES

[1] Hasty Atashzar, Atefeh Torkaman, Marjan Bahrololum, Mohammad H. Tadayon , "A Survey on Web Application Vulnerabilities and Countermeasures," ResearchGate, 2016.

[2] Nilaykumar Kiran Sangani and Haroot Zarger, "Machine Learning in Application Security," INTECH, Dubai, 2021.

[3] Y. Wang, "Vulnerability analysis and improvement of RASP technology," International Symposium on Advances in Informatics, Electronics and Education, Beijing, 2022.

[4] Anton Konev , Alexander Shelupanov *, Mikhail Kataev , Valeriya Ageeva and Alina Nabieva, "A Survey on Threat-Modeling Techniques: Protected Objects and Classification of Threats," Symmetry, MDPI, Tomsk, 2022.

[5] H. Abusaimeh, M. Shkoukani, "Survey of Web Application and Internet Security Threats," Semantic Scholar, 2012.

[6] StackHawk, "Importance of Web Application Security: Three Benefits," Stackhawk, 17 November 2022. [Online]. Available: https://www.stackhawk.com/blog/importance-of-web-application-security-three-benefits/. [Accessed 10 July 2023].

[7] B. Drake, "igicybersecurity," 23 September 2020. [Online]. Available: https://blog.igicybersecurity.com/origins-and-evolution-of-vulnerability-management. [Accessed 10 July 2023].

[8] W. Chai, "A Timeline of Machine Learning History," WhatIs.com, 2020.

[9] Deva, "Evolution Of Natural Language Processing(NLP)," Medium, 2021.

[10] K. Joshi, "What is Classification in Machine Learning and Why is it Important?," emeritus, 2022.

[11] "CIS Benchmarks List," [Online]. Available: https://www.cisecurity.org/cis-benchmarks. [Accessed 16

July 2023].

[12] R. d. Fremery, "The Evolution of Multi-Factor Authentication," LasrPass, 21 December 2021. [Online]. [Accessed 15 July 2023].

[13] Marian Gawron(B), Feng Cheng, and Christoph Meinel, "Automatic Vulnerability Classification Using Machine Learning," ResearchGate, Potsdam.

[14] Petar ýisar* and Sanja Maraviü ýisar**, ResearchGate, Belgrade-Zemun, 2016.

[15] Sandeep kaur ,Gaganpreet kaur , Mohammad Shabaz , "A Secure Two-Factor Authentication Framework in Cloud Computing," Hindawi, Arba Minch, 2022.

AUTHORS BIOGRAPHY

Pasindu Bandara,
Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Methmi Ranasinghe,
Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Hansa Ranaweera,
Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Senura Rathnayake,
Undergraduate, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Amila Senarathne,
Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Kanishka Yapa,
Lecturer, Faculty of Computing, Sri Lanka Institute of Information Technology, Western Province, Sri Lanka.

Citation of this Article:

Aththanayaka P.A.G.P.B., Ranasinghe M.H., Ranaweera H.N.K, Rathnayake S.D., Amila Senarathne, Kanishka Yapa, “WebGuardian: Holistic Approach to Address Dynamic Web Application Threat Landscape” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 9, pp 37-42, September 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.709004

****