Web Application Firewall using Machine Learning Technique

Miss.D.R.Surwase
PG Student, Department of Computer Science & Engineering
SKN Sinhgad COE, Korti, Pandharpur
Maharashtra,India.
drsurwae1615@gmail.com

Prof.Subhash V. Pingale
Department of Computer Science & Engineering
SKN Sinhgad COE, Korti, Pandharpur
Maharashtra, India.
Subhash.pingale@sknscoe.ac.in

Abstract- In recent trends of internet technology many users are moving towards the online applications. Many websites are untrusted hence attackers are trying to find out the vulnerabilities in website if they found they are hacking the website. most of the applications are ruining on application layers by using HTTP protocol hence there is need to develop web application firewall. In this paper we have proposed the technique of web application firewall using machine learning we proposed the a web application firewall which shows higher accuracy than the existing approaches our proposed web Application firewall shows 83.5 percent accuracy

Keywords— Web Application Firewall, preprocessing, Machine Learning ,Dataset

I. INTRODUCTION

Initially many websites are developed by using simple ML, CSS Script due to this simple scripting its very simple to hack the website just by embedding simple code. if the website is not validated then its very simple to hack the website as we know that top ten attack which most famous such as cross site scripting, SQL injection attack and many more. web server infrastructure. Many online forums are available due to this it increased dramatically the use of applications. OWASP Top Ten identifies the following web application security risks: - Injection – in this attack simply execute the SQL query in the code. SQL injection attack return the output by embedding 0 or 1 Data Exposure – it exposes the important data such as financial and healthcare.in this attack most important data is exposed

External Entities (XXE) – it exposes the private internal data or it may be passed for Denial of service attack. Broken Access Control – in this attack the authentication allows attacker to access the unauthorized functionality or data. Security Misconfiguration – in his method insecure setting allows the incorrect access to unauthorized data on the system. Cross-Site Scripting – in this method attacker embed the data into web page and allow it to user to execute script arbitrarily[6].

Insecure Deserialization – in this method attacker can inject the malicious data or code into application leading remote execution Insufficient Logging & Monitoring – in this method attackers does not show their presence to network administrator

II. LITURATURE SURVEY

This literature is mainly related to the advancement in web application firewall. also covered the dataset used in web application firewall while implementation. Mainly focused on the dataset named as CSIC

Anomaly Based WAFs Torrano-Giménez et al. [1] it describe the anomaly based web application firewall he explained the method of XML file in which it describes the how to configure the web application firewall using XML file XML file contain the different rules used including HTTP request, HTTP header, verbs web application present an anomaly-based WAF.

Moosa et al. [2] states that the most of te web applications are vulnerable. Also he explained about the how it reduces the effectiveness of signature based firewall such as Mod security.

Mod security is open source web application firewall but it has some limitation

Moosa proposes advance firewall which is user friendly and easy to update

Pałka et al. [3] describes the use of machine learning web application firewall to secure web application from different attack such as SQL Injection it exploit the use of unusual refereed browser history browsing history. They also explore issues resulting from WAF data storage and how this could represent a vulnerability in its own right if sensitive data is stored or processed within the WAF algorithms.

Epp, N. et al. [4] explore the comparison of triggered learning and continuous learning scheme triggered learning has lower learning retention rate however it is less adaptable but requires a less initial setup. However, continuous learning may trigger false positives with changing usage patterns and could be vulnerable to attacks that target the learning process itself. They implemented a machine-learning-based WAF in Apache based on parameter length and character distribution, parameter class (i.e., numerical, URL, email etc.) and enumerated type.

Nguyen et al. [5] focuses on important feature selection to their effectiveness he considered 30 feature which more related to web attack

III Proposed Web Application Firewall using machine learning model

Now let's see how intrusion detection systems can be developed. Few of the important steps are:

1. Collected the dataset of CSIS 2010
2. Pre-processed this dataset by using feature engineering
3. Applied a machine learning algorithm
4. Train our model with the dataset.
5. Test and evaluate the model.

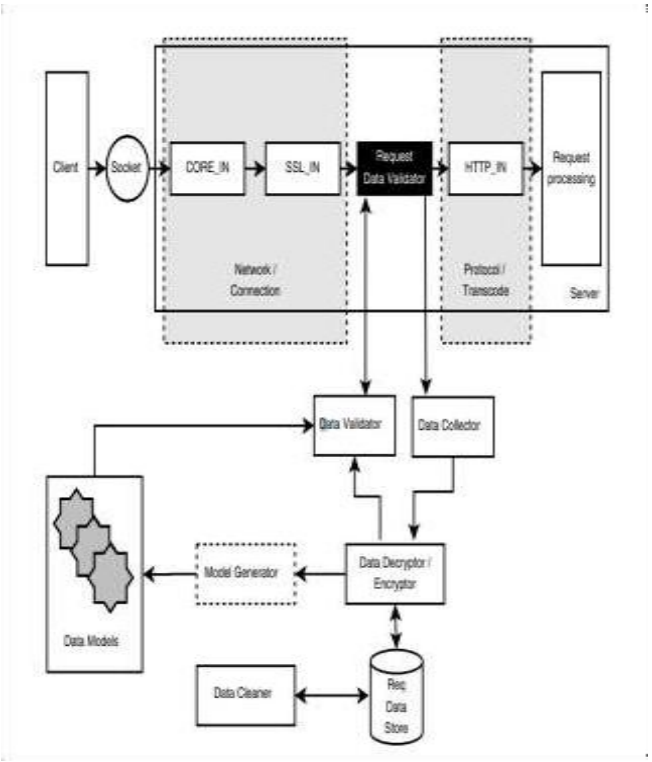

Fig 1. Block Diagram of Proposed WAF

In the machine learning model first we need to do preprocessing on the data. processing means removal of unwanted features once preprocessing completed then we can apply the machine learning algorithm in this paper we applied the machine learning algorithm we focused on the dataset named as CSIC 2010 we obtained better results as compared to other approaches

IV Datasets:

In our implementation we have used the standard dataset named as CSIC 2010 We analyzed the different features from this dataset and passed to the proposed firewall we got the better results as compared to existing approaches that is support vector machine and 48balgorithm

come up with such structures. For limited domains, it should plot appropriate structures, even though output depends on open-domain semantic analysis.

V- Result and Discussion

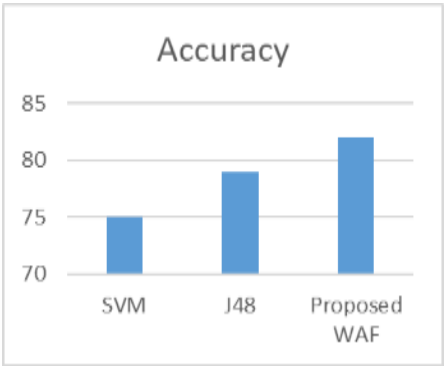

Figure – Comparison of Accuracy

References

[1] Torrano-Giménez, C., Nguyen, H.T., Alvarez, G. and Franke, K., Combining expert knowledge with automatic feature extraction for reliable web attack detection. Security and Communication Networks, 8(16), pp.2750-2767. 2019

[2] Moosa, A., Artificial neural network-based Web Application Firewall for SQL injection. International Journal of Computer and Information Engineering, 4(4), pp.610-619. 2017

[3] Pałka, Learning Web Application Firewall-benefits and caveats. In International Conference on Availability, Reliability, and Security (pp. 295-308). Springer, Berlin, Heidelberg. 2015

[4] Epp, N., Funk, R., Cappo, C. and Lorenzo–Paraguay, S., (2017) Anomaly-based Web Application Firewall using HTTP-specific features and one-class SVM. In Workshop Regional de Segurança da Informação e de Sistemas Computacionais 2018

. [5] Nguyen, H.T., Torrano-Giménez, C., Alvarez, G., Petrović, S. and Franke, K., Application of the generic feature selection measure in detection of web attacks. In Computational Intelligence in Security for Information Systems (pp. 25-32). 2015

[6] Betarte, G., Giménez, E., Martinez, R. and Pardo, A., Improving Web Application Firewalls through anomaly detection. 17th IEEE International Conference 2018