L-015

A Case Study to Detection DDoS Attacks Using a Support Vector Machine Integrated with Random Forest in SDN Network

Yuqi Li† Yuichi Goto†

1. Introduction

Software Defined Networking (SDN) is a network architecture based on software, where a centralized controller manages all network traffic using protocols like OpenFlow to communicate with switches at the infrastructure layer [5]. This helps network engineers to control and manage their network from a centralized controller and to troubleshoot the network at a better pace. With the benefits that SDN provides, it also has some vulnerabilities to attacks like Distributed Denial of Service (DDoS) that can disrupt the availability of the network [4]. DDoS attacks are a significant threat to internet services and can have devastating consequences on website and web application availability, often leading to shutdowns. The financial implications of such attacks can be dire for businesses that rely on internet-based operations. The disruption of communication channels, including access to critical emergency and financial systems, further underscores the significance of DDoS attacks. DDoS attack detection is becoming more and more important.

Singh et al. [4] demonstrated that the detection rate of a model for DDoS detection model using support vector machine integrated with random forest (SVM+RF for short) is significantly higher than that of individual models, achieving an accuracy of up to 99.1%. However, it still faces issues such as insufficient flexibility in feature selection and low detection efficiency.

In this paper, to improve the low detection efficiency, we proposed a new feature set for SVM+RF and confirmed the usefulness of the selected feature set by the experiment on the Mininet. Firstly, we investigated 54 features used by Singh et al. [4] and Mohammad et al. [1], and then we chose features based on the transmission rate of traffic. Finally, we compared detection efficiency between the original features in Singh et al. and our selected features by the experiment on the Mininet.

By simplifying feature selection, our study not only reduces computational complexity and data processing burden but also demonstrates higher detection rates (99.22%) and recall rates (98.54%) in practical applications. These improvements make our SVM-RF model not only excellent in detection performance but also more convenient and efficient in practical use.

2. Related Work

Singh et al. [4] proposed a new technique for detecting and mitigating DDoS attacks in SDN using a hybrid machine learning model (SVM-RF). By constructing their own DDoS dataset and experimenting with multiple machine learning algorithms, they demonstrated the advantage of the hybrid model in detection accuracy. However, the paper still has some shortcomings in feature selection. Firstly, some selected features have high correlations (e.g., packet length and packet count), which increases computational complexity and may lead to model overfitting. Secondly, some features (e.g., source IP, destination IP) have low effectiveness in scenarios involving dynamic IP allocation and IP address spoofing, which hinders the stability of detection performance in dynamic network environments [3].

Mohammad et al. [1] adopted a hybrid model combining unsupervised and supervised learning, using the CICDDOS2017 dataset as the training data without needing to label the data. However, this approach has several drawbacks: the use of a large number of feature vectors can make the model very complex, and training with too many types of feature vectors can lead to overfitting, increased computational complexity, and excessive resource consumption. Additionally, the unsupervised learning method may perform poorly in handling noise and outliers, which can adversely affect the overall performance of the model.

3. Proposed Method

This paper continues to use the SVM-RF model from Singh et al. and refers to the feature selection methods of Mohammad et al. and Singh et al. In our research, we focus on the rate of change in traffic transmission and the fluctuations in the overall number of traffic entries. Through simple calculations, we have formed the following seven features as the new feature set for this paper.

1) SFE (Speed of Flow Entries): calculate the change in the number of traffic entries between two calls. This can help monitor the rate of increase or decrease in network traffic.
2) SSIP (Speed of Source IPs): monitor the rate of change in the number of unique source IP addresses. This helps identify new devices or changes in the network.
3) RFIP(Ratio of Flow Pair): calculate the ratio of bidirectional traffic pairs, i.e., whether there is data transmission in both directions (source and destination IP).
4) Total Byte Count: calculate the total number of bytes in all flows.
5) Total Packet Count: calculate the total number of packets in all flows.
6) Average Byte Per Flow: calculate the average number of bytes per flow.
7) Average Packet Per Flow: Calculate the average number of packets per flow.

4. Experiment

4.1 Experimental environment

To implement attack traffic monitoring, we used Mininet 2.2.2-5 and Ryu 4.30 to simulate an SDN network architecture on Ubuntu 20.04 within VMware. We employed sFlow-rt 3.0-1703

† Graduate School of Science and Engineering, Saitama University

to monitor the traffic of hosts in the SDN network topology. A detection system based on SVM-RF was developed in Python 3.7 and operated within the Ryu controller.

4.2 Experimental Method

The experiment is divided into two phases: data collection and training phase and traffic detection and anomaly mitigation phase.

Phase One: we used traffic generation scripts written with the hping3 tool to simulate attacks on the target host

Table 1 Features for SVM+RF

	Proposed	Original [4]
Num. of features	7	12
Type of feature	SFE (Speed of Flow Entries) SSIP (Speed of Source IPs) RFIP (Ratio of Flow Pair) Total Byte Count Total Packet Count Average Byte Per Flow Average Packet Per Flow	Source IP Destination IP Source port Destination port Highest layer Transport layer Packet length Time Packet count Timestamp IP flag Transport flag
Feature Source	➢ Changes in flow rate and volume ➢ Data size	➢ Flow sources and destinations ➢ Network protocol ➢ Time node

and captured the network traffic data using Wireshark. This process generated a dataset containing both normal and DDoS attack traffic. To train the model, we manually labeled the simulated attack traffic and converted the generated pcap files to CSV format for subsequent use.

Phase Two: we referred to the study in reference [4] and utilized the open-source dataset CICDDoS2019. We replayed traffic on hosts within a custom topology to simulate real-world data transmission. Using the TCP replay tool on Ubuntu, we replayed traffic in the SDN network and employed the trained SVM-RF hybrid model for detection. The system allowed normal traffic to pass through, while abnormal traffic triggered the system's traffic mitigator, which deployed flow rules to block abnormal traffic from the source port.

4.3 Experimental Results

Table 2 Performance evaluation of different algorithms

Model	Accuracy	Detection Rate	False Alarm Rate	Recall Rate
Proposed	98.5%	99.22%	0.015%	98.54%
SVM-RF	99.1%	98.32%	0.011%	98.32%
SVM	85.99%	87.77%	0.127%	87.77%
RF	97.31%	95.63%	0.034%	95.63%

The experimental results are presented in Tables 2. We compared 4 pairs of a model and a set of features: “Proposed” is (SVM-RF, our proposed features), “SVM-RF” is (SVM-RF, original features), “SVM” is (SVM, original features), and “RF” is (RF, original features).

The advantage of this paper lies in its use of fewer feature vectors, achieving efficient detection with only seven features. The features used in–Singh et al.'s study, while comprehensive, may be inadequate in scenarios involving dynamic IP allocation and IP address spoofing, especially for features like source IP and destination IP. In contrast, the features selected in this study focus more on flow rate, flow contrast ratio, total bytes, total packets, average bytes, and average packets. These features more directly reflect abnormal traffic changes, resulting in higher sensitivity and lower false positive rates in detecting DDoS attacks.

5. Conclusion

Despite the superior performance of hybrid models in detecting DDoS attacks compared to single models, many challenges remain. In the future, different datasets and problem types may require distinct fusion strategies, making the selection of an appropriate fusion strategy crucial.

Reference

[1] N. Mohammad, S. Zarifzadeh, S. Mostafavi, "A hybrid machine learning approach for detecting unprecedented DDoS attacks", J Supercomput 78, pp. 8106 – 8136 (2022).
[2] N. Mohammad, S. Zarifzadeh, S. Mostafavi, "DDoS attacks and machine-learning-based detection methods: A survey and taxonomy", Engineering Reports, Vol. 5, NO. 12(2023).
[3] O. Rahman, M. A. G. Quraishi, C. -H. Lung, "DDoS Attacks Detection and Mitigation in SDN Using Machine Learning," 2019 IEEE World Congress on Services (SERVICES), pp. 184-189 (2019).
[4] A. Singh., H. Kaur, N. Kaur, "A novel DDoS detection and mitigation technique using hybrid machine learning model and redirect illegitimate traffic in SDN network", Cluster Comput, Vol.27, No.2 (2024).
[5] I. Sharafaldin, A. H. Lashkari, S. Hakak, A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy", 2019 International Carnahan Conference on Security Technology (ICCST), pp. 1-8 (2019).
[6] P. S. Saini, S. Behal, S. Bhatia, "Detection of DDoS Attacks using Machine Learning Algorithms," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 16-21 (2020).
[7] A. R. Wani, Q. P. Rana, U. Saxena, N. Pandey, "Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques", 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 870-875 (2019).