2018 15th International Symposium on Pervasive Systems, Algorithms and Networks

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

YANG Lingfeng
School of Computer Science and Software Engineering
East China Normal University
ShangHai, China
51174500057@stu.ecnu.edu.cn

ZHAO Hui
Shanghai Key Laboratory of Trustworthy Computing
School of Computer Science and Software Engineering
East China Normal University
Shanghai, China
hzhao@sei.ecnu.edu.cn

***Abstract*—SDN (Software Defined Network) has attracted great interests as a new paradigm in the network. Thus, the security of SDN is important. Distributed Denial Service (DDoS) attack has been the plague of the Internet. Now, it is a threat in some SDN applied scenarios, such as the campus network. In order to alleviate the DDoS attack in the campus network, we propose an SDN framework to identify and defend DDoS attacks based on machine learning. This framework consists of 3 parts which are traffic collection module, DDoS attack identification module and flow table delivery module. Traffic collection module extracts traffic characteristics to prepare for traffic identification. Utilizing the flexible and multi-dimensional features of SDN network architecture in deploying DDoS attack detection system, the controller extracts the network traffic characteristics through statistical flow table information and uses the support vector machines (SVM) method to identify the attack traffic. Then the flow table delivery module dynamically adjusts the forwarding policy to resist DDoS attacks according to the traffic identification result. The experiment is conducted using KDD99 dataset. The experiment results show the effectiveness of the DDoS attack identification method.**

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

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 DDoS attack on its services. Amazon EC2 cloud servers were attacked by a massive DDoS attack[1]. Thus, strengthening DDoS attack detection and defense is an 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 network infrastructure is abstracted from the 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 framework consists of 3 parts which are traffic collection module, DDoS attack identification module and flow table delivery module. Traffic collection module extracts traffic characteristics to prepare for traffic identification. 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 main work 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.

The rest of the paper is organized as follows. In Section II, We describe the scenario assumptions and propose the DDoS attack identification and defense framework. In Section III, The process of establishing a DDoS attack detection model is presented, including feature selection and model training. Section IV shows the result of experiments using KDD99 dataset. Section V introduces some of the current research on

2375-527X/18/$31.00 ©2018 IEEE
DOI 10.1109/I-SPAN.2018.00036

DDoS attack defense and compares the differences between our work. Finally, we conclude the research in section VI.

II. DDoS ATTACK *IDENTIFICATION* AND DEFENSE FRAMEWORK

A. Scenario Assumptions

Figure 1 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.

- 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.

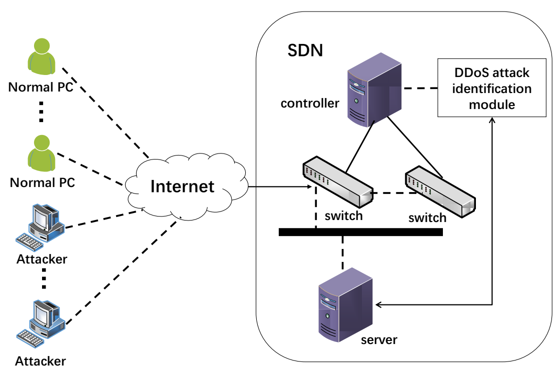

Figure1. The system architecture in the hypothetical scenario

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 showed in figure 2.

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 module, DDoS attack identification module and flow table delivery module. 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.

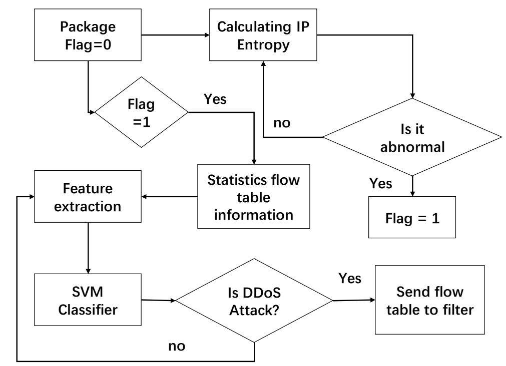

Figure 2. Specific detection process

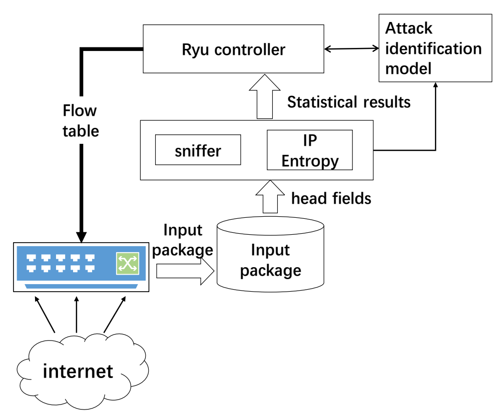

Figure 3. DDoS attack recognition and defense framework

D. Tools used

We use Ryu[6] to build our controller. The reason for choosing it is that Ryu is an open source controller based on Python, which facilitates system integration and has high portability and extensibility. We use the SVM model to establish the DDoS attack detection module. The specific process is described in section III. Generate DDoS attack traffic using the open source tool Ddosflowgen[7]. The web server is simulated with the Damn Vulnerable Web App (DVWA)[8]. DVWA is a PHP/MySQL web application that is damn vulnerable. Its web environment is similar to the school website. In the experimental phase, we use mininet[9] to simulate an SDN network and build an experimental topology.

III. 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. The process of switch processing packets is in figure 4 in detail.

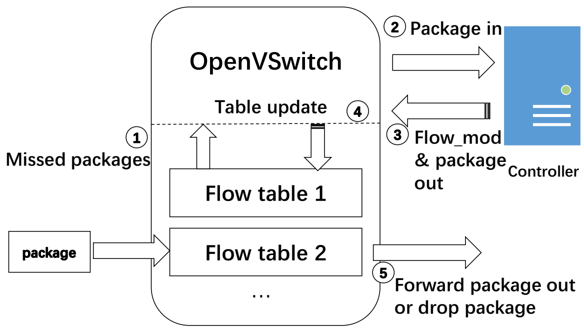

Figure 4. The process of switch processing packets

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_count	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_rerror_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 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 figure5.

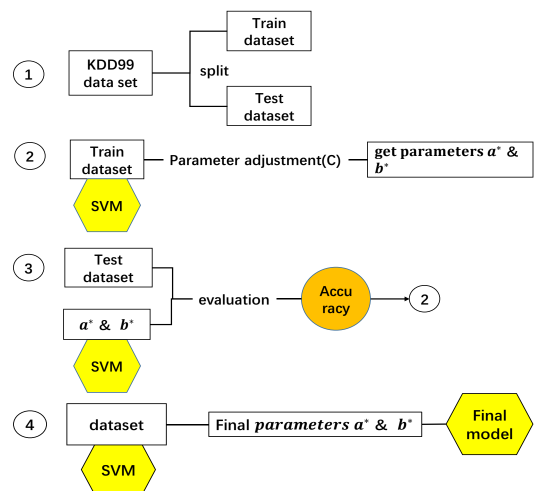

Figure5. The traffic classifier construction process

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 $X=\{x_1,x_2,x_3,..,x_n\}$, and x_i 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.

1) 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 $K(x,y) = x^T y + c$ 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. $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ is the LaGrange multiplier vector. It is the best solution of above.

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

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

$$0 \le \alpha_i \le C, \quad i = 1,2, \dots N$$

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T \quad (3)$$

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

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

3) *Step 3:* Construct decision function.

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

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[10]. It is widely used in academic research, such as IDS and machine learning studies[11].

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.

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.

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

Table 3 Experiment data division

Dataset	Type of attribute	Total instances	Percent
All	Normal & attacks	(768670+1074241)1842911	100%
Training	Normal& attacks	(576842+805342)1382184	75%
Testing	Normal & attacks	(191828+268900)460728	25%

B. Experiment result

All experiments are run on a personal computer which is equipped with a quad-core Intel Core i5-8200U 1.8GHz processor and 8G of memory. The basic code of the SVM classifier is adopted from scikit-learn[12] and is revised to finish our work.

The accuracy is used to evaluate the effectiveness of the DDoS attack classifier.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

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 4. The identification results

Type	The number of
TP	191598
FP	553
FN	230
TN	268347

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[13]. 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[14]. 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) [15]. 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[16]. 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[17]. 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[18]. 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[19]. 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 & FUTURE WORK

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 traffic characteristics to prepare for traffic identification. 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.

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 an d elastic DDoS defense[C]// Usenix Conference on Securi ty 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] Ddosflowgen.[Online].Available: https://github.com/GaloisInc/ddosflowgen

[8] Damn Vulnerable Web App (DVWA). [Online].Available: http://www.dvwa.co.uk/

[9] Mininet.[Online].Available: http://mininet.org/

[10] KDD Cup 1999 Data.[Online].Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[11] Özgür, Atilla, and Hamit Erdem. "A Review of KDD99 D ataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015." PeerJ, vol. 4, 2016.

[12] Scikit-learn.[Online].Available: http://scikit-learn.org/stable

[13] IAO Fu,MA Junqing, HUANG Xunsong, WANG Ruchu an. DDoS attack detection based on KNN in software defi ned networks. Journal of Nanjing University of Posts and Telecommunications, j.cnki.1673-5439.2015.01.013

[14] He Z, Zhang T, Lee R B. Machine Learning Based DDoS Attack Detection from Source Side in Cloud[C]// IEEE, In ternational Conference on Cyber Security and Cloud Com puting. IEEE, 2017:114-120.

[15] Ahmed ME, Kim H, Park M. Mitigating DNS query-base d DDoS attacks with machine learning on software-define d networking[C]// Milcom 2017 - 2017 IEEE Military Co mmunications Conference. IEEE, 2017:11-16.

[16] Li, Chuanhuang, et al. "Detection and Defense of DDoS Attack-Based on Deep Learning in OpenFlow-Based SD N." International Journal of Communication Systems, vol. 31, no. 5, 2018.

[17] Alshamrani A, Chowdhary A, Pisharody S, et al. A Defen se System for Defeating DDoS Attacks in SDN based Net works[C]// ACM International Symposium on Mobility M anagement and Wireless Access. ACM, 2017:83-92.

[18] Hong, Kiwon, et al. "SDN-Assisted Slow HTTP DDoS At tack Defense Method." IEEE Communications Letters, 20 17, pp. 1–1.

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