DDOS Attack Detection Accuracy Improvement in Software Defined Network (SDN) Using Ensemble Classification

Alireza Shirmarz
Dep. of the Faculty of Eng
Ale-Taha University
Tehran, Iran
0000-0003-4296-0002

Ali Ghaffari
Dep. of Computer Faculty
Islamic Azad University
Tabriz, Iran
A.Ghaffari@iaut.ac.ir

Ramin Mohammadi
Dep. of Computer Engineering,
Ondokuz Mayıs University,
Samsun, Turkey
ram_moh1@yahoo.com

Sedat Akleylek
Dep. of Computer Engineering,
Ondokuz Mayıs University,
Samsun, Turkey
0000-0001-7005-6489

2021 International Conference on Information Security and Cryptology (ISCTURKEY) | 978-1-6654-0776-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISCTURKEY53027.2021.9654403

***Abstract*— Nowadays, Denial of Service (DOS) is a significant cyberattack that can happen on the Internet. This attack can be taken place with more than one attacker that in this case called Distributed Denial of Service (DDOS). The attackers endeavour to make the resources (server & bandwidth) unavailable to legitimate traffic by overwhelming resources with malicious traffic. An appropriate security module is needed to discriminate the malicious flows with high accuracy to prevent the failure resulting from a DDOS attack. In this paper, a DDoS attack discriminator will be designed for Software Defined Network (SDN) architecture so that it can be deployed in the POX controller. The simulation results present that the proposed model can achieve an accuracy of about 99.4% which shows an outstanding percentage of improvement compared with Decision Tree (DT), K-Nearest Neighbour (KNN), Support Vector Machine (SVM) approaches.**

Keywords—DDOS attack, SDN, POX controller, accuracy.

I. INTRODUCTION

The rate of Internet availability has been increased sharply; therefore, security measures are more required compared with the past. One of the most significant issues is to detect the DDOS malicious flows with more accuracy to avoid the resources from failure by discarding the malicious flows.

A. Software Defined Network

The current Internet situation is derived from the complexity of the traditional network architecture that makes the ongoing network configuration and network control impossible. Hence, the network specialists and scientists have proposed a new independent architecture called Software Defined Network (SDN) as the future Internet architecture [1]. The main idea of this architecture is the data and control flows separating. This architecture is composed of three layers including application, control and data. This architecture makes the network much more programmable, flexible, and manageable [2][3]. In addition to the three layers, there are three APIs that consist of northbound and southbound that are used to connect the application, controller and data layers while the east-west API is used to expand the number of controllers as Controller Placement Problem (CPP) [4][5][6]. This architecture caused the researchers to be able to propose a novel model for network security, performance and Quality of Service (QoS) improvement because traditional network architecture had limited the innovation to the network hardware vendors.

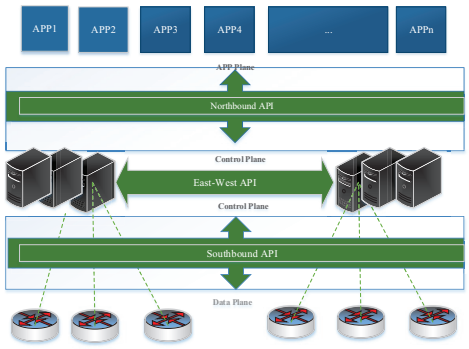

Fig. 1. SDN architecture

In this paper, SDN will be used for the novel DDOS attack detection method.

B. Problem Statement

Many approaches have been proposed to mitigate the DDOS attacks in the network that have been categorized [7]. These proposed methods can generally be classified into Machine Learning (ML)-based and statistical ones.

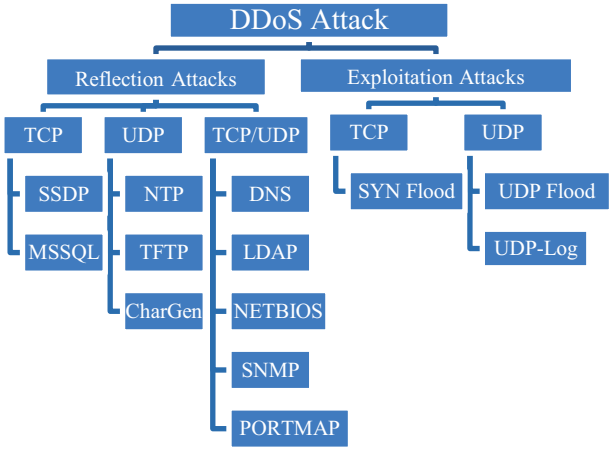

Fig. 2. DDOS attack taxonomy [7]

The DDoS attack is done using malicious network traffics making the network inaccessible with making the resources (server and bandwidth) overwhelmed by these malicious network flows. According to Cisco's annual internet report which has been published in 2020, the

number of DDoS attacks will double to 15.4 million by 2023 globally [8]. This report shows that DDoS attacks will increase in the future and should be paid attention to more than past. For instance, the DDoS attacks which happened in well-known companies and organizations like CNN, Netflix, Twitter caused a denial of service in 2006 [9]. The most significant component of the Security Operation Center (SOC) is to detect the DDoS malicious flows accurately in the first step. In the next step, these malicious flows should be discarded to make the resource safe so that they can provide their services. DDoS flow detection is an issue that will be addressed to ameliorate its accuracy in this paper. This module, to detect the DDoS malicious flows, will be implemented in the controller which is (physical or conceptual) centeralized in SDN.

C. Proposed Approach

In this paper, an ensemble classifier is proposed to improve the DDoS flows detection accuracy. Three classification algorithms, including decision tree, Neural Network (NN) and Support Vector Machine (SVM), are used to discriminate the DDoS flows in this paper. Finally, a boasting ensemble which is the combination of these three algorithms is proposed to improve the classification accuracy. The results of each classifier are aggregated with a voter. The model uses the dataset CICDDoS2019 to extract the model. The 90% of this dataset is considered as the train set and 10% of the dataset is used as a test set. The results show that the proposed model can classify the flows into legitimate or DDoS flows accurately. The results will be compared with accuracy, precision, recall and f-score parameters.

II. Related work

This section addresses the DDoS attack detection works which have been proposed in recent papers. This review is categorized into two sections, including statistical analysis methods and ML-based solutions.

A. Statistical Approaches

Statistical analysis has been used to detect the DDoS attack in SDN. B. Yuan and et al. have worked on a critical point in SDN which is the size of flow-table on which can be posed threat by DDOS malicious flows [10]. They have used practical mathematical ways to detect DDoS attacks.

Conditional entropy has been used to detect the DDoS attack in SDN and could reach the detection accuracy morae than 99% [11]. A. Ahalawat et al. have proposed an entropy-based method to detect DDoS attacks in SDN while they were using a Floodlight controller [12]. Entropy is a parameter to measure stochastic behaviour. More entropy shows that the behaviour of the flow is more likely to be normal. On the contrary, less entropy value indicates that the flow has an abnormal behaviour, so it should be considered as a threat.

A threshold is needed to be set for the value of entropy; hence, S. Mousavi et al. have proposed this method to detect the DDOS flow in the centralized controller [13]. They could detect the DDoS flow from the first five hundred received packets in the controller. In each network, every node must receive new packets with the same probability that this condition can cause high entropy. If one or some nodes are more probable to receive packets, the entropy decreases and it states a DDoS attack probability. S. Mousavi et al. have worked on the DDoS attack and they have proposed a threshold for entropy and stated that the value below the threshold expresses the DDoS flow likelihood [14]. The entropy-based solutions can work as mentioned in the reviewed paper, but it is based on the expert person to set the threshold for the value of entropy. This threshold has been addressed in the related work. In the following, machine learning approaches will be expressed.

B. ML-Based Approaches

The growth of machine learning knowledge and its penetration among researchers has caused many to use ML to detect DDoS attacks. To use machine learning, real and appropriate datasets are required to be used because ML needs a strong train set to extract the pattern of DDoS malicious flows' behaviour. The feature selection which is a significant stage in data mining and ML has been focused on by M. Wang et al. [15]. They have proposed a dynamic method to select the optimal features. They have used Multi-Layer Perceptron (MLP) to implement sequential feature selection for DDoS flows' detection process. Machine learning techniques in cybersecurity are helpful by recommending the proper decision for analysis and even doing the proper action automatically, such as artificial neural networks (ANN), Bayesian networks, Decision trees (DT), clustering, ensemble and so on. Phan et al. have proposed a Distributed Self-Organizing Mapping (DSOM) method to detect DDoS attacks in SDN [16]. Each DSOM in each switch processes the incoming traffic; hence, the processing load on the controllers will be divided between the switches. There is also a DSOM component in the application layer that is responsible for managing the performance of DSOM on switches. In this method, the attack detection point is located in the switches and the data layer, it is necessary to check all the packets passing through the switch by SOM, which consumes a lot of processing time. Braga et al. in [17] introduce the lightweight DDoS flooding attack method that detects a DDoS flow based on tracking suspicious input flows using. The control plane architecture is centralized with a NOX controller. M. Dey et al. have examined varied Anomaly detection accuracy and achieved that Deep Neural Network (DNN) could express better accuracy in intrusion detection compared with random forest and LSTM [18] in OpenFlow-based SDN. R. Santos et al. have worked on four classifiers, including SVM, MLP, Decision Tree, and Random Forest, to discriminate DDoS flows in SDN [19]. According to the reviewed papers, the researchers in the network have strived to propose a novel method to detect the DDoS malicious flows with more accuracy. To improve the accuracy, two principle proposals consisting of more real datasets including DDoS attacks and a more accurate classifier. In this paper, we propose a boosting ensemble to improve the classification accuracy, in order to extract an accurate model to detect DDoS flows using existing public datasets.

III. Proposed DDoS attack classification model

In this section, the model that is proposed will be discussed in three subsections including dataset, feature extraction and classifier.

A. Training and Test Sets

This paper uses the CICDDoS2019 dataset [7] collected by the University of New Brunswick. This dataset includes normal traffic and the most up-to-date distributed denial-of-service attack traffic including various types of attack traffic such as DNS, LDAP, SYN, etc. To extract the feature, the CICFlowMeter two-way traffic flow generation tool was used. The output of this tool contains 76 features. The samples distribution of the dataset is presented in Table 1.

TABLE I. SAMPLES DISTRIBUTION IN THE TRAINING SET

Sample Type	*The number of Samples*
Legitimate	56863
DDoS_DNS	5071011
DDoS_LDAP	2179930
DDoS_MSSQL	4522492
DDoS_NetBios	4093279
DDoS_NTP	1202642
DDoS_SNMP	5159870
DDoS_SSDP	2610611
DDoS_SYN	1582289
DDoS_TFTP	20082580
DDoS_UDP	3134645
DDoS_UDP_Lag	366461
DDoS_webDDoS	439

This dataset has two classes that consist of normal flow or DDoS malicious flow which had been labelled before. Given that the number of DDoS flows is less than normal flows, so this set is unbalanced and causes the probability of normal flows to increase. Therefore, among the unbalanced sets methods include oversampling, undersampling and weight sampling, we use weight sampling in this paper. To make the samples balanced, weight is considered for DDoS flows as mentioned in (1) and normal flows as shown in (2). These weights make the samples balanced; hence, it can make the probability of DDoS flows and normal flows equal.

$$W_{DDDoS\ Flows} = \frac{No_{Total\ Samples}}{2 \times NO_{DDoS\ Flows}} \quad (1)$$

$$W_{Legitimate\ Flows} = \frac{No_{Total\ Samples}}{2 \times NO_{Normal\ Flows}} \quad (2)$$

In this paper, the training set is 90% of samples and 10% of samples is used to test the model in the evaluation phase. In the following, the feature extraction component is discussed.

B. Feature Extraction & Selection

As mentioned, this dataset contains 76 features. All these features can be used for classification, but their importance is not equal while the high number of features can cause the training and testing phase. Therefore, the number of features decreases to the number 24 in this paper using the significant component analysis (PCA) method.

C. Boosting Classification For DDoS Detection

There are many algorithms for classification; DT is a simple algorithm that can classify with high accuracy. KNN is another simple algorithm that is less complex in comparison with other algorithms. SVM can classify accurately but its training phase takes more time compared with other classification algorithms.

In this paper, three classifications including DT, KNN and SVM classify the normal flows and DDoS malicious flows. There are two fundamental methods to improve the classification accuracy called ensemble that includes bagging and boosting. The ensemble bagging does classification with divided data set into smaller groups of samples, so the result will be determined with average or voting. The boosting ensemble uses more than one classifier and the result will be presented with voting. The proposed model is shown in Fig. 3.

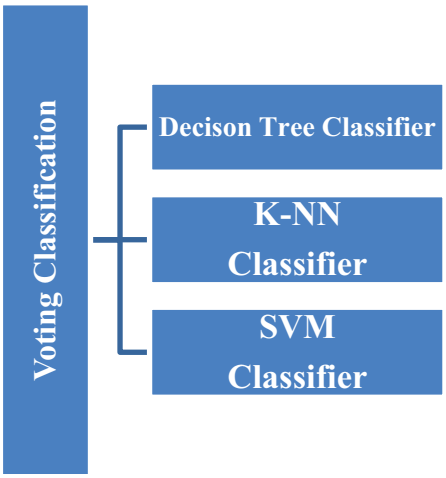

Fig. 3. The proposed boosting ensemble for DDoS detection

In this model, the model has been trained with the training set to extract the model with three classifications separately. Three trained models are combined and the result is the major result. On the other hand, the flow that is detected as the DDoS traffic with two of these three models is the DDoS attack and should be discarded. The model DDoS attack detection performance is evaluated with accuracy, precision, recall and f-score.

IV. EXPERIMENT & SIMULATION

The simulation condition is done with the following hardware and software situation that is presented in Table II.

TABLE II. EXPERIMENTAL SETUP

Hardware & Software	*Specs*
CPU	CPU Corei72.5 GHZ
Ram	12 GB
Python IDE	Spyder
Programming Language	Python
Package	Scikit learn

A. Evaluation Variables & Metrics

The proposed mode needs to be evaluated with the metrics which have been defined in Table III.

TABLE III. EVALUATION VARIABLES

Evaluation Variables	*Definition*
True Positive (TP)	Attack traffic identified as an attack
False Positive (FP)	Attack traffic detected as normal
True Negative (TN)	Normal traffic that is identified as normal
False Negative (FN)	Normal traffic identified as an attack

The accuracy, precision, recall and f-score are calculated as presented in (2), (3), (4) and (5).

$$\text{Accuracy} = \frac{tp+tn}{tp+fp+tn+fn} \quad (2)$$

$$\text{Precision} = \frac{tp}{tp+fp} \quad (3)$$

$$\text{Recall} = \frac{tp}{tp+fn} \quad (4)$$

$$\text{F-Score} = \frac{2 \times P \times R}{P+R} \quad (5)$$

B. Classification Results

The simulation results for each classifier are reported in Table IV. These results refer to the simulation which has been done under the simulation condition explained in this section.

TABLE IV. CLASSIFICATION RESULTS

Algorithm	***Training Set***		***Evaluation Set***	***Test Set***
	Accuracy %	F-Score %	Accuracy %	Accuracy %
DT	93.6	92.5	92.3	91.2
K-NN	95.67	95.83	97.70	96.3
SVM	92.42	93.12	91.6	90.7
Boosting Model	99.3	99.6	99.3	99.4

The simulation results show that the proposed model can detect the DDoS attack with an accuracy of more than 99.4% which is very high accuracy. The boosting ensemble could improve the accuracy near the 100% which is very significant.

C. Classification Accuracy & F-Score chart

The classification metrics including the accuracy of the training phase presented in Fig. 4.

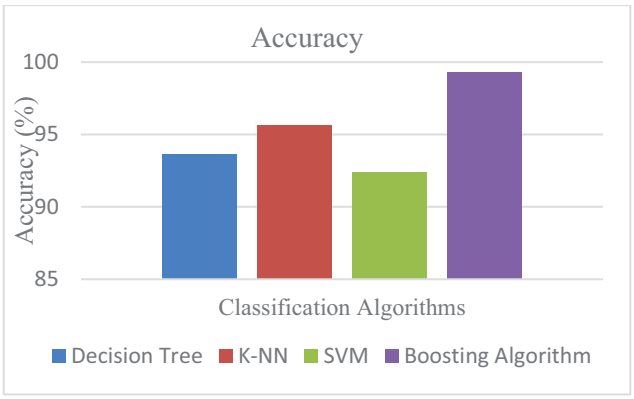

Fig. 4. Accuracy of evaluation samples in the training phase

The other metric is F-score which is evaluated in the training phase in Fig. 5.

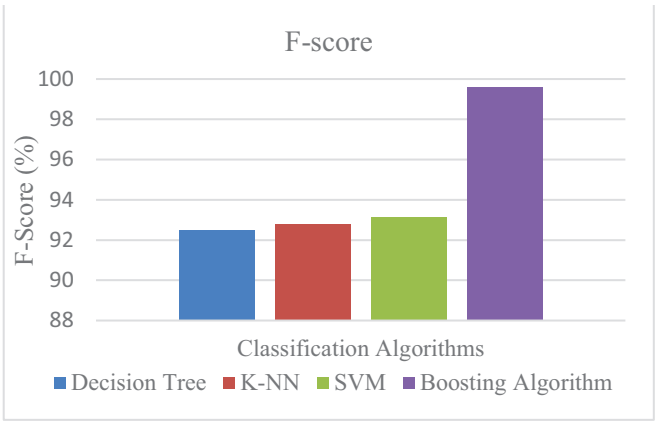

Fig. 5. F-Score of evaluation samples in the training phase

The training phase is done with cross-validation to avoid the model from overfitting. The cross-fold-10 is used to train and evaluate three models and the proposed model. The evaluation that is reported for each classifier is the average evaluation value for 10 iterations that have been done.

Finally, three classifiers and the proposed model are evaluated with accuracy that shows how accurately the model can detect the DDoS malicious flows meticulously.

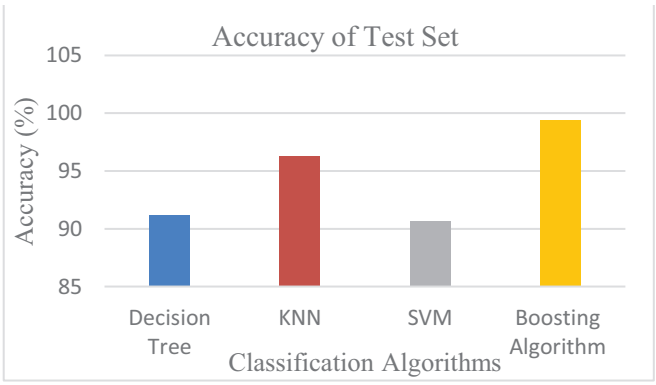

Fig. 6. The DDoS flows detection performance for training & test set

The results show that the proposed model can detect the DDoS flows with an accuracy of more than 99% which is better performance in comparison with the three other classifiers. This model can achieve accuracy near 100 percent that is the network security challenge nowadays.

V. Conclusion

In this paper, a boosting ensemble including decision tree, K-NN and SVM has been proposed to improve the DDoS attack detection accuracy. To extract the pattern which exists in the DDoS malicious flows, the dataset called CICDDoS2019 has been used. The dataset had been gathered for DDoS flows. The dataset is imbalanced; hence, the weighted method has been used to make the training set balanced.

The DDoS detection model has been trained with DT, K-NN and SVM. The boosting ensemble has been proposed to improve classification performance including accuracy of DDoS flows detection.

The training set is limited; therefore, the cross-validation technique has been used in the training phase. The result showed that the accuracy of malicious traffics detection has been improved by a prominent amount. This model has been implemented in the POX controller and is tested. The POX had been developed with Python version 2, so the proposed model has been converted into version 2. This conversion is a challenge in this paper. The proposed model has been used three DT, KNN and SVM. The main reasons for these classification algorithms selections are based on training phase time consuming, algorithm simplicity and accuracy that had been evaluated in the previous research. The other classification algorithms' combination is proposed for future works, and the malicious flows detection time in addition to its accuracy can be considered for future work.

References

[1] A. Shirmarz and A. Ghaffari, "An Autonomic Software Defined Network (SDN) Architecture With Performance Improvement Considering," *J. Inf. Syst. Telecommun.*, vol. 8, no. 2, pp. 1–9, 2020.

[2] A. Shirmarz and A. Ghaffari, "Performance issues and solutions in SDN-based data center: a survey," *J. Supercomput.*, vol. 76, pp. 7545–7593, 2020.

[3] A. Shirmarz and A. Ghaffari, "An adaptive greedy flow routing algorithm for performance improvement in a software-defined network," *Int. Numer. Model. Electron. networks, Devices, Fields-Wiley online Libr.*, vol. 33, no. 1, pp. 1–21, 2019.

[4] A. Shirmarz and A. Ghaffari, "Taxonomy of controller placement problem (CPP) optimization in Software Defined Network (SDN): a survey," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–26, 2021.

[5] G. Ramya and R. Manoharan, "Enhanced Multi-Controller Placements in SDN," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–5, 2020.

[6] K. Sood and Y. Xiang, "The controller placement problem or the controller selection problem?," *J. Commun. Inf. Networks*, vol. 2, no. 3, pp. 1–9, Sep. 2017.

[7] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2019-Octob, 2019.

[8] T. Cisco and A. Internet, "Cisco Annual Internet Report," 2020.

[9] "Defending against Distributed Denial of Service (DDoS) attacks," 2020. [Online]. Available: https://www2.deloitte.com/ca/en/pages/risk/articles/DDoSattacks.html.

[10] B. Yuan, D. Zou, S. Yu, H. Jin, W. Qiang, and J. Shen, "Defending against flow table overloading attack in software-defined networks," *IEEE Trans. Serv. Comput.*, vol. 12, no. 2, pp. 231–246, 2019.

[11] M. Xuanyuan, V. Ramsurrun, and A. Seeam, "Detection and mitigation of DDoS attacks using conditional entropy in software-defined networking," *Proc. 11th Int. Conf. Adv. Comput. ICoAC 2019*, pp. 66–71, 2019.

[12] A. Ahalawat, S. S. Dash, A. Panda, and K. S. Babu, "Entropy Based DDoS Detection and Mitigation in OpenFlow Enabled SDN," *Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking, ViTECoN 2019*, pp. 1–5, 2019.

[13] S. M. Mousavi and M. St-hilaire, "Early Detection of DDoS Attacks against SDN Controllers," in *International Conference on Computing, Networking and Communications, Communications and Information Security Symposiu*, 2015, pp. 77–81.

[14] S. M. S. Mousavi and M. St-Hilaire, "Early Detection of DDoS Attacks in Software Defined Networks Controller," *Thesis*, pp. 77–81, 2014.

[15] M. Wang, Y. Lu, and J. Qin, "A dynamic MLP-based DDoS attack detection method using feature selection and feedback," *Comput. Secur.*, vol. 88, p. 101645, 2020.

[16] T. V. Phan, N. K. Bao, and M. Park, "Distributed-SOM: A novel performance bottleneck handler for large-sized software-defined networks under flooding attacks," *J. Netw. Comput. Appl.*, vol. 91, pp. 14–25, 2017.

[17] R. Braga, E. Mota, and A. Passito, "Lightweight DDoS flooding attack detection using NOX/OpenFlow," in *Proceedings - Conference on Local Computer Networks, LCN*, 2010, pp. 408–415.

[18] T. Khalil, "A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning," pp. 372–378, 2014.