iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53133

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

DDoS Attack Detection using ML

Pradnya Kasture[1], Tejas Chougule [2], Dhananjay Patil[3], Prithviraj Mahapure[4]
Department of Computer Engineering, First-Fifth SPPU

Abstract: ***DDoS attacks are an attempt to prevent the service from being unavailable by overloading the server with malicious traffic. In the past few years, distributed denial of service attacks is becoming the most difficult and burdensome problem. The number and magnitude of attacks have increased from few megabytes of data to 100s of terabytes of data these days. As there are different attack patterns or new types of attacks, it is difficult to detect such attacks effectively. New techniques for generating and mitigating distributed denial of service attacks have been developed in the present paper, which demonstrate that they are far superior to those currently used. In addition, in order to carry out a thorough investigation of the challenges presented by distributed denial of service attacks, we classify DDoS attack methods and techniques used for their detection. We're comparing the attack module to a few other tools out there.***

Keywords: DDOS, SVM, Random Forest

I. INTRODUCTION

The SVM algorithm is utilized for DoS attack detection by extracting flow statistics associated with such attacks. This method demonstrates advantages in terms of low resource consumption and a high detection rate. The crucial aspect lies in extracting the time interval information. However, a drawback of this approach is the presence of detection hysteresis, leading to potential delays and less accurate identification of attack behaviours. The authors have proposed a framework designed for detecting and mitigating DoS attacks in large-scale networks, which may not be suitable for smaller deployments.

In another study, a mechanism for DoS attack detection is introduced, which relies on a database containing legitimate source and destination IP addresses. By employing the non-parametric cumulative algorithm CUSUM, this method analyzes the abnormal characteristics exhibited by source and destination IP addresses during a DoS attack, effectively identifying such attacks. However, the approach requires adjusting and determining the appropriate threshold for optimal performance.

Regarding DoS attack detection in SDN networks, it is observed that information entropy and the utilization of data mining algorithms, particularly the SVM algorithm, play a significant role. Nevertheless, the information entropy approach suffers from a high false positive rate, while the SVM algorithm necessitates determining the number of neurons in advance. Consequently, this paper summarizes the characteristics of several DoS attacks, collects information from the switch flow table, extracts a matrix of characteristic values based on a six-tuple representation, and establishes an SVM classification model for detection purposes.

II. RELATED WORK

There has been significant research and work done on detecting and mitigating Distributed Denial-of-Service (DDoS) attacks. DDoS attacks are more sophisticated and challenging to handle compared to traditional DoS attacks because they involve multiple attacking sources distributed across different networks. Here are some related works and approaches in the field of DDoS attack detection:

1) *Traffic Analysis:* Many studies focus on analyzing network traffic patterns to detect DDoS attacks. These approaches involve monitoring and analyzing various network parameters such as packet rates, traffic volume, and flow characteristics. Deviations from normal traffic behavior can indicate the presence of a DDoS attack.
2) *Flow-based Detection:* DDoS attacks can be identified by analysing flows or connections between network entities. Flow-based detection methods examine flow attributes like source/destination IP addresses, port numbers, and protocol types to identify suspicious traffic patterns that indicate a DDoS attack.
3) *Collaborative Detection:* Collaboration among multiple network entities is employed to detect and mitigate DDoS attacks. This approach involves sharing information and coordinating actions among different network domains or organizations to collectively identify and respond to DDoS attacks. Collaborative detection enhances the ability to detect attacks distributed across multiple networks.
4) *Behaviour-based Detection:* Behaviour-based detection techniques aim to identify abnormal behavior in network traffic. These methods establish baseline traffic patterns and then monitor for deviations that might indicate a DDoS attack. Anomalous behaviours could include sudden traffic spikes, unusual communication patterns, or abnormal resource consumption.

5) *Traffic Filtering and Rate Limiting:* Network-based solutions involve filtering or rate limiting traffic at various network levels. These techniques employ mechanisms like Access Control Lists (ACLs), firewalls, and Intrusion Prevention Systems (IPS) to identify and filter out malicious traffic associated with DDoS attacks.
6) *SDN-based Detection and Mitigation:* Software-Defined Networking (SDN) has been leveraged to enhance DDoS attack detection and mitigation. SDN allows centralized network control, enabling more efficient monitoring and response to attacks. Dynamic flow management and traffic engineering techniques in SDN architectures have been explored to detect and mitigate DDoS attacks effectively.
7) *Anomaly Detection:* Anomaly-based detection techniques aim to identify DDoS attacks by detecting deviations from normal network behavior. These methods often involve statistical analysis, machine learning algorithms, or heuristics to establish a baseline of normal behavior and flag any anomalous patterns or activities as potential DDoS attacks.
8) *Traffic Classification:* DDoS attacks can be detected by classifying network traffic into different categories based on their characteristics. This classification can help distinguish legitimate traffic from malicious traffic associated with DDoS attacks. Various methods such as port-based classification, payload analysis, and behavior-based classification have been explored for this purpose.
9) *Collaborative Defense Systems:* Collaborative defense systems involve the cooperation of multiple network entities, such as ISPs, organizations, and security service providers, to collectively detect and mitigate DDoS attacks. These systems facilitate information sharing, early attack detection, and coordinated response actions, improving the overall effectiveness of DDoS mitigation.
10) *Reputation-Based Approaches:* Reputation-based techniques utilize reputation or trust scores assigned to network entities (e.g., IP addresses, domains) to identify potential sources of DDoS attacks. By maintaining reputation databases and analyzing the historical behavior of entities, these methods can block or limit traffic from entities with poor reputations.
11) *Hybrid Approaches:* Some approaches combine multiple detection techniques to enhance the accuracy and efficiency of DDoS attack detection. For example, combining flow-based analysis with machine learning algorithms or integrating anomaly detection with traffic classification can provide more comprehensive and effective detection capabilities.

III. PROPOSED WORK

We succeeded in building a machine learning models like Support Vector machine (SVM) and random forest (RF). It uses the dataset for training and the testing the models which contains the packet headers (source/destination IP addresses, source/destination ports, protocol types), traffic rates (packet rates, byte rates), flow characteristics (flow duration, number of packets in a flow), and statistical metrics (mean, standard deviation, entropy, etc.) calculated over specific time windows or flows. It also includes the labels of each request. Using this dataset for training our model can detect if the attack is performed or not.

A. *Here is How our SVM Model Works*

1) *Data Collection:* The first step is to collect labelled network traffic data, including both normal and DDoS attack instances. This data serves as the training set for the SVM model.
2) *Feature Extraction:* Relevant features need to be extracted from the network traffic data. These features can include packet headers, traffic rates, flow characteristics, protocol information, and statistical metrics.
3) *Feature Selection:* It's crucial to select the most informative features that contribute significantly to the discrimination between normal and attack traffic. Feature selection techniques, such as correlation analysis or information gain, can help identify the most relevant features.
4) *Data Preprocessing:* The collected data may require preprocessing steps like normalization, scaling, or handling missing values to ensure optimal performance of the SVM model.
5) *Model Training:* The preprocessed data is then used to train the SVM model. The SVM algorithm aims to find an optimal hyperplane that separates the normal and attack instances in the feature space.
6) *Model Evaluation:* The trained SVM model is evaluated using test data that was not used during the training phase. Various evaluation metrics, such as accuracy, precision, recall, and F1-score, can be used to assess the performance of the model.
7) *Threshold Determination:* Depending on the SVM model output (e.g., distance to the hyperplane or decision scores), a threshold can be set to classify instances as either normal or DDoS attack. This threshold can be adjusted based on the desired trade-off between detection rate and false positive rate

8) *Real-Time Detection:* Once the SVM model is trained and the threshold is determined, it can be deployed for real-time DDoS attack detection. Network traffic is continuously monitored, and incoming instances are classified by the SVM model. Instances exceeding the threshold are flagged as potential DDoS attacks.

B. And here is how our Random Forest algorithms Works

1) *Data Collection:* Similar to SVM, the first step is to collect labeled network traffic data, including both normal and DDoS attack instances. This data will be used to train the Random Forest model.
2) *Feature Extraction:* Relevant features need to be extracted from the network traffic data. These features can include packet headers, traffic rates, flow characteristics, protocol information, and statistical metrics. It is important to select informative features that can effectively distinguish between normal and attack traffic.
3) *Data Preprocessing:* The collected data may require preprocessing steps such as normalization, scaling, handling missing values, or encoding categorical variables to ensure optimal performance of the Random Forest model.
4) *Train-Test Split:* The labeled data is split into a training set and a separate test set. The training set is used to train the Random Forest model, while the test set is used to evaluate its performance.
5) *Model Training:* The Random Forest model is trained using the training set. Random Forest is an ensemble learning algorithm that constructs multiple decision trees. Each decision tree is built on a random subset of the training data and a random subset of features. The trees are trained independently and their predictions are aggregated to make the final decision.
6) *Model Evaluation:* The trained Random Forest model is evaluated using the test set. Various evaluation metrics such as accuracy, precision, recall, and F1-score can be used to assess the performance of the model in detecting DDoS attacks.
7) *Real-Time Detection:* Once the Random Forest model is trained and evaluated, it can be deployed for real-time DDoS attack detection. Network traffic is continuously monitored, and incoming instances are classified by the Random Forest model. The aggregated predictions from the ensemble of decision trees determine if an instance is classified as normal or indicative of a DDoS attack.

IV. SYSTEM ARCHITECTURE

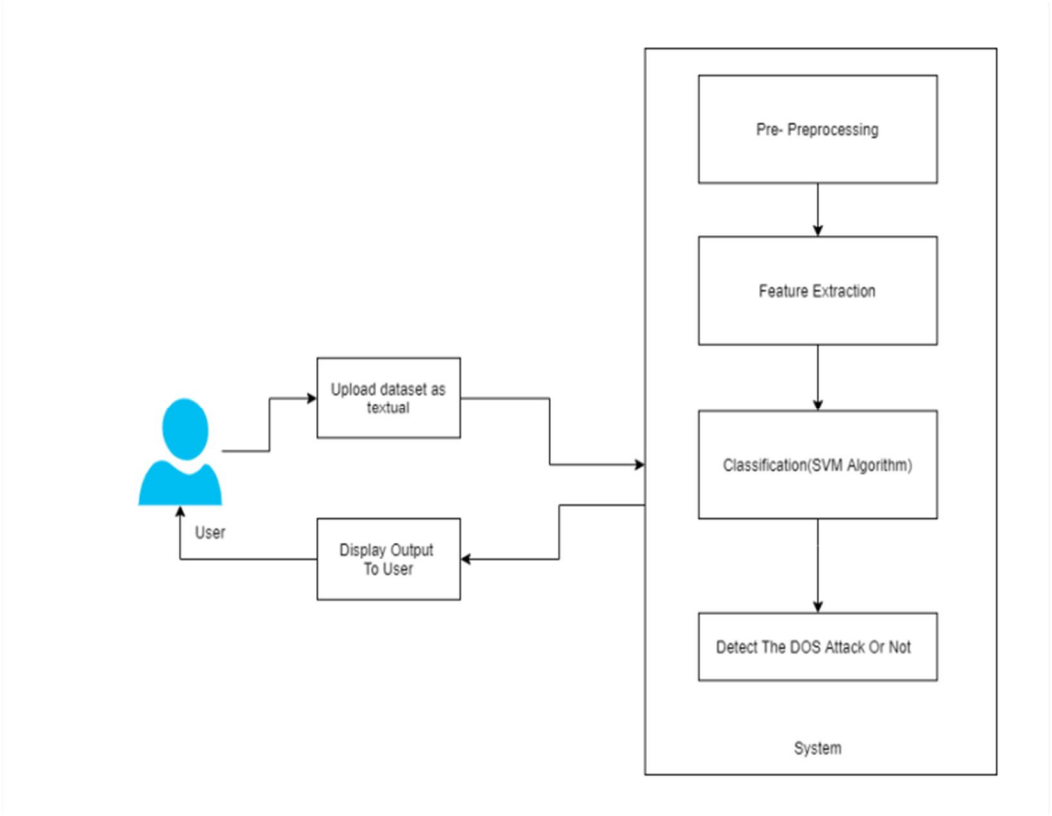


Fig. System Architecture

V. CONCLUSION

In conclusion, the project of DDoS attack detection using SVM and Random Forest demonstrates the effectiveness of machine learning algorithms in identifying and mitigating DDoS attacks. Both SVM and Random Forest are powerful classifiers that can analyse network traffic data and distinguish between normal traffic and DDoS attack patterns. The SVM algorithm utilizes support vectors to find an optimal hyperplane that separates the two classes, while Random Forest constructs an ensemble of decision trees to make aggregated predictions. By collecting a labelled dataset containing network traffic instances representing normal and attack traffic, relevant features can be extracted and used to train the SVM and Random Forest models. The models are then evaluated using test data to assess their performance in detecting DDoS attacks. Both SVM and Random Forest have their strengths and weaknesses. SVM is known for its ability to handle high-dimensional data and find optimal decision boundaries, while Random Forest excels at capturing complex relationships and is robust against overfitting. The project highlights the importance of feature selection, data preprocessing, and model evaluation to ensure accurate and reliable detection results. Additionally, ongoing monitoring and periodic updates are necessary to adapt to evolving DDoS attack techniques. Overall, the project provides insights into the application of machine learning algorithms in DDoS attack detection, showcasing the potential of SVM and Random Forest as effective tools in mitigating and protecting against DDoS attacks in network environments

REFERENCES

[1] T. Subbulakshmi, K. BalaKrishnan, S. M. Shalinie, D. AnandKumar, V. GanapathiSubramanian and K. Kannathal, "Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset," *2011 Third International Conference on Advanced Computing*, Chennai, India, 2011, pp. 17-22, doi: 10.1109/ICoAC.2011.6165212.

[2] A. E. Krasnov, D. N. Nikol'skii, D. S. Repin, V. S. Galyaev and E. A. Zykova, "Detecting DDoS Attacks Using the Analysis of Network Traffic as Dynamical System," *2018 International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC)*, Moscow, Russia, 2018, pp. 1-7, doi: 10.1109/MoNeTeC.2018.8572034.

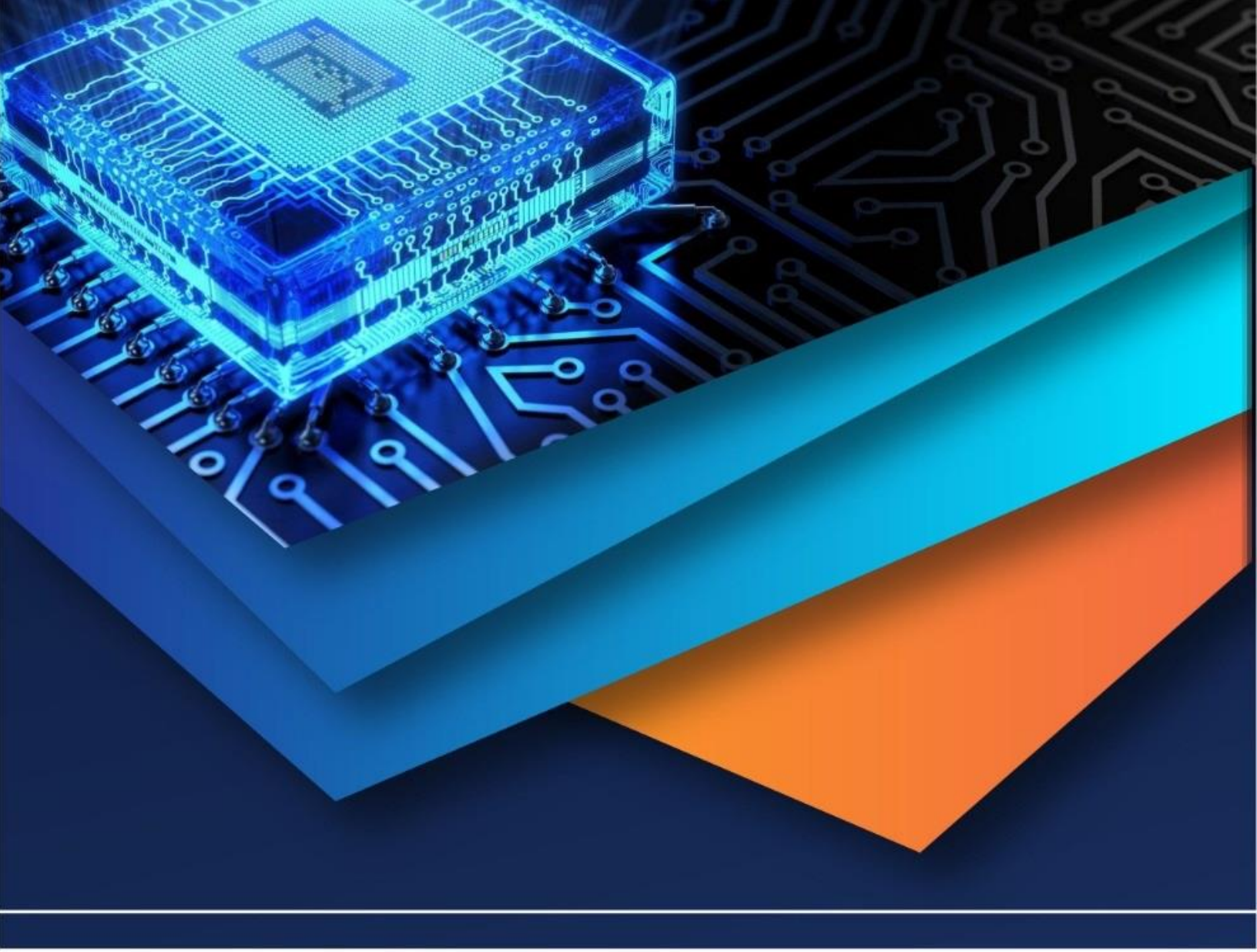
[3] M. A. T. Laksono, Y. Purwanto and A. Novianty, "DDoS detection using CURE clustering algorithm with outlier removal clustering for handling outliers," *2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC)*, Bandung, Indonesia, 2015, pp. 12-18, doi: 10.1109/ICCEREC.2015.7337029.

[4] T. -C. Leung and C. -N. Lee, "Flow-Based DDoS Detection Using Deep Neural Network with Radial Basis Function Neural Network," *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Chiang Mai, Thailand, 2022, pp. 1774-1779, doi: 10.23919/APSIPAASC55919.2022.9980171.

[5] M. S. E. Sayed, N. -A. Le-Khac, M. A. Azer and A. D. Jurcut, "A Flow-Based Anomaly Detection Approach With Feature Selection Method Against DDoS Attacks in SDNs," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 4, pp. 1862-1880, Dec. 2022, doi: 10.1109/TCCN.2022.3186331

[6] Y. Chen, K. Hwang and W. -S. Ku, "Collaborative Detection of DDoS Attacks over Multiple Network Domains," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 12, pp. 1649-1662, Dec. 2007, doi: 10.1109/TPDS.2007.1111.

[7] R. Yogesh Patil and L. Ragha, "A rate limiting mechanism for defending against flooding based distributed denial of service attack," *2011 World Congress on Information and Communication Technologies*, Mumbai, India, 2011, pp. 182-186, doi: 10.1109/WICT.2011.6141240.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY