Low rate DDoS Attack Identification and Defense using SDN based on Machine Learning Method

Rameshwari Khamkar, Kshitij Thakre, Aditya Kotkar, Priti Jadhav,Rohini Hanchate

D.Y. Patil Institute of Engineering, Ambi

Abstract: Software defined networking architectural framework eases the life of the network administrators by isolating the data plane from the control plane. This facilitates easy configuration of the network, provides a programmable interface for developing applications related to management, security, logging etc. and the centralized logical controller gives more control over the entire network, which has the total visibility of the network. These advantages of SDN also expose the network to the vulnerabilities and therefore the impact of the attacks are much severe in comparison to standard networks, where the network devices in itself provided protection against different attacks and limits the scope of the safety threats. During this paper, we explore various attacks which will be launched on SDN controller at different layers and secure the SDN against threats.A Distributed Denial of Service (DDoS) attack may be a DoS attack utilizing multiple distributed attack sources. Increase in randomness causes decrease in vulnerabilities of system. To extenuate this threat, this paper proposes to use different techniques for the central control of SDN for various attack detection and introduces an answer that's effective and light-weight in terms of the resources that it uses. More precisely, this project shows how DDoS attacks can exhaust controller resources and provides an answer to detect such attacks supported the variation of the destination IP address. Traffic characteristics through statistical flow table information and uses the support vector machines (SVM) method to identify the attack traffic.The experiment is conducted using KDD99 dataset.

Keywords—Software Defined Network (SDN), Distributed Denial of Service (DDoS), Machine Learning (ML), Support Vector Machines (SVM), security threats.

I. Introduction

Security has been regarded as the dominant barrier of the development of Internet service. Denial of Service (DoS) attacks and Distributed Denial of Service(DDoS) attacks are the main methods to destroy availability of Internet service. DDoS attacks refer to the use of client/server technology to combine multiple computers as an attack platform to launch a DoS attack on one or more targets. Thus, the power of DoS attacks mainly used IDC is multiplied to forge source IP attacks. DDoS attacks are common in these years. These attack incidents incurred heavy downtime, business losses, to name but a few. There are some noted attack examples. In 2015, Lizard Squad attacked cloud-based game services of Microsoft and Sony, leading to the decline of QoS on Christmas day. Cloud service provider, Rackspace, was targeted by a massive amount of DDoS attack on its servers. Amazon EC2 cloud servers were attacked by a massive DDoS attack[1]. Thus, strengthening DDoS attack detection and defense isan urgent task. The security of the campus network is paid much attention by the government[2]. Denial of Service (DoS) attacks and Distributed Denial of Service(DDoS) flooding attacks are the main methods to destroy availability of campus network. In traditional networks, hardware and software applications based on DDoS attack detection and defense are expensive and difficult to deploy[3]. Software Defined Network (SDN) has attracted great interests as a new paradigm in the network. In SDN, the control planes and data planes are decoupled. Network intelligence and Network state are logically centralized. The underlying SDN infrastructure is abstracted from the specific applications. SDN can improve network manageability, scalability, controllability and dynamism[4]. Thus, SDN can dynamically modify forwarding rules to defend DDoS traffic and improve network security. To mitigate the DDoS attacks and reduce the restrictions, traffic classification needs to be performed to identify attack traffic. Machine learning technology based network traffic classification has become a hot topic and has achieved encouraging results in intrusion detection[5].In this paper, we propose an SDN framework to identify and defend DDoS attacks based on machine learning for the campus network. This system framework consists of 2 phases which are network traffic collection module, DDoS attack identification module.Traffic collection module extracts network traffic characteristics to prepare for traffic identification in the system. The Support Vector Machine (SVM) is applied to identify the DDoS traffic. The Ryu controller[6] is employed to build the flow table decision delivery module.

The objective of this paper is as follows:

➢ Combining the characteristics of the SDN network, we propose network features that are easy to extract in the SDN environment.

- Abstracting the DDoS attack detection problem as an attack traffic classification problem and using SVM to establish a classification detection model.
- Designing and implementing the attack detection and prevention framework using Ryu controllers.

II. Problem Statement

With a listing of different security concerns in Software Defined Networks, one of the main security threats we are concentrating upon in this research work is on Distributed Denial-Of-Service. When a large number of packets are forwarded to a network device with an intent to either stop the service or decrease the performance then such attacks are termed as Distributed Denial-of-service attacks.In DDoS attacks, an outsized number of packets are sent to a number or a gaggle of hosts during a network. If the source addresses of the incoming packets are spoofed, which they typically are, the switch won't find a match and has got to forward the packet to the controller. The collection of legitimate and therefore the DDoS spoofed packets can bind the resources of the controller into continuous processing that exhausts them. This will ultimately causes controller to be unreachable for the newly arrived valid packets and may crash the controller causing the loss of the SDN architecture. Even if there is another one a backup controller in the SDN, it has to face the same issue as the previous one.

This kind of attacks can be detected at an early stage by monitoring few hundreds of packets based on the entropy changes. The early detection of DDOS attack prevents the controller going down. The term "early" is subjected to tolerance level and traffic being handled by the controller [7] [8]. If detection happens early say first few hundreds of packet then, the impact of flooding of malicious packets can be controlled significantly. The early detection mechanism must be of light weight and should have a high response time. The high response time saves the controller in the period of attack to regain the control by terminating the DDOS attack.

III. Motivation

The main objective of this research is detecting a DDoS attack in its early stages using various machine learning techniques. The term early depends on the network itself. Since the controller software are often run on a laptop or a strong desktop, the term early would depend upon the tolerance of the device and traffic properties. However, if the detection happens within the first few hundred packets, the mitigation is applied before the controller is totally swamped with the massive number of malicious packets.

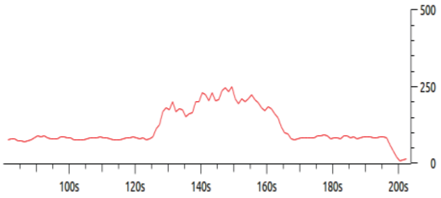

Figure 1 Sample attack on the controller

Figure 1 shows a high DDoS attack on the SDN controller where the normal incoming packet rate is around 100 packets per second. When the attack happens, the speed rises sharply to, approximately, 250 packets per second. The simulated DDoS attack was directed to a SDN controller that's connected to a network with 64 hosts and nine switches. This attack lasted for 40 seconds and sent over 500 packets with spoofed source addresses all destined for one host over nodes. For the aim of this research, all packets will have spoofed IP addresses. Because of this way, the switches do not have a match and all the packets are sent to the controller and works effectively.To accomplish this goal, a fast and effective method is needed that works within the controller.One of the functions of the controller is collecting statistic.

In this study, this attribute is used for adding another set of statistics collection to the controller; various destination IP addresses.

I. DDOS ATTACK *IDENTIFICATION* AND DEFENSE FRAMEWORK

A. Scenario Assumptions

Figure 2 shows the simplified illustration of the system architecture in the hypothetical scenario. The system is composed of a web server, an SDN controller and the DDoS attack identification module running on the controller and two OpenFlow switches. In addition, there are some normal visitors and some attackers. In order to better describe the DDoS attack identification and defense framework, we give the following assumptions about the above system architecture.

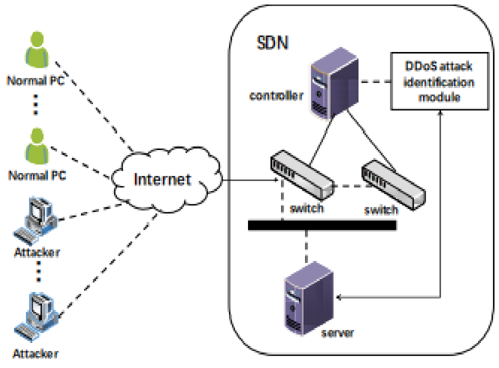

Figure 2 The system architecture in the hypothetical scenario

- All attackers come from external networks.
- DDoS attacks are HTTP flood attacks against web servers.
- Web server is used to simulate the website of a university.

B. The detection process

The detection process is divided into two steps. Firstly, the IP entropy is detected to determine whether a DDoS attack has been generated. If the IP entropy detection result is a DDoS attack, the system sets the flag to 1. After that, the traffic collection module performs feature extraction based on the flow table entries and the message packets when the flag is 1. Then the DDoS attack detection is applied to perform DDoS recognition. When a DDoS attack is identified the controller sends out the flow table to filter this packet. The detection process is done.

C. The DDoS attack identification and defense framework

According to the above DDoS attack detection process, we propose a DDoS attack identification and defense framework showed in Figure 3.The framework consists of 3 parts which are traffic collection, DDoS attack identification and flow table delivery. The sniffer is used to collect statistical information from the package-in message and flow tables then convert statistical information into a feature vector which the classifier can handle. Assuming that the originator of the attack is all in the external network, the sniffer only needs to check the flow table of the switch1 connected to the external networks and the package-in message initiated by this switch. The attack identification model passes the recognition result to the flow table delivery model. The flow table delivery model conducts the control strategy: The traffic will be forwarded as usual unless it is dropped because of having the DDoS attack packet.

II. THE DDOS TRAFFIC IDENTIFICATION MODEL

A. Feature Extraction

When a packet arrives at the switch, if there is no matching rule in the flow entry, the switch will send a package-in message to the controller. After the controller receives the package-in message, it will design the forwarding rule through the internal decision and return flow table to the switch with the package-out message. After the switch updates the flow table, the packets are processed according to the matching rules. The traffic statistics information can be obtained from the package- in message and flow table.

Combining the characteristics of the OpenFlow flow table,we extract 8 features based on the original data set features. Table1 depicts the features. All data packet features are collected and extracted by the sniffer.

Table 1 The description of features which we extract

Label	Description
count	In the past two seconds, the number of connections between the target hosts and the current connections is the same.
srv_count	In the past two seconds, the number of connections with the current connection has the same service.
same_srv_rate	In the past two seconds, the percentage of connections that have the same service as the current connection has the same service as the current connection.
dst_host _coun t	In the first 100 connections, the number of connections with the same target host is the same as the current connection.
dst_host _srv_ count	In the first 100 connections, the number of connections that have the same target service as the current connection is the same as the current connection.
dst_host_same_src_port_rate	In the first 100 connections, the connection with the current connection has the same target host, the same service and the same port.
dst_host_serror_rate	In the first 100 connections, the percentage of SYN error connections is the same as the current connection with the same target host.
dst_host_ serror_rate	In the first 100 connections, the percentage of REJ error connections is the same as the current connection with the same target host.

B. Support Vector Machine based DDoS traffic Identification Model

The real-time requirement of DDoS attack detection is relatively high. Support Vector Machine (SVM) is a supervised learning method with associated learning algorithms that analyze data used for classification and regression analysis. In this study, the main goal is to classify each packet as an attacker or a normal one. Therefore, we select SVM to establish the DDoS attack recognition model. SVM has a higher robustness than other machine learning algorithms. The traffic classifier construction process is showed in figure3.

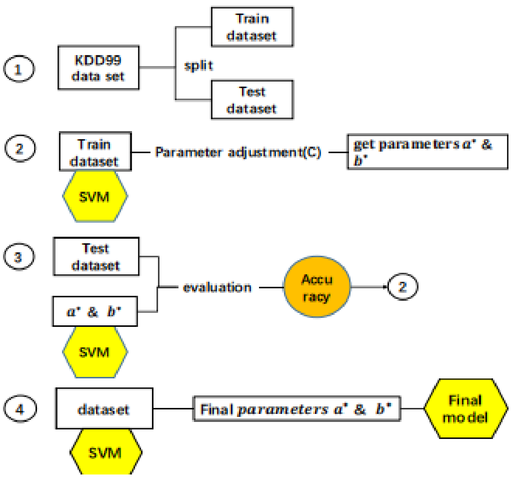

Figure 3 The traffic classifier construction process

Firstly, the raw dataset is separated into train dataset and test dataset. Then the training dataset is used to build the DDoS attack recognition model. And determining the best parameters through repeated tests.

The traffic data is collected from the flow table based flow table entries. The traffic dataset *X* has *N* traffics *X*={*x1,x2,x3,.., xn*}, and *xi* denotes a TCP connection that is made up of 8 features. These features denote the host-based network traffic features and time-based network traffic features. We use -1 to represent “attacker packet” for the packet from the attacker pool, and we use 1 to represent “normal packet” for the packet from normal pc.

Step 1: we use SVM to solve the following the optimization problem. The linear kernel function *K(x,y)* and the appropriate parameter *C* is selected to build the SVM model. The linear kernel function K(x, y) = x^Ty+ c is used to map input space the to high-dimensional feature space. *C* is the regularization parameter, which must be greater than zero. We let the C is 1. $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots., \alpha_N^*)^T$ is the LaGrange multiplier vector. It is the best solution of above.

$$\min\left(\frac{\text{the } 1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K\left(x_{i,} x_j\right) - \sum_{i=1}^{N} \alpha_i\right) (1)$$

$$\text{s.t} \sum_{i=1}^{N} \alpha_i y_i = 0 \ (2)$$

$$0 \le \alpha_i \le C, i = 1,2, \ldots. N$$

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots., \alpha_N^*)^T \ (3)$$

Step 2: Select a positive component from $\alpha^*\left(0 \le \alpha_j^*\right)$to j calculate b^* which is the parameter of objective function:

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i \, K\left(x_i . x_j\right) (4)$$

Step 3: Construct decision function.

$$f(x) = \text{sgin}\left(\sum_{i=1}^{N} \alpha_i^* y_i \, K\left(x_i . x_j\right) + b^*\right) (5)$$

Finally, we get the decision function. The decision function

is used to decide how to forward packets. If the output of the

function is -1, it is a DDoS attack packet, otherwise, it is a

normal packet.

IV. EXPERIMENT

A. Training and Test Dataset

To explain the effectiveness of the DDoS identification method based on SVM, We select the KDD99 dataset as training and test Dataset [9]. It is widely used in academic research, such as IDS and machine learning studies [10].

The KDD99 dataset includes five major categories, which are normal, DoS, Probe, R2L, U2R. It uses 41 features to describe a connection. The statistical analysis of the dataset is shown in table 2.

Where *TP* indicates the true positives, *TN* indicates negatives, *FN* indicates false negatives and *FN* indicates false negatives. The experiment results are shown in table 4. that denotes the effectiveness of our model. We can get the accuracy is 0.998. It can be seen that our model for identifying DDoS attacks has a high recognition rate.

Table 2 KDD99 Data Set

Type	Number
Normal	12056
DoS	46024
Probe	839
R2L	3277
U2R	9
TOTAL	62205

We focus on the HTTP flood attacks in this paper. Thus, the data with DoS and Normal label are selected. And the TCP network connections are used as a dataset.

Table 3 Experiment data division

Datase t	Type of attribute	Total instances	Percent
All	Normal & attacks	(768670+1074241) 1842911	100%
Trainin g	Normal&	(576842+805342)1 382184	75%

	attacks		
Testing	Normal & attacks	(191828+268900)460728	25%

Then we divided the dataset into train dataset and test dataset after feature selection. Table3 describes the datasets in details respectively.

V. RELATED WORK

A. DDoS Defense Based on Machine Learning

Machine learning method based DDoS attack detection are paid much attention. Most frequently used algorithms include Naive Bayes, Decision Tree, K-Nearest Neighbor (KNN) and Support Vector Machine. IAO Fu et al. propose an improved KNN algorithm to classify the attack traffic[11]. However, this method is suitable for offline detection. He Z et al. propose a DDoS attack detection algorithm based on machine learning to prevent attacks on the source side in the cloud[12]. They evaluate nine machine learning algorithms and carefully compare their performance. They found that machine learning methods had a good effect in identifying DDoS attacks. They only did experiments about the effectiveness of the detection algorithm without proposing the way to defend against DDoS attacks.

Ahmed ME et al. propose a method for mitigating DNS Query-Based DDoS attacks based on DPMM(Dirichlet Process Mixture Model) [13]. Although the method has a good effect on mitigating DNS Query-Based DDoS attacks, the miscarriage rate is high.

Chuanhuang Li and Yan Wu et al. propose a DDoS attack detection and defense method based on deep learning, and they apply it to OpenFlow-based SDN[14]. The result shows deep learning is a good method to detect DDoS attack.

B. Research on defense DDoS in SDN based Networks

Alshamrani A et al. propose a defense system for defeating DDoS attacks in SDN based networks[15]. The system unlike most of the existing ML-based approaches, the extensive range of prediction features are used to cover more types of DDoS attacks as well as to ensure better DDoS detection accuracy. However, the extracted features are based on the subset of valid features. The easy accessibility of the features is not actually considered in the SDN environment.

Hong, Kiwon et al. propose an SDN-assisted DDoS attack defense method that can detect and mitigate Slow HTTP DDoS attacks(SHDA), which relies on an SHDA in the SDN controller[16]. They use a proprietary controller so that the portability is poor.

Yang Xu and Yong Liu studied how to utilize SDN to detect DDoS attacks by capturing the flow volume feature as well as the flow rate asymmetry feature[17]. But their methods only consider one factor. We propose a framework to identify and defend DDoS attacks that based on SDN and machine learning for the campus network. This framework can be deployed online to identified the DDoS attack and defense DDoS attack. Our framework enables online real-time detection of DDoS attacks and corresponding defense strategies. This framework design does not depend on other hardware and has good portability.

VI. CONCLUSION

Protecting the operating system of SDN (i.e. the controller) by detecting LR-DDoS attacks is the primary objective of this research. In this paper, we design an SDN framework to identify and defend against DDoS attacks. This framework consists of 2 parts which are traffic collection module, attack identification module and flow table delivery module. Traffic collection module extracts the features to prepare for traffic identification from the data. Currently, we have applied SVM to DDoS traffic identification. The experiment results on the KDD99 dataset show the effectiveness. This classification model is deployed on the simulated SDN environment for campus network as a DDoS detection module. All traffic is identified by this model. If attack traffic is identified, the controller will discard packets according to the predefined rule. If the packet is not attacked, the forwarding policy will be executed normally. In the future, we will optimize that the ratio of convection and single flow are used to judge whether the growth of network traffic is DDoS. And we will improve the flow table delivery module and deploy the model to the SDN environment for the campus network.

VII. FUTURE WORK

One limitation that our method has is the detection of attacks when the entire network is being targeted by DDoS. It might slow down the entire detection process till the network traffic cleared.

Having addressed the detection in one controller network, two more tasks to be done are:
i) Detection of attack in a multi-controller SDN structure.
ii) Mitigation of the attack.

In SDN, networks are connected to controllers and, several controllers might be connected to each other. Detecting an attack in one of them could show the source of the attack and make discovery of the source much easier. This method requires an inter-controller communication that sends the threat alert to all the controllers. Adding this communication process to SDN will be an extension to the current work and a topic for future work. Mitigation of Various in SDN will be next objective for this research. The first step of mitigation will be detection of the source or sources of the attack. Adding more specific statistics collection to the controller will enable it to monitor the packet flow rate at the switch level where attack flows are directed to the controller. Then, more elaborate techniques can be used to pinpoint the malicious hosts. This is a very interesting future work that can be a baseline for any detection scheme in SDN structure.

REFERENCES

[1] Somani, Gaurav, et al. “DDoS Attacks in Cloud Computing.” Computer Communications, vol. 107, 2017,

pp. 30–48.

[2] Key Points of Education Informationization in 2017.[Online].Available:http://www.edu.cn/edu/zheng_ce_gs_gui/zheng_ce_wen_jian/zong_he/201702/t20170221_1491075.shtml

[3] Fayaz S K, Tobioka Y, Sekar V, et al. Bohatei: flexible and elastic DDoS defense[C]// Usenix Conference on Security Symposium. USENIX Association, 2015:817-832.

[4] Yan Q, Yu F R, Gong Q, et al. Software-Defined Networking (SDN) and Distributed Denial of Service (DDoS) Attacks in Cloud Computing Environments: A Survey, Some Research Issues, and Challenges[J]. IEEE Communications Surveys & Tutorials, 2016, 18(1):602-622.

[5] Perera P, Tian Y C, Fidge C, et al. A Comparison of

Supervised Machine Learning Algorithms for Classification of Communications Network Traffic[C]// International Conference on Neural Information Processing. Springer, Cham, 2017:445-454.

[6] Ryu controller.[Online]. Available: http://osrg.github.io/ryu/resources.html

[7] D. Huang, L. Xu, C. Chung T. Xing, "SnortFlow: A openflow-based Intrusion Prevention System in Cloud Environment," Second GENI Research nad Educational Experiment Workshop, pp. 89-92, 2013.

[8] W. Su, L. Wu, Y. Huang, S. Kuo Y. Hu, "Design of Event-Based Intrusion Detection System on OpenFlow Network," in IEEE International Conference on Dependable Systems and Networks (SDN), 2013, pp. 1-2.

[9]KDDCup1999Data.[Online].Available:http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] Özgür, Atilla, and Hamit Erdem. “A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015.” PeerJ, vol. 4, 2016.

[11] IAO Fu,MA Junqing, HUANG Xunsong, WANG,uchuan. DDoS attack detection based on KNN in software defined networks. Journal of Nanjing University of Posts and

Telecommunications, j.cnki.1673-5439.2015.01.013

[12] He Z, Zhang T, Lee R B. Machine Learning Based DDoS

Attack Detection from Source Side in Cloud[C]// IEEE, International Conference on Cyber Security and Cloud Computing. IEEE, 2017:114-120.

[13] Ahmed ME, Kim H, Park M. Mitigating DNS query-based DDoS attacks with machine learning on software-define

d networking[C]// Milcom 2017 - 2017 IEEE Military Communications Conference. IEEE, 2017:11-16.

[14] Li, Chuanhuang, et al. “Detection and Defense of DDoS

Attack-Based on Deep Learning in OpenFlow-Based SDN.” International Journal of Communication Systems, vol. 31, no. 5, 2018.

[15] Alshamrani A, Chowdhary A, Pisharody S, et al. A Defense System for Defeating DDoS Attacks in SDN based Networks[C]// ACM International Symposium on Mobility Management and Wireless Access. ACM, 2017:83-92.

[16] Hong, Kiwon, et al. “SDN-Assisted Slow HTTP DDoS Attack Defense Method.” IEEE Communications Letters, 2017, pp. 1–1.

[17] Xu Y, Liu Y. DDoS attack detection under SDN context [C]// INFOCOM 2016 - the, IEEE International Conference on Computer Communications, IEEE. IEEE, 2016:1-9.