Savitribai Phule Pune University

Modern Education Society’s Wadia College of Engineering, Pune

19, Bund Garden, V.K. Joag Path, Pune – 411001.

Accredited by NBA and NAAC with ’A++’ grade

Department of Computer Engineering

A

Seminar Report

ON

Study and Critical Analysis of Web Application Firewall

T.E. (Computer)

SUBMITTED BY

Acharya Ayush Surendra

GUIDED BY

Dr.(Mrs.) S.K. Wagh

(Academic Year: 2023-24)

Date: 03/11/2023

Savitribai Phule Pune University

Modern Education Society’s Wadia College of Engineering, Pune

19, Bund Garden, V.K. Joag Path, Pune – 411001.

Accredited by NBA and NAAC with ’A++’ grade

Department of Computer Engineering

CERTIFICATE

This is to certify that

Acharya Ayush Surendra

has been completed Seminar entitled

Study and Critical Analysis of Web Application Firewall

As a partial fulfillment of the Third Year of Bachelor degree in “Computer Engineering” as prescribed by the Savitribai Phule Pune University in TE COMP I in the Semester - I of academic year 2023-24.

Seminar Guide	**HOD**
Dr.(Mrs.) S.K.Wagh	**Dr.Nuzhat F. Shaikh**

Place: Pune

Date: November 03, 2023

Date: November 03,2023

Abstract

This literature survey offers a concise yet comprehensive exploration of Web Application Firewalls (WAFs) and their pivotal role in modern cybersecurity. It covers the historical development, deployment best practices, and introduces a robust framework for HTTP attack detection.

In the ”Bot Scanning” phase, it emphasizes the importance of swift identification and response to malicious bots and scanners, underlining the WAF’s role in enhancing web application security.

This resource is a valuable reference for those seeking a deeper understanding of WAFs and their practical application in safeguarding web applications in today’s evolving threat landscape.

Keywords:

WAF: Web Application Firewall

DoS: Denial of Service

DDoS: Distributed Denial of Service

XSS: Cross Site Scripting

OWASP: Open Web Application Security Project

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude towards my seminar guide **Dr. (Mrs.) S.K.Wagh** for her support, continuous guidance and being so understanding and helpful throughout the seminar.

I furthermore thank Computer Department HOD **Dr. (Mrs.) N.F.Shaikh** to encourage me to go ahead and for continuous guidance. I also want to thank **Dr. (Mrs.) S.P. Deore** and **Prof. S.S. Raskar** for all their assistance and guidance for preparing the report.

I would like to thank all those who have directly or indirectly helped me for the completion of the work during this seminar.

Acharya Ayush Surendra
TE COMP 1
F21111036

Contents

Chapter 1

Introduction

1.1 Overview

1. Firewall as a Network Security Measure:

A firewall is a critical component in network security. It acts as a barrier between your internal network and the external internet. Its primary aim is to filter and control incoming and outgoing network traffic based on an organization's previously established security policies.

2. Single Choke Point:

Firewalls serve as a single choke point for enforcing security policies. By having all traffic pass through the firewall, you can apply consistent security measures and monitor all network traffic at a centralized location. This simplifies the management of security policies.

3.Layer of Defense:

Firewalls provide an additional layer of defense for your network. They can block malicious traffic, detect intrusion attempts, and prevent unauthorized access to your internal systems. This layered approach helps to minimize the risk of security breaches.

4.Web Application Firewall (WAF) Introduction:

A Web Application Firewall is a specialized security solution that focuses on protecting web applications specifically. It does this by filtering and monitoring HTTP requests and responses. Unlike a traditional firewall, which works at the network level, a WAF operates at the application layer, making it more effective in protecting web applications.

5.Protection from Online Threats:

WAFs are designed to safeguard web applications from a wide range of online threats. These threats include SQL injection, cross-site scripting (XSS), cross-site request forgery (CSRF), and more. They identify and block these attacks before they can exploit vulnerabilities in your web applications.

6.Data Breach Prevention:

WAFs play a crucial role in preventing data breaches. By detecting and blocking malicious requests, they can prevent unauthorized access to sensitive data stored within web applications. This is vital for protecting both user data and an organization's proprietary information.

7.Continuous Monitoring and Reporting:
A critical aspect of WAFs is continuous monitoring and reporting. They provide real-time insights into web traffic, including traffic patterns, potential threats, and security incidents. This data is valuable for identifying and mitigating emerging threats.

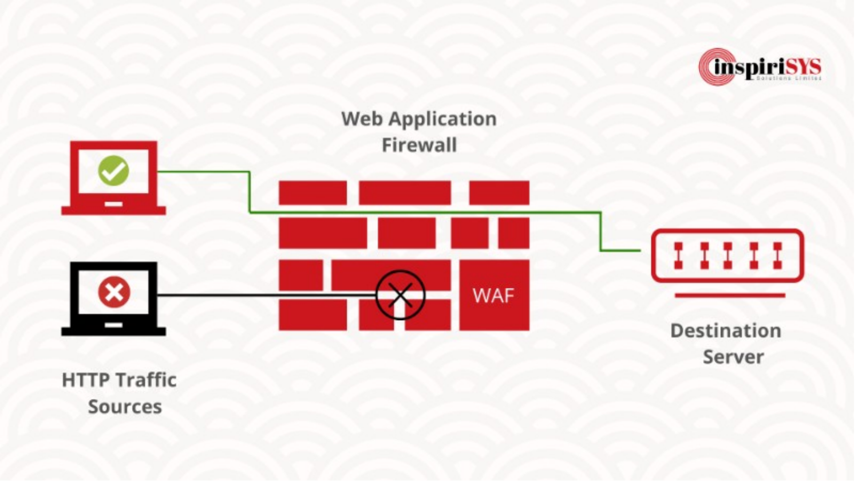

Figure 1.1: A Web Application Firewall

1.2 Introduction

The ubiquity of internet has resulted in a momentous growth in its usage all across the globe resulting in roughly half of humanity having an internet connection (Max Roser and Ortiz-Ospina, 2020). This perpetual digital transformation in human lives has intrigued end users, corporates, small businesses, large organizations and in fact everyone to share their data and information via websites. The number of websites in the world have surpassed 1.7 Billion and are still growing. These websites are connected to backend databases in real time where the later, although not visible or directly accessible to the user are updated, parsed and accessed by the web application(s) as a result of some user activity.Besides catalyzing this massive e-transformation, the overwhelming use of webpages and reliance on websites has also stirred the attackers to compromise these sites in order to get hold of important data that might be of some financial or reputational gain. According to 2019 Verizon Data Breach Investigations Report, web attacks are the most prevalent sort of cyber-attacks, happening once every 39 s . These attacks often result in massive data, financial and reputational loss and are further aggravated due to the absence of cybercrime regulations in many parts of the world.

Security experts have developed ways and means to tackle the grave threat of chronic web attacks. On the other hand, attackers have equipped themselves with more sophisticated approaches to circumvent these defensive mechanisms. OWASP (Open Web Application Security Project) has listed the top 10 web attacks (The Open Web Application Security Project, 2017) that pose a great deal of challenge to security experts . These attacks have accentuated the need to not only go for secure coding in web applications but to also rely on security mechanisms that can help thwart such attacks in real time.

One obvious choice is to use a WAF (Web Application Firewall), as shown in Fig. 2, that acts as a layer of defense standing between the attacker and the website. WAFs analyze user requests in order to filter the illicit traffic out of the routine one. This analysis of the HTTP requests is based on a set of predefined rules designed to discriminate the malicious traffic.

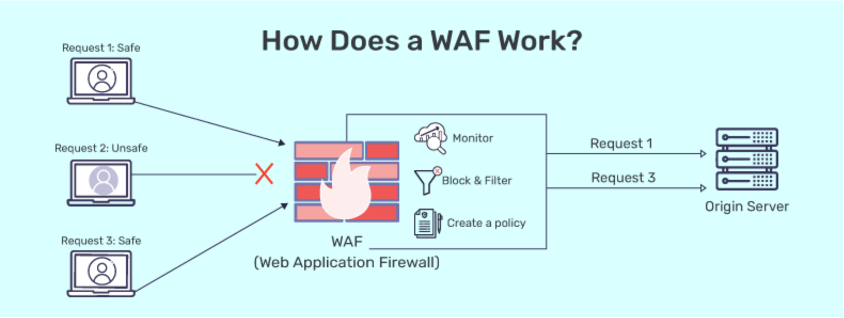

Figure 1.2: A Web Application Firewall

WAFs perform better against these scanners because of known signatures and payloads, but often fail against experienced hackers who prefer writing their own pieces of code. Even the most popular WAFs such as ModSecurity (Trustwave, 2020) fails in detecting such HTTP requests. These issues with WAFs are either because of misconfigurations, their rule-based nature, lack of semantic analysis and behavioral detection, failure against permutation of known attacks and no protection against zero-day attacks . Moreover, tired of false positives, many security administrators keep their WAFs on the Alert Mode, in which attacks are never blocked in real time. According to a survey conducted by Ponemon institute (Vicente, 2019), only 9 percent of the WAF users say that their solutions have never been compromised, whereas, 65 percent complain that attackers can bypass their solutions without much of an effort.

Therefore, development of any web attack detection system must consider achieving certain security goals. Firstly, the system should have very high accuracy and precision values. The proposed web attack detection framework achieves all these security goals by addressing all performance and optimization issues related to Deep Learning based classifiers. The main research contributions of this research work are listed as follows:

1. A novel framework has been proposed which is based on a hybrid approach which nests Deep Learning model with a Cookie Analysis Engine for web attacks detection, mitigation and attacker profiling in real time.

2. The Convolutional Neural Network based deep learning classifier is trained using HTTP request parameters like Content Type, Length, Requested URL and Data etc. on a large dataset specifically generated for this purpose.

3. A Cookie Analysis Engine that has been designed to check all incoming cookie(s) for integrity failures, mutations and failed sanitization checks and informs the user about probable privacy infringement by third party cookies.

4. An efficient framework which saves useful processing time when deployed in real time as the attacker profiling feature limits the execution of deep learning classifier for every incoming HTTP request without degrading attack detection capability. The proposed framework gives an accuracy of 99.94dataset and 98.74% on a publicly available benchmark dataset.

Chapter 2

Motivation

The study and critical analysis of Web Application Firewalls (WAFs) are of utmost importance in the realm of cybersecurity. In an era where web applications serve as the backbone of digital interactions, the security of these applications has never been more critical. WAFs offer a specialized defense against a multitude of web-based threats, from SQL injection to cross-site scripting attacks. To truly understand their efficacy and applicability, it is imperative to delve into a comprehensive examination of their features, limitations, and evolving capabilities. By critically analyzing WAF technology, we can ensure the robust protection of web applications and data, safeguard user trust, and mitigate the ever-evolving threats that target our digital landscape.

The Motivation for this topic relates to :-

1. Rising Cybersecurity Concerns :

Cybersecurity threats have seen a significant uptick in recent years, with attackers becoming increasingly sophisticated and relentless in their efforts. The exponential growth of the digital landscape has expanded the attack surface, making organizations vulnerable to a wide range of threats.

2. Web Application Vulnerabilities:

Web applications often harbor vulnerabilities that malicious actors can exploit. Common vulnerabilities include SQL injection, which allows attackers to manipulate databases, and cross-site scripting (XSS), which enables the injection of malicious code into web pages, potentially impacting users.

3. Criticality of Web Applications:

Web applications underpin a wide array of essential services, from online banking and e-commerce to healthcare and government services. They are the primary interface through which users interact with organizations, making their security a matter of paramount importance.

4. Data Breach Consequences:

Data breaches can have severe consequences for organizations. They result in financial losses due to remediation efforts, legal repercussions, and regulatory fines. Furthermore, they erode the trust of customers and clients, potentially leading to reputational damage.

5. Regulatory Requirements:

Regulations like the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and Payment Card Industry Data Security Standard (PCI DSS) impose strict data protection requirements. Non-compliance can lead to significant fines and legal penalties.

6. Emerging Threats:

Cyber threats are continually evolving. Attackers develop new techniques and exploit vulnerabilities. To counter these emerging threats, WAFs are incorporating machine learning and AI, enabling them to detect and mitigate previously unseen attacks.

7.Relevance for Various Sectors:

WAFs are highly relevant across industries. For example, financial institutions need to secure customer data, while healthcare providers must protect patient records. E-commerce platforms require defenses against online fraud and data theft.

Chapter 3

Literature Survey

The comprehensive understanding of Web Application Firewalls (WAFs) and their role in contemporary cybersecurity is of paramount significance in the digital age. As the utilization of web applications continues to surge across various sectors, so too does the complexity of potential threats and vulnerabilities. This literature survey serves as a foundational component of our research, as it is essential to delve into the wealth of existing knowledge and insights concerning WAFs. Our objective is to examine and analyze the current state of the field, identify gaps in the literature, and leverage this knowledge to further explore and contribute to the realm of WAF technology.

This Literature survey endeavors to offer a comprehensive overview of the historical development of WAF technology, key features and capabilities, their interplay with the Open Web Application Security Project (OWASP) Top Ten vulnerabilities, and the challenges they address. Furthermore, it will examine the evolving landscape of WAFs in the face of emerging trends and innovations in cybersecurity, such as the application of artificial intelligence and machine learning.

By synthesizing and critically assessing the existing body of knowledge surrounding WAFs, this literature survey lays the groundwork for our subsequent research objectives. We aim to address unexplored dimensions and contribute novel insights that enhance the effectiveness and practicality of Web Application Firewalls in protecting web applications in an ever-evolving threat landscape.

3.1 Research Paper I

Title : A critical review of the techniques used for anomaly detection of HTTP-based attacks: taxonomy, limitations and open challenges.

Authors : Jesus E. Verdiogo, Rafael Alonso , Antonio Alonso, German Madinabeitia

Journal : Computer and Security Journal Volume 124

Year Of Publication : 2023

Remarks :-

The introduction of a novel framework for detecting HTTP-based attack anomalies signifies a significant advancement in web security. This framework introduces innovative techniques for identifying web-based attacks and addresses key challenges found in the existing literature. It emphasizes the need for standardized datasets, which serve as the foundation for accurate anomaly detection, essential for training and testing machine learning models. Additionally, the framework highlights the importance of robust detection methods capable of adapting to evolving attack strategies and evasion tactics used by attackers. It introduces standardized evaluation metrics for fair comparisons and precise measurements of detection method effectiveness.

Furthermore, the framework acknowledges the challenges of scaling and obsolescence in real web service applications, underscoring the need for adaptable detection methods. In summary, this framework not only introduces innovative techniques but also provides solutions to critical web security challenges, contributing to ongoing efforts to enhance web service security.

3.2 Research Paper II

Title : A deep learning assisted personalized deception system for countering web application attacks.

Authors : Waleed Shahid, Baber Aslam, Haider Abbas, Hammad Afzal, Saad Khalid

Journal : Information Security and Applications Journal Volume 67

Year Of Publication : 2022

Remarks :-

The high interaction web deception system provides a comprehensive web security solution that surpasses traditional measures. By engaging attackers in a deceptive environment and utilizing advanced deep learning techniques, it ensures early detection of potential security threats. Additionally, it incorporates a cookie analysis engine for further enhancing security and actively profiles attackers in real-time, allowing for a proactive response while minimizing false positives.

Efficiency and scalability are at the forefront of this system's design, optimizing resource usage and adapting to evolving web environments. Its compatibility with Internet of Things (IoT) networks positions it as a valuable asset for securing these often-vulnerable devices, contributing significantly to overall cybersecurity.

In summary, the high interaction web deception system combines advanced technology with efficient design to offer a comprehensive and adaptable defense strategy for web environments. Its deep learning-based approach, cookie analysis, and real-time attacker profiling make it a powerful tool in the ongoing battle against web-based threats, ensuring the security of web applications and data.

3.3 Research Paper III

Title : An enhanced deep learning based framework for web attacks detection, mitigation and attacker profiling.

Authors : Waleed Shahid, Baber Aslam, Haider Abbas, Hammad Afzal, Saad Khalid

Journal : Networks and Computer Applications Journal Volume 198

Year Of Publication : 2022

Remarks :-

The hybrid framework at the forefront of web security leverages deep learning and cookie analysis for real-time web attack detection and mitigation. It excels in promptly identifying a wide range of web attacks, crucial for preventing potential damage and data breaches in a constantly evolving threat landscape.

Furthermore, the framework's efficient real-time attacker profiling minimizes resource requirements, enhancing overall system efficiency and cost-effectiveness. By conserving processing resources and reducing false alarms, it becomes a versatile solution adaptable to various web application environments.

Additionally, the hybrid framework provides comprehensive protection against a broad spectrum of web attacks, including those involving POST requests. Its adaptability ensures it can recognize and counter both known and unknown attack patterns, enhancing web application security against evolving threats. In conclusion, this advanced hybrid framework is a significant step in fortifying web application security and reducing the risk of cyber threats, making it a valuable asset for organizations.

3.4 Research Paper IV

Title : DNS Intrusion Detection (DID) — A SNORT-based solution to detect DNS Amplification and DNS Tunneling attacks.

Authors : Sanjay Adiwal, Balaji Rajendran, Pushparaj Shetty, Sithu Sudarsan

Journal : Franklin Open Journal Volume 2

Year Of Publication : 2023

Remarks :-

The integration of DNS Intrusion Detection (DID) with SNORT is a groundbreaking advancement in DNS attack detection and mitigation. This integrated system effectively addresses the challenges posed by DNS attacks, with a particular focus on amplification and tunneling attacks, which are significant threats to network security.

The collaboration between DID and SNORT is a notable feature of this system, with SNORT serving as a renowned open-source Intrusion Detection System (IDS). This partnership enhances the system's capacity to address DNS-specific threats, thereby boosting its overall effectiveness in safeguarding network resources.

DID's innovation lies in its use of tailored IDS signatures for the detection of empirical DNS attacks. These signatures are meticulously designed to identify patterns and characteristics associated with DNS amplification and tunneling attacks, significantly reducing false positives. In conclusion, the integration of DNS Intrusion Detection with SNORT offers a powerful solution for DNS attack detection, with a specific focus on amplification and tunneling attacks.

3.5 Research Paper V

Title : Dynamic Web Application Firewall detection supported by Cyber Mimic Defense approach.

Authors : Mariusz Sepczuk

Journal : Networks and Computer Applications Journal Volume 213

Year Of Publication : 2023

Remarks :-

The integration of Cyber Mimic Defense with Web Application Firewalls (WAF) is a pivotal advancement in web application security, addressing the challenges of detecting unknown and evolving cyber threats. This innovative approach significantly enhances detection capabilities, surpassing traditional security measures. Cyber Mimic Defense introduces unpredictability, making it difficult for attackers to succeed.

Experiments validating this concept have shown promising results, notably increasing the detection of unknown attacks and dynamically adapting WAF rules in real-time to counter emerging threats. Additionally, this concept reduces malicious web traffic at the edge, improving system efficiency and fortifying web application security.

In conclusion, the integration of Cyber Mimic Defense with WAF strengthens web application security by focusing on detecting unknown attacks, enhancing adaptability, and reducing malicious web traffic. This concept is a major step forward in safeguarding web applications against the ever-evolving threat landscape.

3.6 Research Paper VI

Title : HTTP request pattern based signatures for early application layer DDoS detection: A firewall agnostic approach.

Authors: Amit Praseed, P. Santhi Thilagam

Journal: Information Security and Applications Journal Volume 65

Year Of Publication : 2022

Remarks :-

The Early Detection Module (EDM) is a significant advancement in web security, particularly in countering Application Layer Distributed Denial of Service (AL-DDoS) attacks. These attacks pose a substantial threat to web services, targeting the application layer. EDM focuses on improving AL-DDoS attack detection by using HTTP request patterns as signatures, a novel and effective approach.

A key challenge in combating AL-DDoS attacks is reducing detection latency, as these attacks can overwhelm web services rapidly. EDM addresses this challenge by employing Sample Entropy, which enhances the detection of temporal patterns, identifying similarities in the timing and sequences of HTTP requests. This capability is invaluable for early detection, ensuring the availability and security of web services in the face of evolving and sophisticated attack strategies.

In summary, the introduction of the Early Detection Module, with its focus on HTTP request patterns and Sample Entropy, signifies a significant milestone in web security. EDM offers a crucial solution for reducing detection latency and safeguarding web services against the ever-present and rapidly evolving threat of AL-DDoS attacks.

3.7 Research Paper VII

Title : Hybrid unsupervised web-attack detection and classification – A deep learning approach.

Authors: Seema Pillai, Dr. Anurag Sharma

Journal: Computers Standards and Interfaces Journal Volume 86

Year Of Publication : 2023

Remarks :-

This research introduces a groundbreaking approach that combines unsupervised deep learning methods with a sophisticated DBM-BiLSTM-based classification model. Its core strength lies in its autonomous web attack detection and efficient classification, achieving an impressive F1-Score of 97.62, highlighting its exceptional performance.

The real-world significance of this approach is evident in the ever-present threat landscape where web applications face a constant barrage of attacks. This technique offers a robust defense mechanism, not only identifying attacks but also classifying them with remarkable accuracy, enhancing web application security significantly.

Compared to existing methodologies, this approach demonstrates its superiority by providing enhanced protection for web applications in an environment where threats continually evolve. The results of this research indicate that this approach can set new standards in web application security, offering a highly accurate defense against a wide array of attack vectors. With its outstanding F1-Score of 97.62, it represents a significant leap forward in securing web applications against malicious attacks.

3.8 Research Paper VIII

Title : Software-defined network-based HTTP flooding attack defender.

Authors: Reeza Mohammadi, Chaggan Lal, Mauro Conti, Lokesh Sharma

Journal: Computers & Electrical Engineering Journal Volume 101

Year Of Publication : 2022

Remarks :-

This research presents SHFD, a Software-Defined Networking (SDN) based defense against HTTP flooding attacks, a significant threat to web services in the era of Distributed Denial of Service (DDoS) attacks.

SHFD operates as a proactive sentinel integrated into the SDN controller, tasked with real-time identification and counteraction of HTTP flooding attacks. What sets SHFD apart is its dynamic adaptability, making use of SDN's capability to configure networks on-the-fly. This adaptability results in a remarkable 13% improvement in detection time and a notable 29% increase in the number of malicious flows successfully blocked compared to existing methodologies.

The efficiency and effectiveness of SHFD can be attributed to its utilization of the SDN paradigm, outperforming traditional defenses by proactively detecting and mitigating HTTP flooding attacks. This research underscores the utility of it in enhancing web services' security against persistent threats, highlighting the crucial role that dynamic network configuration plays in securing web applications amid ever-evolving cyber threats.

3.9 Research Paper IX

Title : Study of Web Application Firewalls (WAFs).

Authors: Namit Gupta, Abakash Saikia , Dr. Dheeraj Sanghi

Journal: IIT Kharagpur Project Paper

Year Of Publication : 2007

Remarks :-

The significance of Web Application Firewalls cannot be overstated. As the digital landscape rapidly evolves, web applications face a growing array of cyber threats. Traditional security measures no longer suffice, as the focus shifts from network layer security to the complex realm of application layer security. This transition is driven by the increasing diversity and sophistication of cyberattacks that target vulnerabilities within the applications themselves.

WAFs play a pivotal role in this shifting landscape. They are adept at handling HTTP requests, a fundamental element of the internet's architecture. WAFs act as vigilant gatekeepers, meticulously scrutinizing and managing these requests. Their primary function is to filter incoming traffic, allowing legitimate requests to pass while swiftly identifying and thwarting suspicious or malicious ones.

In addition to handling HTTP requests, WAFs feature built-in checks, predefined rules, and validations. These checks ensure incoming requests meet established standards and security parameters. WAFs also allow administrators to create user-defined rules, customizing defense for individual web applications. In summary, WAFs are crucial in the ongoing fight against cyber threats, offering strong protection at the application layer, where many vulnerabilities are targeted.

Chapter 4

Methodology

4.1 Types Of Cyber Attacks

Cyberattacks encompass a broad spectrum of malicious activities designed to compromise the confidentiality, integrity, and availability of computer systems, networks, and data. These attacks are perpetrated by individuals, groups, or organizations with various motives, ranging from financial gain to political or ideological agendas. Understanding the different types of cyberattacks is crucial for implementing effective cybersecurity measures. Here are descriptions of some common cyberattacks:

1. Malware Attacks: Malware, short for malicious software, includes various types of software designed to infiltrate or damage computer systems. This category encompasses viruses, worms, Trojans, ransomware, and spyware. Viruses attach themselves to legitimate programs and replicate when executed, while worms self-replicate without user intervention. Trojans disguise themselves as legitimate software but often create backdoors for attackers. Ransomware encrypts files and demands a ransom, while spyware secretly collects data.

2. Phishing Attacks: Phishing attacks typically involve the use of deceptive emails, messages, or websites to trick users into revealing sensitive information such as login credentials, credit card numbers, or personal details. Email phishing targets a broad audience, while spear phishing customizes attacks for specific individuals or organizations. Vishing occurs through phone calls, and smishing through text messages.

3. Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) Attacks: These attacks aim to make a targeted system or network unavailable to users. In DoS attacks, a single system overwhelms the target, while DDoS attacks involve multiple compromised systems that flood the target with traffic.

4. Man-in-the-Middle (MitM) Attacks: MitM attacks occur when an attacker intercepts and potentially alters communications between two parties without their knowledge. This can lead to data theft or tampering.

5. SQL Injection: Attackers exploit vulnerabilities in web applications to manipulate a database. By injecting malicious SQL code, they can gain unauthorized access or manipulate data within the database.

6. Cross-Site Scripting (XSS) Attacks: In XSS attacks, malicious scripts are injected into websites, and when viewed by other users, these scripts can steal their data or perform other malicious actions.

7. Zero-Day Exploits: Zero-day exploits target unpatched vulnerabilities in software before developers can release security fixes, making them particularly dangerous.

8. Brute Force Attacks: Attackers employ brute force methods to repeatedly guess passwords or encryption keys, testing all possible combinations until they find the correct one.

4.2 Deployment of WAF

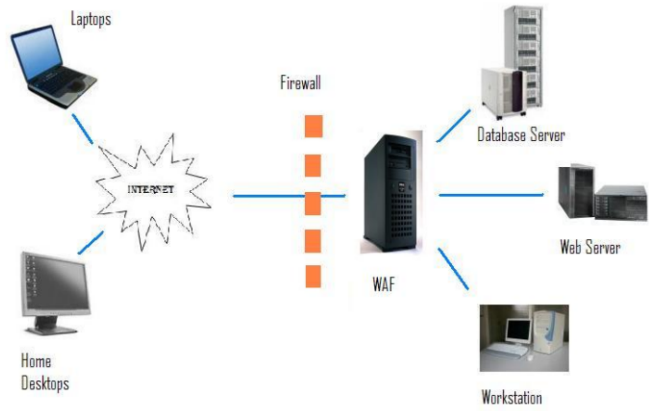

Figure 4.1: Deployment of Web Application Firewall

The deployment of a Web Application Firewall (WAF) is a critical aspect of fortifying the security of web applications and protecting against a wide range of cyber threats. When deploying a WAF, several key considerations and best practices come into play:

1. Deployment Mode:

Reverse Proxy Mode: In this mode, the WAF is positioned in front of the web application server. It acts as an intermediary between client requests and the application server, inspecting and filtering traffic before it reaches the application. This mode provides a strong defense against web-based attacks.

Inline or Bridge Mode: Here, the WAF is placed in-line with the web traffic, actively intercepting and filtering requests. This mode allows for real-time protection and monitoring of traffic but may require careful configuration to avoid disruptions.

2. Physical or Virtual Deployment:
WAFs can be deployed as physical appliances or virtual appliances, depending on the organization's infrastructure and

needs. Virtual deployments are often favored for their scalability and flexibility, while physical appliances may be chosen for dedicated, high-performance requirements.

3. Traffic Routing:
Traffic from clients is typically routed through the WAF, which inspects and filters requests for security threats. Careful consideration is given to how traffic is redirected, ensuring that all incoming requests pass through the WAF.

4. Logging and Monitoring:
WAF deployments should be configured to log and monitor web traffic and security events. This data is valuable for incident response, compliance reporting, and ongoing security analysis.

The deployment of a WAF is a pivotal step in safeguarding web applications from an ever-evolving threat landscape. Careful planning, configuration, and ongoing monitoring are essential to ensure that the WAF provides robust protection while allowing legitimate traffic to flow unhindered.

4.3 Working Process of WAF

The Working Process of WAFs is as follows:-

4.3.1 HTTP request Prolifering:

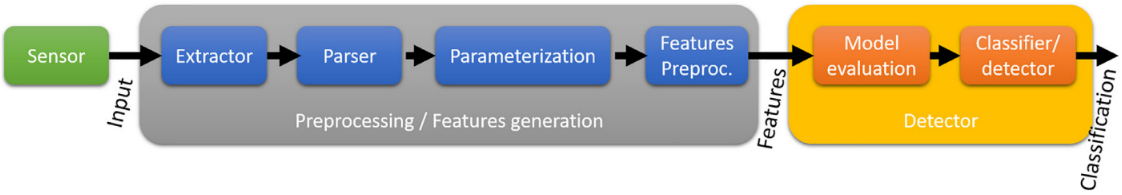

Fig. 1. Process for detecting anomalies in the URL.

Figure 4.2: HTTP Anomaly detection

The blocks in the above diagram have the following functionality:-

1. Sensor

The ”Sensor” serves as a crucial component in HTTP attack detection systems. Its primary role is to generate observations of incoming HTTP requests, which are subsequently used as input for the detection system. Sensors are responsible for monitoring and capturing data related to the web traffic directed towards a web application or network.

These observations typically encompass a wide range of attributes and information, including request headers, payloads, source IP addresses, and patterns of requests. By collecting this data, the sensor allows the detection system to analyze and assess the traffic for signs of potential attacks, anomalies, or malicious behavior.

2. Extractor

The ”Extractor” is an integral part of the HTTP attack detection process. Its primary function is to retrieve the specific data of interest from the observations generated by the sensor. This data typically includes crucial elements like Uniform Resource Identifiers (URIs), cookies, and HTTP response codes.

The Extractor acts as the component that filters and isolates these essential pieces of information from the raw HTTP request data collected by the sensor. By extracting this relevant data, the detection system gains insights into the characteristics and attributes of the incoming requests, allowing for more precise analysis and evaluation. This, in turn, facilitates the accurate identification of potential threats, anomalies, or malicious activities within the web traffic, ultimately bolstering the security of web applications and networks. The Extractor's role is pivotal in streamlining the detection process and ensuring that the detection system focuses on the critical elements necessary to assess and respond to potential security risks.

3. Parser

The ”Parser” is a critical component in the process of HTTP attack detection. It takes the output extracted by the ”Extractor” and dissects it, breaking down data like URIs into their essential elements, which can include protocol versions and headers. This structured format allows for in-depth analysis, aiding in the identification of patterns, anomalies, or potential threats within web traffic. By providing this detailed analysis, the Parser significantly enhances the security of web applications and networks, serving as a crucial bridge between raw data and the detection system's capabilities.

4. Parameterization

In the context of HTTP attack detection, ”Parameterization” serves as a vital component that focuses on the analysis of the output generated by the ”Parser.” Its primary role involves scrutinizing the properties of the parsed data. This analysis often entails assessing probabilities, identifying patterns, and examining n-grammars within the data.

Parameterization’s objective is to derive meaningful insights from the structured data provided by the Parser. By evaluating probabilities, identifying recurrent patterns, and analyzing n-gram structures, it aids in recognizing potential threats, anomalies, or malicious activities within the web traffic. This data-driven analysis empowers the detection system to make informed decisions regarding the security of web applications and networks. The Parameterization component, through its data assessment and probabilistic analysis, plays a pivotal role in enhancing the overall security posture by identifying and responding to potential security risks with precision and accuracy.

5. Features Preprocessor

The ”Preprocessor” is a critical element in the HTTP attack detection process, tasked with reducing the complexity of the data obtained from the previous step, typically the ”Parameterization.” It achieves this by applying techniques such as Principal Component Analysis (PCA) and clustering to the data.

The Preprocessor’s primary role is to streamline and simplify the data for more effective analysis and detection. Techniques like PCA help in reducing the dimensionality of the data while retaining its essential characteristics, making it more manageable for subsequent analysis. Clustering allows for the grouping of similar data points, aiding in the identification of patterns and anomalies. This preprocessing step enhances the detection system’s efficiency by preparing the data for in-depth analysis, ultimately contributing to the security of web applications and networks. The Preprocessor plays a pivotal role in ensuring that the detection process is focused on the most relevant data, facilitating accurate threat identification and response.

6. Model Evaluation

"Model Evaluation" is a pivotal phase in the context of HTTP attack detection. This component plays a crucial role in assessing incoming data by comparing it to a pre-established model that represents normal behavior. The primary objective is to obtain metrics that gauge the similarity of the input data with respect to this model.

During Model Evaluation, the system quantifies any deviations, anomalies, or patterns that do not align with the established model of normality. By analyzing these differences, the detection system can effectively identify potential threats or malicious activities within the incoming web traffic. These metrics enable the system to differentiate between legitimate and potentially harmful data, guiding security responses based on the degree of deviation from the established model. In essence, Model Evaluation is instrumental in maintaining the security of web applications and networks by allowing the system to make informed decisions about potential security risks.

7. Classifier/Detector

The "Classifier/Detector" marks the conclusive stage in HTTP attack detection, where it uses metrics and evaluations from earlier phases to make informed decisions. It plays a crucial role in categorizing incoming data as either potentially posing a threat or representing legitimate traffic. This categorization facilitates immediate security responses.

When deviations from the established model of normal behavior are detected, the Classifier/Detector classifies the data as potentially threatening, indicating that it exhibits characteristics associated with security risks or malicious activity. Conversely, if the data aligns with the expected normal behavior, it is categorized as legitimate and safe.

This pivotal component ensures that the security posture is maintained by swiftly responding to potential security threats.

It empowers the system to take appropriate actions, such as blocking suspicious requests or permitting unimpeded passage for legitimate traffic. In essence, the Classifier/Detector is the final line of defense, ensuring that web applications and networks remain secure in the face of evolving cyber threats.

4.3.2 Bot Scanning

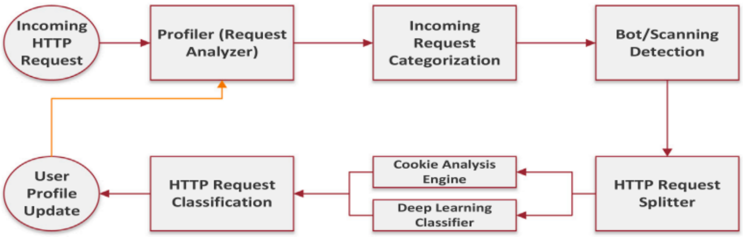

Fig. 2. Block diagram of the proposed framework.

Figure 4.3: Attack Detection Framework

In the ”Bot Scanning” phase, the initial step of attack detection, incoming requests are rigorously examined to ensure web application security. This scrutiny begins with an analysis by a bot and scanning detection engine, which evaluates various aspects of the request, including its timing, the requested URL, and the socket information of the requester.

Following this preliminary assessment, the Web Application Firewall (WAF) takes charge to perform further checks on the HTTP request. These checks consist of a set of fundamental built-in validations, meticulously defined within the system. Notably, they encompass intricate examinations such as URL encoding and Unicode encoding validations, which serve to uncover potentially malicious activity.

Should the WAF identify a bot or scanner in the incoming request, it takes immediate action by rejecting the request and profiling the user as potentially malicious. This swift response is essential to maintaining the security of web applications. Additionally, for heightened security, the request is divided into two segments. Cookies are subjected to in-depth analysis by the Cookie Analysis Engine, while other HTTP parameters are routed to a deep learning classifier for more comprehensive examination. This layered approach ensures a thorough assessment of the request and bolsters the protection of web applications and networks against potential security threats.

4.3.3 Execute Matching of rules

In the "Execute Matching of Rules" phase, the system leverages a set of predefined rules, which are expressed as Regular Expressions, to identify and mitigate specific types of threats. These rules are carefully crafted by administrators to address vulnerabilities within the system. During this phase, each rule is compared with the information contained in the packet, which represents an incoming data request.

If a rule matches the content of the packet, it triggers a security response by closing the socket connection. For example, if there is a rule that filters requests with a remote address of '203.200.95.130' and requests for the filename '/bin/sh,' it would be designed to deny such requests. This immediate and targeted action helps in preventing potential security threats from progressing further, enhancing the security of the system.

In summary, the "Execute Matching of Rules" phase employs predefined Regular Expressions-based rules to identify and block specific threats by comparing them with incoming data packets. This proactive approach is instrumental in maintaining the security of the system by swiftly responding to potential vulnerabilities.

4.3.4 Intrusion Logging

”Intrusion Logging” stands as a pivotal element in the realm of web application security. It plays a dual role, maintaining logs of communication both from the client to the server and vice versa. These logs capture a wealth of information, encompassing 18 key fields such as the communication protocol, source and destination IP addresses, port numbers, communication methods, response statuses, timestamps, and more.

The significance of these logs extends beyond mere recordkeeping. They serve as a valuable resource for identifying and addressing potential security threats. To harness this wealth of data effectively, the log files undergo a training process using Clustering Algorithms. These algorithms are designed to analyze the logged information systematically, searching for deviations or patterns that may indicate intrusion attempts or abnormal behavior. This proactive approach empowers the system to respond swiftly to any potential security breaches, contributing to a robust security infrastructure.

4.4 Clustering Algorithms

Clustering algorithms are essential in intrusion detection systems. They operate without prior labeled data, forming clusters of similar data points and identifying anomalies as data points that significantly deviate from these clusters. Several types of clustering algorithms are employed, each with unique strengths and applications. These algorithms rely on distance metrics to measure data point proximity and make assumptions that normal data instances cluster together, while anomalies are distant outliers. In intrusion detection, challenges arise due to rare anomalies and adversarial tactics to blend with normal traffic. To address these challenges, clustering algorithms are often combined with other detection methods in ensemble approaches, ensuring accurate and robust security monitoring.

Clustering algorithms fall under the umbrella of unsupervised learning, which means they don't require prior labeled data to identify anomalies. Instead, they group data points based on inherent similarities or dissimilarities. These algorithms form clusters or groups of data points based on common attributes. In the context of intrusion detection, normal activities and behaviors typically share common features, leading to the formation of clusters of normal data. Any data point that deviates significantly from these clusters may be considered an anomaly. Clustering algorithms often operate based on the assumption that normal data instances exhibit similarity and proximity in the feature space, whereas anomalies are distant or isolated. To measure this proximity, distance metrics like Euclidean distance or Mahalanobis distance are employed.

Intrusion detection presents unique challenges for clustering algorithms. Anomalies may be rare and scattered throughout the dataset, making their detection more challenging. Additionally, adversaries may employ various tactics to blend in with normal traffic, further complicating the detection process. Clustering algorithms can sometimes produce false positives (identifying normal instances as anomalies) or false negatives (failing to detect actual anomalies). Balancing these aspects is crucial to maintain the accuracy and reliability of the intrusion detection system. To enhance accuracy and robustness, clustering algorithms are often used in conjunction with other intrusion detection techniques, such as signature-based methods and anomaly-based methods, creating ensemble approaches that combine multiple detection strategies.

4.4.1 Types of Clustering Algorithms

1. Leonid Portnoy Algorithm

The Leonid Portnoy Algorithm, also known as the K-Nearest Neighbor Clustering (KNN-C) algorithm, is a clustering approach that operates based on the concept of nearest neighbors.

It assigns data points to clusters by considering their proximity to other data points. The algorithm calculates distances between data points and assigns them to clusters based on their similarity to neighboring points.

KNN-C is particularly useful in identifying clusters of data points that share common features, making it suitable for anomaly detection by highlighting data points that significantly differ from their neighbors.

This algorithm is versatile and can adapt to different types of data, making it a valuable tool in intrusion detection for identifying anomalies and outliers. **2. K-Means Algorithm**

The K-Means algorithm is one of the most popular clustering techniques used in various domains, including intrusion detection.

It operates by partitioning a dataset into 'K' distinct clusters, where 'K' is a user-defined parameter. The algorithm iteratively assigns data points to clusters and recalculates cluster centroids to minimize the variance within each cluster.

K-Means is a centroid-based algorithm, and its effectiveness depends on selecting an appropriate value for 'K' and the quality of the initial cluster centroids.

In intrusion detection, K-Means can be applied to group similar network traffic patterns and identify deviations as anomalies. It is particularly useful when the number of clusters or patterns is not known in advance.

Both the Leonid Portnoy Algorithm (KNN-C) and the K-Means Algorithm are important tools in clustering and anomaly detection. They enable the grouping of similar data points into clusters and the identification of anomalies as data points that differ significantly from the norm. These algorithms are versatile and can be adapted for various intrusion detection scenarios.

4.5 Comparison between WAFs

Table 4.1: Comparison of Web Application Firewalls (WAFs)

Feature	**ModSecurity**	**Cloudflare WAF**	**Akamai Kona Site Defender**	**Imperva Incapsula**
Deployment	On-Premises	Cloud	Cloud	Cloud
DDoS Protection	No	Yes	Yes	Yes
Bot Mitigation	Limited	Yes	Yes	Yes
OWASP Top 10 Support	Yes	Yes	Yes	Yes
Real-time Monitoring	Yes	Yes	Yes	Yes
Custom Rules	Yes	Yes	Yes	Yes
Geo-Blocking	Yes	Yes	Yes	Yes
API Protection	Limited	No	Limited	Yes
Content Delivery	No	Yes	Yes	No
SSL/TLS Offloading	No	Yes	Yes	Yes

In the realm of Web Application Firewalls (WAFs), four notable options stand out for organizations seeking robust security solutions. ModSecurity offers on-premises deployment but provides limited DDoS protection and basic bot mitigation capabilities. On the other hand, Cloudflare WAF is a cloud-based solution renowned for its robust DDoS protection, powerful bot mitigation, and comprehensive support for the OWASP Top 10 vulnerabilities. Akamai Kona Site Defender, another cloud-based WAF, also excels in DDoS protection and bot mitigation while providing extensive support for OWASP Top 10 vulnerabilities. Imperva Incapsula is a cloud-based WAF with a focus on comprehensive API protection but lacks content delivery capabilities. All four solutions offer real-time monitoring, custom rule creation, and geo-blocking features. The choice among them should be based on an organization's specific security requirements and deployment preferences.

Chapter 5

Applications

Web Application Firewalls (WAFs) are a fundamental component of modern cybersecurity strategies, designed to protect web applications from a wide array of online threats and vulnerabilities. These security solutions operate at the application layer, which is increasingly becoming the primary target of cyberattacks due to the proliferation of web-based services and the complexity of web applications. WAFs serve as a protective shield by monitoring, filtering, and blocking malicious traffic, ensuring the confidentiality, integrity, and availability of web applications. The underlying theory of WAFs revolves around threat detection, traffic analysis, and rule-based protection mechanisms.

Applications of WAFs:-

1. Traffic Analysis:
WAFs utilize deep packet inspection and content analysis techniques to scrutinize incoming web traffic. This analysis involves assessing the structure and content of HTTP requests and responses, identifying patterns, anomalies, and known attack signatures.

2. Rule-Based Detection:
A core principle of WAF theory is the establishment of security rules. These rules define patterns and behaviors that are indicative of common web attacks, such as SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF). By

comparing incoming traffic to these rules, WAFs can detect and mitigate malicious behavior.

3. Compliance and Data Protection:
WAFs are instrumental in helping organizations meet regulatory compliance requirements. They aid in protecting sensitive data, making them indispensable for industries subject to data protection laws like GDPR and HIPAA.

4. Threat Intelligence Integration:
WAFs often integrate threat intelligence feeds to stay updated on emerging threats. This application ensures proactive protection against the latest attack techniques.

5. API Security:
With the increasing use of APIs in web applications, WAFs have extended their application to protect APIs against attacks and unauthorized access.

In summary, Web Application Firewalls are grounded in theories of traffic analysis, rule-based detection, and anomaly recognition. Their applications encompass web application protection, DDoS mitigation, compliance assurance, threat intelligence integration, and API security. These versatile security solutions are critical in fortifying web applications against the evolving threat landscape and ensuring robust cybersecurity.

Chapter 6

Advantages and Disadvantages

6.1 Advantages

Web Application Firewalls (WAFs) are a critical component of cybersecurity, offering various advantages and disadvantages in safeguarding web applications.

Some Advantages of Web Application Firewalls are:-

1. Application Layer Protection:

WAFs provide security at the application layer, where modern web applications are most vulnerable. This ensures that vulnerabilities like SQL injection, cross-site scripting, and parameter tampering are effectively mitigated.

2. Real-Time Threat Detection:

WAFs continuously monitor incoming traffic in real-time, instantly detecting and mitigating web-based threats and attacks. This rapid response minimizes the potential for damage.

3. Customizable Security Rules:

WAFs allow organizations to define custom security rules to match the specific requirements of their applications. This adaptability ensures a tailored defense against evolving threats.

4. Protection Against Zero-Day Vulnerabilities:

Advanced WAFs employ machine learning and behavior analysis to detect and protect against new, previously unknown threats and zero-day vulnerabilities.

5. DDoS Mitigation:

Many WAFs include Distributed Denial of Service (DDoS) protection features, which can help maintain application availability even during volumetric attacks.

6. Security Event Logging:

WAFs generate comprehensive logs of security events, aiding in post-incident analysis, forensic investigations, and compliance reporting.

6.2 Disadvantages

Some Disadvantages of Web Application Firewalls are:-

1. Complexity:

Implementing and maintaining a WAF can be complex. Configuring and maintaining security rules can be time-consuming, requiring a thorough understanding of the application and its potential vulnerabilities.

2. False Positives and Negatives:

WAFs may generate false positives by blocking legitimate traffic or false negatives by allowing malicious traffic to pass through. Continuous fine-tuning and monitoring are essential to reduce these inaccuracies.

3. Resource Intensive:

WAFs can consume significant computational resources, which may impact the performance of web applications. Careful configuration and optimization are required to minimize resource usage.

4. Cost:

High-quality WAFs can be expensive, involving costs for hardware, software, licensing, and ongoing maintenance. Smaller organizations may find cost to be a limiting factor.

5. No Silver Bullet:

While effective against application-layer attacks, WAFs are not a universal solution. They may not protect against network or infrastructure-level attacks, and other security measures are required for comprehensive protection.

6. Complex Attacks:

Sophisticated attackers can potentially bypass or circumvent

WAFs, especially if the WAF is not properly configured, regularly updated, or if the attacks are designed to specifically evade the WAF's detection.

Understanding both the advantages and disadvantages of WAFs is crucial for organizations looking to enhance their web application security while addressing potential challenges and limitations.

Chapter 7

Conclusion

Understanding the working principles of Web Application Firewalls is fundamental. These security tools operate at the application layer of the OSI model, allowing them to inspect and filter HTTP requests and responses. This focus on the application layer makes them highly effective against application-layer attacks like SQL injection and cross-site scripting (XSS). By analyzing incoming traffic and applying predefined or custom security rules, WAFs can recognize and mitigate malicious traffic in real-time. This real-time threat detection capability provides organizations with proactive defense against web-based attacks, bolstering the overall security posture of web applications.

A thorough comparison of various WAF providers is essential for organizations to make informed decisions. This analysis helps shed light on the diversity in features, deployment options, and strengths of different solutions. For example, some WAFs are cloud-based, while others are on-premises or hybrid. They may offer additional features like DDoS protection, bot mitigation, support for OWASP Top 10, real-time monitoring, custom rule creation, and more. Assessing these features in light of an organization's specific requirements is crucial for selecting the most suitable WAF solution. Understanding the strengths and limitations of each provider ensures that organizations can make the right choice to protect their web applications effectively.

The knowledge and insights gained through the study and analysis of WAFs have significant implications for decision-making. This understanding equips organizations with the ability to select and implement WAFs that align with their security needs and objectives. It empowers them to make informed choices regarding deployment options, features, and capabilities that best suit their web applications and the evolving threat landscape. As a result, organizations can bolster the security and integrity of their web applications, protecting them against a wide range of threats and vulnerabilities.

In summary, the study and analysis of Web Application Firewalls provide organizations with a solid foundation of knowledge. This foundation encompasses the working principles of WAFs and their role in safeguarding web applications. Furthermore, the comparative analysis of different WAF providers helps organizations understand the diversity in features and strengths offered by these solutions. Armed with this knowledge, organizations can make informed decisions, selecting the most suitable WAF to enhance the security and resilience of their web applications in an ever-evolving digital landscape.

Chapter 8

Bibliography

[1] Jesus E. Verdiogo, Rafael Alonso, Antonio Alonso, German Madinabeitia. ”A critical review of the techniques used for anomaly detection of HTTP-based attacks: taxonomy, limitations and open challenges.”

[2] Waleed Shahid, Baber Aslam, Haider Abbas, Hammad Afzal, Saad Khalid. ”A deep learning assisted personalized deception system for countering web application attacks.”

[3] Waleed Shahid, Baber Aslam, Haider Abbas, Hammad Afzal, Saad Khalid. ”An enhanced deep learning based framework for web attacks detection, mitigation and attacker profiling.”

[4] Sanjay Adiwal, Balaji Rajendran, Pushparaj Shetty, Sithu Sudarsan. ”DNS Intrusion Detection (DID) — A SNORT-based solution to detect DNS Amplification and DNS Tunneling attacks.”

[5] Mariusz Sepczuk. ”Dynamic Web Application Firewall detection supported by Cyber Mimic Defense approach.”

[6] Amit Praseed, P. Santhi Thilagam. ”HTTP request pattern based signatures for early application layer DDoS detection: A firewall agnostic approach.”

[7] Seema Pillai, Dr. Anurag Sharma. ”Hybrid unsupervised web-attack detection and classification – A deep learning approach.”

[8] Reeza Mohammadi, Chaggan Lal, Mauro Conti, Lokesh Sharma. ”Software-defined network-based HTTP flooding attack defender.”

[9] Namit Gupta, Abakash Saikia, Dr. Dheeraj Sanghi. ”Study of Web Application Firewalls (WAFs).”