International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 874-876 www.ijaem.net ISSN: 2395-5252

Intrusion Detection System for Web-Based Attacks Using Machine Learning

Mahesh Basavaraj[1], Vanitha S[2], Sreeja K[3], Sharad Mahadev[4]

[1] *Asst. Professor, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.*

[234] *Students, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.*

Date of Submission: 10-04-2023 Date of Acceptance: 20-04-2023

ABSTRACT-Web-based apps are frequently used in modern times. These programs have become somewhat of a difficulty for the web servers due to their enormous variety and the fact that they were developed by programmers with no experience in security, intrusion tactics, or avoiding them. Many research has been conducted to investigate the viability of using machine learning approaches to detect web-based threats. False-positive and false-negative results have been identified as a significant problem that needs to be solved in order to make machine learning-based web attack detection and prevention effective and trustworthy. We tried to pinpoint and fix the primary reason behind the false-positive and false-negative outcomes.

Key Words: Detection of Intrusions, Using Python for machine learning, Algorithms, Network Security, False Positive, false Negative, OneR.

I. INTRODUCTION

Many corporations view web servers as one of their most important and sensitive components. On the other hand, attacks against these web servers have increased in frequency recently due to their significance. However, the application layers are typically created by persons who are not very knowledgeable about web-layer assaults and security, which will lead to redundancy in the app's susceptibility.

On the other hand, because these applications are often accessible on port 80, attacks cannot be stopped by using the current firewalls because the port must be open to everybody if we want to enable access to the relevant websites. Several well-known assaults are found because some intrusion detection systems based on the web are set up in accordance with the signature.

Website and web application vulnerabilities are constantly expanding as a result of widespread use. According to a 2019 poll [1], 68% of web apps could experience sensitive data breaches, and 9 out of 10 web applications are vulnerable. Furthermore, in 8% of instances, illegal access to web servers led to that network breach. The Internet is a hacker's target because of how widely and incalculably it is used. Cross-Site Scripting (XSS), SQL Injection, and other types of web vulnerability detection are the key objectives.

1.1 Objectives

The main goal of the study is to create a framework for machine learning-based automatic detection of abnormalities prevalent in web-based attacks. Do feature selection and choose the right features for a prediction model of web-based attacks as the secondary goal. SVM, decision trees, and random forests, among other low-cost machine learning methods, are used to construct the suggested detection model.

1.2 Problem Statement

Online assaults like SQL injection (SQLi) and cross-site scripting (XSS) have been identified as serious risks to websites, web applications, and web users. These web assaults have the potential to seriously harm websites, web users, and web applications by circumventing authentication mechanisms, stealing private data from databases and users, and even seizing total server system control. To cope with web threats, numerous strategies have been researched and deployed to secure web applications websites, and web users. A viable method for adding defensive layers to protect websites and web applications is the detection of web attacks.

II. RELATED WORKS

The project is divided into four modules: attack prediction and testing, feature selection, model training, and data pre-processing. NSL-KDD is the chosen dataset [2]. One of the most popular datasets

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 874-876 www.ijaem.net ISSN: 2395-5252

for training intrusion detection systems (IDS) and intrusion prevention systems is the KDD Cup 99 dataset (IPS). For network security, there are not enough tagged datasets available.

The selection of features can be done in a variety of ways.

1) Basic Feature Selection
2) Feature elimination through recursion
3) The Principal Component Analysis (PCA)

The models mentioned above are applied, with the dataset being divided into 85% training and 15% testing. The performance metrics used will be

1) Accuracy
2) Precision
3) Recall
4) f1-score

III. IMPLEMENTATION AND WORKING

3.1 Web Application Architecture

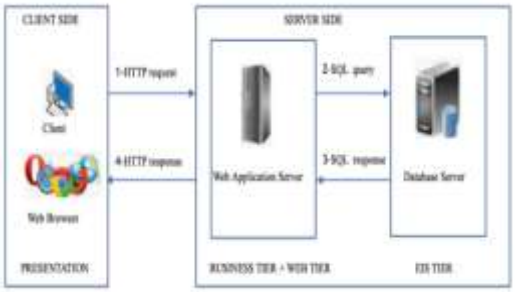


Fig -1: Web Application Architecture

Web-based apps are a required network-based delivery technique for standard online services. These programs are created using client- and server-side development. The server side uses backend scripting languages like.NET, PHP, and JEE and consists of a web server, a web application, and a database server (Jakarta Enterprise Edition). Front-end scripting languages, such as CSS/HTML, JavaScript, etc., are used on the client side to operate on the user's web browser. The HTTP protocol is typically used to connect these two.The server-side and client-side web application architecture is shown in Fig. 1.

3.2 System Framework

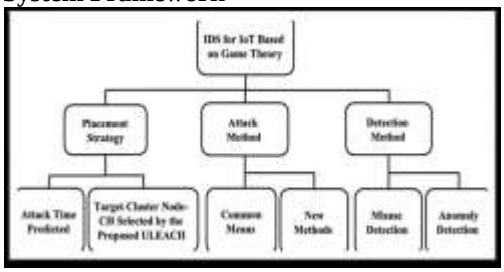


Fig -2: System Framework

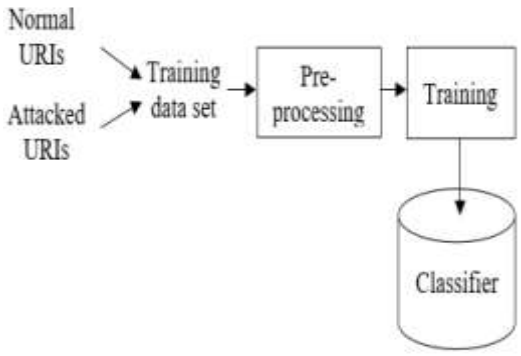
The Intrusion Detection System has recently integrated machine learning techniques, particularly supervised learning approaches (IDS). Most modern IDSs suffer from excessive computational cost and poor performance against unknown assaults due to the limitations of supervised learning in the Internet of Things (IoT). To address these issues, we suggest UTEN-IDS, a novel IDS built on unsupervised methods. The network data is handled by UTEN-IDS using an ensemble of autoencoders, and an isolation forest technique is used to detect anomalies. Two benchmark datasets are used to confirm the effectiveness of the suggested strategy. As compared to other methodologies, the results demonstrate that our methodology has considerable advantages in classification performance and demonstrates its applicability in the IoT network.

3.4 Proposed System

Phase 1 of the experimentation, which was conducted in three stages, concentrated on dataset preparation and pre-processing to identify sub-categories of anomalous traffic for a refined training set that helped us find the missing features for frequent attacks. Before being processed for the following stage, the data set was corrected using Python programming language for some errors, Latin characters, blank spaces, and rows at the end of the lines.

Phase 2 was devoted to extracting features from the dataset, which contains fields like GET and POST requests, content-length, the number of special characters, login name, and password, as well as the number of keywords for SELECT, DROP, UNION, DELETE, and MODIFY. Before feeding the cleaned dataset to Weka 3.8, an open-source machine learning and data mining application, it was placed file.

Fig -3: Diagram of proposed system



International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 874-876 www.ijaem.net ISSN: 2395-5252

3.5 Hardware requirements

Table -1: List of Requirements

Processor	Intel i3
Ram	4GB or above
System Type	64/32 -bit OS
Hard disk	40GB 0r more

3.6 Software requirements

1. Operating System: Windows 7,8/10/11, Linux Ubuntu
2. Programming language: Python 3.10
3. Libraries: matplotlib, NumPy, pandas, sklearn
4. Interface/IDE: Jupyter Notebook / VS code

IV. FUTURE SCOPE

A detective tool called an intrusion detection system (IDS) is used to find harmful (including policy-violating) activities. A preventive tool, an intrusion prevention system (IPS) is primarily made to both identify and prevent harmful activity. Provide centralised administration for the attack correlation. act as another layer of defence for the business. It examines various attacks, recognises their patterns, and supports in organising and putting in place efficient control.

V. CONCLUSIONS

In order to discriminate between regular and abnormal traffic, we filtered and labelled the CSIC HTTP 2010 dataset before conducting our experiments. A python script was used to perform fine-grained pre-processing on the data to find any features that a typical attack would be missing. Additionally, by using various Machine Learning classifiers like J48, Bayesian Network, OneR, and Decision tables that use evaluation metrics to find the accuracy using Weka 3.8, feature extractionfrom thedataset proved to be important in identifying malicious activity and the attack types like SQL injection (SQLi), Cross-Site Scripting (XSS), and Buffer Overflow. Additionally, by using fine-tuned feature set engineering, 20 features were retrieved with better web-based attack detection, raising the true positive rate. Finally, the J48 decision tree algorithm was shown to be the top performing algorithm with the best attack detection rate of 94.5% in our testing results using three machine learning algorithms (J48, Naive Bayes, and OneR).

REFERENCES

[1]. C. Turano-Gimenez, A. Pérez-Villegas and G. Alvarez, “An Anomaly-Based Approach for Intrusion Detection in Web Traffic,” The Allen Institute for Artificial Intelligence, 2009.

[2]. G. Betarte, E. Giménez, R. Martínez, and A. Pardo, “Machine learning-assisted virtual patching of web applications,” [Online]https://arxiv.org/abs/1803.05529, Mar 2018.

[3]. J. Liang, W. Zhao, W. Ye, “Anomaly-Based Web Attack Detection: A Deep Learning Approach,” ICNCC 2017, pp. 80-85, December 8–10, 2017, Kunming, China.

[4]. Y. Pan, F. Sun, Z. Teng, J. White, D.C. Schmidt, J. Staples and L. Krause, “Detecting web attacks with end-to-end deep learning,” Journal of Internet Services and Applications, vol. 10:16, SpringerOpen, 2019.

[5]. Black-box detection of XQuery injection and parameter tampering vulnerabilities in web applications. Deepa, G.; Thilagam, P.S.; Khan, F.A.; Praseed, A.; Pais, A.R.; Palsetia, N. Int. J. Inf. Secur. 2018, 17, 105–120. [CrossRef]

[6]. Yan, J.; Liu, T.; Qi, Y.; Shi, L. Locate-then-DetecT: Using attention-based deep neural networks, real-time web assault detection. Pages 4725–4731 are found in the proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19), which was held in Macao, China, from 10–16 August 2019.

[7]. Jia, L.; Melicher, W.; Fung, C.; Bauer, L. A Lightweight Hybrid Method for Machine Learning-Based DOM XSS Vulnerability Detection. Pages. 2684–2695 in Web Conference 2021 Proceedings, Ljubljana, Slovenia, 19–23 April 2021. [CrossRef]