International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Review Paper Volume-9, Issue-12, 2022 Regular Edition E-ISSN: 2349-7084

Machine Learning-Based DDoS Saturation Attack Detection and analysis in SDN Environment

[1]P Sandeep Kumar Reddy, [2]M Sri Raghavendra,[3] K Sreenivasulu,[4]T N Balakrishna

[1] *M.Tech Student , G.Pullaiah College of Engineering and Technology , Kurnool*
[2]*Assistant Professor of CSE, G.Pullaiah College of Engineering and Technology, Kurnool*
[3]*Professor of CSE,G.Pullaiah College of Engineering and Technology , Kurnool*
[4]*Assistant Professor of CSE , G.Pullaiah College of Engineering and Technology , Kurnool*

Email: sandeeppalannagari@gmail.com, sr.meeniga@gmail.com , sreenu.kutala@gmail.com, balakrishna.tn@gmail.com

Corresponding author : ***P Sandeep Kumar Reddy***

Available online at: http://www.ijcert.org

Received: 21/10/2022, Revised: 14/11/2022, Accepted: 26/12/2022, Published: 8/01/2023

Abstract: To create dynamic, adaptable, manageable, and cost-effective computer networks, Software Defined Network (SDN) has been developed as a new methodology. As a result, the security of SDN is essential. Switches in SDN can match incoming packets to flow tables but not process anything. To identify SDN, DDoS, and saturation attacks, different Machine Learning-based detection methods have recently been presented. This method detects and analyses DDoS saturation attacks using Machine Learning in an SDN environment. The presented model utilizes a variety of Machine Learning (ML) methods, including AdaBoost, Decision Tree (DT), and Support Vector Machine (SVM). Experimental results clearly express that the described Machine Learning model provides more Accuracy, Precision, Recall and F1-Score compared to simple Machine Learning models. The combined Machine Learning (SVM+ DT+ AdaBoost) accuracy is 97.6%, precision, recall, F1-score values are 96.6%, 97.4%, 98% respectively.

Keywords: Software Defined Network (SDN), Distributed Denial of Service Attacks (DDoS), Machine Learning, SVM, DT and AdaBoost.

---

1. Introduction

By dividing the control plane from the data plane, Software-Defined Networking (SDN) enables a more straightforward approach to network management and operation.It facilitates the creation and deployment of network applications by offering programmable interfaces. Alternatively, SDN is susceptible to DoS saturation attacks [1]. By creating recognition systems based on Machine Learning, this issue has been solved in a number of studies. SwitchGuard uses a nonlinear auto-encoder classifier to recognize and prevent anticipated and well-known saturation attacks on OpenFlow switches.SDN architecture makes the controller directly programmable and divides network control from forwarding devices [2]. It enables network managers to dynamically modify the flow of network traffic. Because of the network's distributed architecture, the controller can observe the entire system. The SDN environment offers excellent reliability, simplicity, and adaptability with the aid of these strong qualities [3].

Data plane, control plane and application plane of SDN consists of three planes.Network traffic is carried by the data plane based on a controller determination. The control plane evaluates routing tables to determine traffic flow. The application plane controls other applications like load balancers, firewalls, and Quality of Service (QoS) apps. [4]. SDN architecture increases network performance by separating network control and forwarding tasks. Several routers in the network are managed by control programmes operating on a normal control system. An SDN controller is at risk from Distributed Denial-of-Service (DDoS) attacks. DDoS attacks can be difficult to detect because attack packets may be mistaken for legitimate traffic. A DDoS attack targets a single system with the many compromised zombies or bots [5] or bots. The best categorization approach for identifying DoS and DDoS attacks on the controller has been carefully investigated.

There are two forms of DDoS detection: Signature-based Detection and Anomaly-based Detection.

• Signature–based Detection:Rule-based or filter-based methods are used in this approach to detect previously known attacks, since they are usually stored in the database unless a wave is detected and its limit is with unknown new attacks [6].

• Anomaly–based Detection:In contrast to the first method, this detection system is focused on behaviour and primarily reports on routine usage. Otherwise, it declares an attack and possibly detects the next attack; However, it takes a lot of time to train normal behaviour.

Network traffic classification based on Machine Learning technology has gained popularity and shown excellent performance in the detection of intrusions [7]. With the use of Machine Learning (ML), it is possible to research and learn while evaluating the given data to draw conclusions about things like prediction, diagnosis, remote control, and recognition. SDN with an integrated Machine Learning application could also be used as a reference model in research on the integration of 5G networks.

UPD flood:

User Datagram Protocol (UDP) packets were repeatedly sent to a predetermined or random port in an attempt to attack the victim machine.

ICMP flood attack:

The victim receives a significant amount of Internet Control Message Protocol (ICMP) echo request (ping flood) packets with fake source IP addresses.

TCP SYN attack: This attack makes use of a weakness in the Transmission Control Protocol (TCP). An attacker tries to attack the server with many SYN (sync) requests. The server waits for the client to transmit an ACK packet before responding to the request with a SYN + ACK (Acknowledge) packet. The server is currently waiting for a hypothetical ACK because the attacker has not yet submitted an ACK packet. A full limited buffer queue on the server causes it to refuse valid inbound requests.

Smurf attack: These amplifiers can redirect ICMP packets with a spoofed source IP address to routers and servers targeted for amplification attacks. The victim host IP address will be the spoof address. The sources of UDP and ICMP flood attacks are simple to trace, but the sources of Smurf attacks are more difficult.

This approach is organized in the following section: literature survey is introduced in Section II, Machine Learning-based DDoS saturation attack detection and analysis on SDN is described in Section III. Section IV includes the analysis of experimental results and finally the approach concludes with Section V.

2. Literature Survey

Alshamrani, A. Chowdhary, A. Pisharody, S. Lu, D. Huanget al. [8] Performance of the current DDoS attack protection has been evaluated. As a result, the impact of harmful and unique flow attacks on SDN has been studied. To respond with rapid traffic changes occurring in the SDN architecture during an attack, Machine Learning classification algorithms was utilized to monitor traffic data periodically gathered from transmission devices on the data plane. As a result, during the attack, Packet_In messages that flow between the controller and the transmitting device are used. For classification, the Naive Bayes (NB), J48, and Support Vector Machine (SVM) algorithms are utilized.

Z. Z. Abaid, M. A. Kaafar, and S. Jha et. al. [9] The potential for developing escape attacks against Android malware classifiers with varying levels of adversary capability and awareness is being looked at. According to the displayed statistics, when hazardous mobile applications are simply concealed without further attacker information, the detection rates of the Random Forest and ANN (Artificial Neural Networks) classifiers can be decreased by 100%.

H. Lin and P. Wanget al. [10] suggestedaSDN-based DDoS attack detection technology has been developed. Attacks known as Denial of Service (DoS) bombard networks with a massive volume of machine-generated packets. Three OpenFlow management tools were combined with Flow to detect unexpected network activity. The recommended method's implementation and operation are therefore complicated.Barki, Lohit, et al. [11] Different machine learning techniques like Naive Bayes, K-Nearest Neighbor, K-Means, K-medoids have been used to detect DDoS attack. It is found that the Naive Base model outperforms other algorithms regarded as having the highest accuracy.

Nanda. S, Zafari. F, DeCusatis. C, Wedaa. E, Yang. B et al. [12] Reduce security risks, it has been recommended that Machine Learning algorithms be used to identify potentially hostile connections and possible targets. These algorithms are trained on historical network attack data. Along with C4.5, the Bayesian Network (BayesNet), Decision Table (DT), and Naive Bayes algorithms are used. It is stated that utilizing Machine Learning methods, it is possible to identify illegal users at the data plane. Verifying user identity is crucial for network efficiency and stability since it enables the SDN controller to implement new rules to stop attacks in an effective manner.

N. F. Haq, A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah, and D.M. Farid, et. al. [13] The study was conducted to examine the use of single, hybrid, and ensemble classifiers in IDS (Intrusion Detection System) from 2009 to 2014 using Machine Learning techniques. The analysis is divided into three sections: the IDS construction methodology, a summary of the articles published within a given time period, and the datasets used for the analysis. According to the results of the investigation, using several classifiers is an effective and efficient technique to design an IDS.

Gisung Kim, Seungmin Lee, Sehun Kim et.al. [14] To detect DDoS attacks and protect OpenFlow switches, a hybrid learning model is developed.Using the C4.5

decision tree, a harmful detection model and atypical detection models were created, and it was determined that the model performs effectively even for unidentified attacks (1-class SVM).

B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Srndi c, P. Laskov, G. Giacinto, and F. Roli et. al. [15] Provide a straightforward technique based on the gradient method to create evasion attacks against malware detection systems for PDF files, with the presumption that adversary knowledge and capacity will grow. The SVM's detection performance was drastically decreased by the analysis are presented.

3. DDoS Saturation Attack Detection Model Using Machine Learning

Figure 1 below shows a block diagram for machine learning-based DDoS saturation attack detection and analysis in SDN architecture.

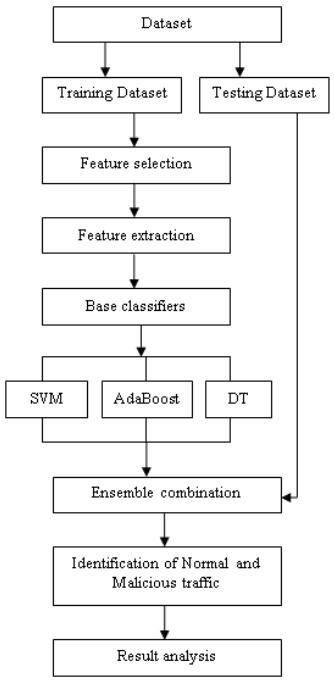

Fig. 1: DDOS Saturation Attack Detection Model

Traffic and datasets containing thousands of records of DDOS and general network traffic including TCP, UDP, ICMP packets, and community-based attack by the Center for Applied Internet Data Analysis (CAIDA 2016). There are 650 total occurrences of both benign and malicious traffic in the collection. In the SDN architecture, the DDoS attack is directed at multiple hosts, and Machine Learning algorithms support in identifying the malicious attack. 25% of the acquired data is used as a test set, while 75% is used as a training dataset.

Feature selection techniques were used to choose the features in the dataset. Some features have a greater impact on the classification results while some features have a lesser impact on the classification results. Time and process costs are also increased by using low-impact features for classification. The aim is to minimize the least effective features and use the most effective features for classification. In this model filter-based feature selection method is used and focuses on intrinsic features for their features. The contribution of feature selection to the achievement of feature results is calculated using filter-based statistical methods.

The fundamental thing for data classification is feature extraction. Four features are extracted using the feature extractor: (1) the number of Packet-In messages sent from OpenFlow switches to the SDN controller, (2) the number of Packet-Out messages sent from the SDN controller to the OpenFlow switches, (3) the number of Flow-Mod messages sent from the SDN controller to the OpenFlow switch and(4) TCP -numbers of ACK messages forwarded by the switches to the SDN controller, these four features are extracted by collecting the feature extractor OpenFlow.

A multi-classifier system is used to describe the ability of models to combine several Machine Learning algorithms and provide better accuracy than individual models that learn about a descriptive method. SVM, DT and AdaBoost Machine Learning classifiers were used.

Regression analysis and classification are utilised by SVM, a supervised learning model, to define and analyse patterns that are found in learning files. SVMs are used to find the hyper-plane that divides a collection of training data most effectively. The nonlinear regression case and the non-linear separable case are the two situations that need to be handled. The first scenario searches through the group of hyper-planes that split the training instances to find the best hyper-plane. The second problem is addressed by converting training data into a high-dimensional feature space using kernel functions.

One of the best and most used techniques for classification and prediction is the decision tree (DT). Decision nodes and leaf nodes, which represent the classifications that the tree is able to classify, make up DT. A decision tree's nodes each offer a test on a specific instance attribute, and each branch descending from the node represents one of the attribute's possible values. With a decision being made at the leaves and the flow starting at the root node, DT can be represented as a flowchart.

Ada stands for adaptable. The performance of this model improves as it gains knowledge from previous poor classifiers. It may be less susceptible to the over-fitting issue than other learning algorithms in particular situations. It can be shown that the final model converges to a strong learner even if each individual learner's performance is just slightly better to random prediction.

 https://doi.org/10.22362/ijcert/2022/v9/i12/v9i1206

Based on the value of the output flag, the traffic dataset is divided into two classes: attack and normal (0 or 1). When an attack occurs, the flow table (flag=1) warns controllers to drop a certain flow. For normal traffic packets the controller creates a routing path. Machine Learning criteria such as accuracy, recall, precision, and F1-score based structure are used to describe DDoS saturation attack detection performance.

4. Result Analysis

A tool for creating virtual hosts, controllers, switches, and hosts is called Mininet. A lightweight, Python-based controller called POX was chosen for the requested task. It is used to set up a wireless network in which one server randomly communicates with another server.MiniEdit was used to generate a network configuration.

Training data requires 75% of the original data, while testing data requires 25% of the original data.The performance of each classifier was then evaluated by extracting the qualitative and quantitative features. The recommended detection approach can detect TCP-SYN, ICMP and UDP saturation attacks. The identification system was evaluated during the attack by examining the following performance indicators. Malicious traffic injection is compared to normal traffic values to determine real positive and false negative numbers. The performance parameters in the presented model are precision, recall, accuracy, and F1-score.

Accuracy:

One of the most commonly used metrics for datasets is accuracy, which is determined using Equation 1 and is defined as the percentage of successfully categorized attacks.

Accuracy=(TP+TN)/(TP+TN+FN+FP)…(1)

Recall:

The term true positive rate, sensitivity, or recall refers to a measurement that indicates the proportion of positive cases.

Recall=TP/(TP+FN)…(2)

Precision:

As shown in Equation (3) below, positive predictive value or precision is the ratio of the number of accurate positive scores to the number of positive scores predicted by the classification method.

Precision=TP/(TP+FP)…(3)

F1-score:

The average of recall and precision is the F1-score. It must be expressed as zero for the classification method's poor performance and one for its strong performance in Equation (4) below.

F1-Score=2*(Precision*Recall)/(Precision+Recall)…(4)

Where,

1) True Negative (TN) -The percentage of regular flows that are appropriately identified.
2) True Positive (TP) -A percentage of attack packets that were correctly classified.
3) False Positive (FP) - The percentage of regular flows that are incorrectly classified as attacks.
4) False Negative (FN) - The percentage of attack streams that are incorrectly classified as authorized.

Performance of individual classifiers with combined classification method is elaborated in below Table 1.The graphical representation of the accuracy and precision parameters is shown in Figure 2, and the recall and F1-Score parameters are shown in Figure 3.

Table 1: PERFORMANCE OF DIFFERENT CLASSIFIERS

ML Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	91.1	90.2	89	90.5
DT	89.6	89.1	91.5	92.6
AdaBoost	90.2	92.3	92.3	93.3
Combined model (SVM+DT+ AdaBoost)	97.6	96.6	97.4	98

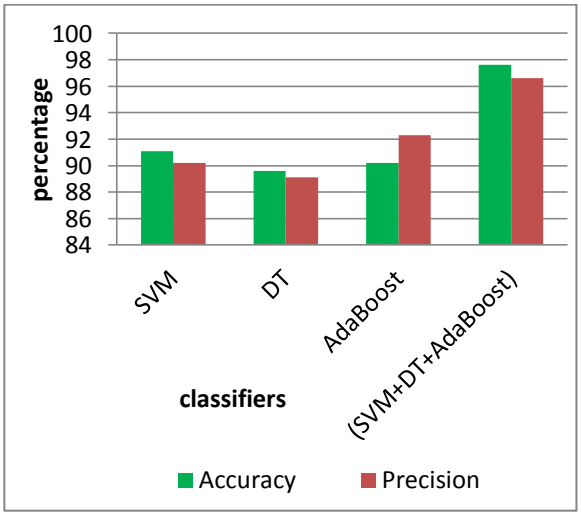

Fig. 2: Performance Analysis In Terms Of Accuracy and Precision

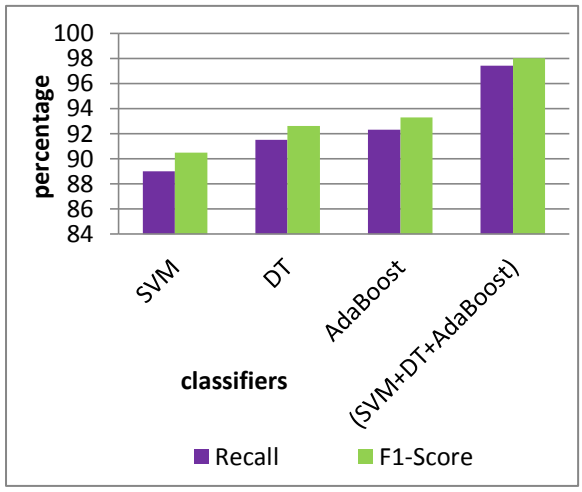

Fig. 3: Performance Analysis In Terms Of Recall and F1-Score

The results demonstrate that integrated Machine Learning outperforms independent classifiers in regards to Precision, Recall, Accuracy, and F1-Score metrics. Obtained Accuracy of combined Machine Learning (SVM+ DT+ AdaBoost) is 97.6% and similarly, Precision, Recall, F1-Score values are 96.6%, 97.4%, 98% respectively.

V. Conclusion

In this approach, Machine Learning based DDoS saturation attack detection and analysis is performed in an SDN environment. Datasets from the Center for Applied Internet Data Analysis (CAIDA 2016) include hundreds of records of TCP, UDP, and ICMP packets sent over the network both as part of DDOS attacks and as regular traffic. Training data requires 75% original data and testing data requires 25% original data. SVM, DT and AdaBoost Machine Learning classifiers were also used. The performance of the suggested technique is evaluated by comparing the Accuracy, Precision, Recall, and F1-Score of each classifier. The accuracy of combined Machine Learning (SVM+ DT+ AdaBoost) is 97.6% and similarly, the precision, recall, F1-score values are 96.6%, 97.4% and 98%, respectively.

References

[1] Shengli Du, Qiushuo Yan, Lijing Dong, JunfeiQiao, "Secure Consensus of Multiagent Systems With Input Saturation and Distributed Multiple DoS Attacks", IEEE Transactions on Circuits and Systems II: Express Briefs, Volume: 69, Issue: 4, Year: 2022

[2] Hamza Chahed, Andreas J. Kassler, "Software-Defined Time Sensitive Networks Configuration and Management", 2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Year: 2021

[3] Adel Alshamrani, "Reconnaissance Attack in SDN based Environments", 2020 27th International Conference on Telecommunications (ICT), Year: 2020

[4] Nitin Varyani, Zhi-Li Zhang, David Dai, "QROUTE: An Efficient Quality of Service (QoS) Routing Scheme for Software-Defined Overlay Networks", IEEE Access, Volume: 8, Year: 2020

[5] Jesús Arturo Pérez-Díaz, Ismael AmezcuaValdovinos, Kim-Kwang Raymond Choo, Dakai Zhu, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning", IEEE Access, Volume: 8, Year: 2020

[6] Abdullah H Almutairi, Nabih T Abdelmajeed, "Innovative signature based intrusion detection system: Parallel processing and minimized database",2017 International Conference on the Frontiers and Advances in Data Science (FADS), Year: 2017

[7] MarwaneZekri, Said El Kafhali, Noureddine Aboutabit, Youssef Saadi, "DDoS attack detection using machine learning techniques in cloud computing environments", 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), Year: 2017

[8] Alshamrani, A. Chowdhary, A. Pisharody, S. Lu, D. Huang,"A defense system for defeating DDoS attacks in SDN based networks", In Proceedings of the 15th ACM International Symposium on Mobility Management and Wireless Access, Miami Beach, FL, USA, 21–25 November 2017.

[9] Z. Abaid, M. A. Kaafar, and S. Jha, "Quantifying the impact of adversarial evasion attacks on machine learning based android malware classifiers," in 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA), IEEE, 2017, pp. 1–10.

[10] H. Lin and P. Wang, ''Implementation of an SDN-based security defense mechanism against DDoS attacks,'' in Proc. Joint Int. Conf. Econ. Manage. Eng. (ICEME), Int. Conf. Econ. Bus. Manage. (EBM), Philadelphia, PA, USA, 2016

[11] Barki, Lohit, "Detection of distributed denial of service attacks in software defined networks." International Conference on Advances in Computing, Communications and Informatics (ICACCI), Sep 2016, pp: 2576-2581

[12] Nanda. S, Zafari. F, DeCusatis. C, Wedaa. E, Yang. B,"Predicting network attack patterns in SDN using machine learning approach", In Proceedings of the 2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Palo Alto, CA, USA, 7–10 November 2016.

[13] N. F. Haq, A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah, and D.M. Farid, "Application of machine learning Approaches in Intrusion Detection system: A survey", international journal of advanced Research in Artificial Intelligence, vol. 4, no. 3, pp. 9 18, 2015

[14] Gisung Kim, Seungmin Lee, Sehun Kim "A novel Hybrid attackDetection method integrating Anomaly Detection with misuse Detection", - journal on Expert

Systems with Applications, 41(4), March 2014, pp: 1690-1700

[15] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Srndi c, P. Laskov, G. Giacinto, and F. Roli, "Evasion attacks against machine learning at test time", in Joint European conference on machine learning and knowledge discovery in databases. Springer, 2013, pp. 387–402

