Research on Detection Method of Abnormal Traffic in SDN

Yabin Xu[1,2(✉)], Chenxiao Cui[2], Ting Xu[2], and Yangyang Li[3]

[1] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing 100101, China
[2] Beijing Information Science and Technology University, Beijing 100101, China
xyb@bistu.edu.cn
[3] China Academy of Electronics and Information Technology, Beijing 100041, China

Abstract. Compared with traditional network, the network architecture and equipment function of SDN have changed dramatically. Thus it is necessary to research more targeted network security strategies. Abnormal traffic detection is the foundation of intrusion detection and intrusion prevention. For this reason, This paper proposes a specific abnormal flow detection method aimed at SDN. The method makes full use of flow-table in SDN switch to extract the features of abnormal flows, and applies information entropy to process non-numerical features of a flow into numerical features. Finally, a BP neural network model previously trained by these numerical features are used for abnormal flows detection. The contrast experiment results show that, this method can detect abnormal traffic in SDN effectively.

Keywords: SDN · Abnormal traffic detection · Entropy · BPANN

1 Introduction

With the arrival of the "Internet+" era, new network applications emerge and make higher demands on the flexibility and convenience of network. Traditional network switch, because of the strong coupling between network control and data transmission, is strictly limit the development of these new network applications. In order to improve the status quo, researchers from Stanford University proposed an OpenFlow protocol [1] in 2008, and gradually extended it as Software Defined Network (SDN).

The core idea of SDN is to decouple network control from data transmission. The control function is provided by SDN controller. The SDN switch only has data transmission function and no control function, so as to simplify the design of switch. Due to the changes of network architecture, network devices and the functions of network device in SDN, the network security problems in SDN should be reconsidered, and a specific solution for SDN is needed. So far, data center network has been one of the main application areas for SDN. In data center network, the abnormal network flows can consume large network resources, making them unable to provide normal network service and even making data center suffer serious data loss. How to detect

X. Sun et al. (Eds.): ICAIS 2019, LNCS 11632, pp. 248–259, 2019.
https://doi.org/10.1007/978-3-030-24274-9_22

abnormal flows in data center network and take action to restrain them has become an urgent problem for network researchers and network managers. Therefore, the detection of abnormal flows in SDN makes sense.

2 Related Work and Our Idea

In traditional network, many detection methods of abnormal traffic are proposed. Huang [2], Zhu [3] and Kong [4] used some methods of machine learning to detect abnormal traffic in network. Cheng [5] defined a network flow abnormal index with the changing rules of new and old IP addresses, and set thresholds to detect DDOS attacks in big data environment. Chang [6] used flow as the basic unit for abnormal detection and a threshold is preset to decide whether the traffic belongs to anomaly. In traditional network, data transmission takes packet as the basic transmission unit. So sampling collection is needed for the method of traffic statistics, which result in extra overhead. In SDN, however, data transmission takes flow as basic transmission unit. So it is suitable for us to take flow as the detection unit whose information can be got directly from flow-table.

Wan [7] proposed an event-based anomaly detection approach which be installed in SDN switches to identify misbehaviors. They used the N-gram model and K-means algorithm to select feature and to extract event sequence, finally trained HMM to identify aberrant behaviors.

The number of successfully matched packets in flow-table of each switch is counted and a threshold is preset to detect abnormal traffic by Zhang [8]. This method is subjective because the threshold needs to be set upon experience and is lack of versatility. It can't dynamically change with the change of traffic.

Braga [9] proposed a detection method for DDoS attacks. The method uses the OpenFlow protocol to collect statistics of each data stream in the flow table and converts this information into a feature vector, finally inputs the data into the self-organizing mapping network. However, the selected data stream characteristics are relatively monotonous, and the distribution of IP addresses and port numbers in the switch is not considered.

Giotis [10] combined OpenFlow protocol with sFlow technology to extract flow-table information and carried out flow detection based on entropy. While this method proposed a static threshold to detect whether the flow is abnormal or not. So some error is inevitable in a dynamically changing network environment.

Several traditional flow detection methods are utilized under SDN architecture by Mehdi in literature [11]. But all these methods are using single traffic feature to classify flows, and they are only aiming at specific anomaly traffic.

Zuo [12] proposed an online traffic anomaly detection method (OpenTAD). The flow table statistic was collected from the controller online and generated the traffic matrix and sample entropy, finally used the PCA to detected the abnormal traffic. While this method also needs take static threshold to detect anomaly flow and cause little deviation for dynamically changing network traffic.

In summary, the research on detection technology of abnormal traffic in SDN architecture is not mature enough. Most existing methods which choose a threshold are

relatively subjective and the detection results can be affected by the static threshold. Besides, existing research is not comprehensive enough in the selection of network traffic characteristics, resulting in a large deviation in the detection results. Faced with these problems, we take advantage of flow-table in SDN to extract versatile flow features, and employ the Artificial Neural Network model to detect abnormal flows.

3 Abnormal Traffic Detection System Design

Routers or switches in traditional network only contain information about the next forwarding node, so the concept of flow cannot be fully utilized to detect abnormal traffic. Under this condition, most of abnormal traffic detection methods utilize NetFlow or sFlow to collect traffic statistics. While in SDN environment, every flow is transmitted according to the flow-tables in OpenFlow switches. So the information in flow-tables of OpenFlow switches can be used to extract the features about flows, and the abnormal traffic in network can be detected directly.

The network architecture of abnormal traffic detection scheme proposed by us is shown in Fig. 1.

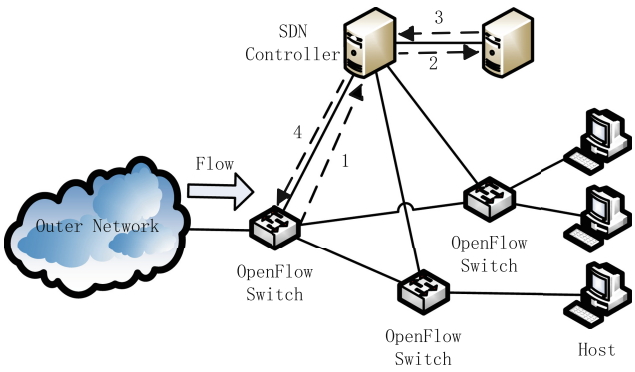

Fig. 1. Network architecture of abnormal traffic detection scheme in SDN

Three-layer network architecture is established in our solution. OpenFlow switches which correspond to the data layer in SDN architecture are in charge of data transmission. SDN controller in control layer collects flow-table information from data layer through the OpenFlow protocol and also provides the fundamental functions, such as topology discovery and flow-table management. Abnormal traffic detection server, which lies in the application layer, detects abnormal traffic by the detection model of abnormal traffic, taking parts of information from flow-tables uploaded by controller as features. The detection results will be sent back to SDN controller which can directly suppress the data forwarding of abnormal traffic.

The specific detection procedure of abnormal traffic (also shown in Fig. 1) is as follows:

(1) SDN controller collects information from each flow-table in OpenFlow switches.
(2) SDN controller transmits the flow-table information to the anomaly flow detection server. The anomaly flow detection server extracts flow features from the flow-table information and uses the machine learning method to detect whether there is abnormal traffic in OpenFlow switches.
(3) Anomaly flow detection server sends the detection results to SDN controller.
(4) SDN controller develops data forwarding control policy according to the detection result and allocates flow-tables to corresponding OpenFlow switch, so as to restrain the entrance of abnormal traffic and forward normal traffic.

4 Abnormal Traffic Features Selection and Processing

4.1 Flow Features Selection

As SDN controller can easily acquire flow-tables from every OpenFlow switches, we can determine whether the flow is abnormal according to the features implicit in the flow-table. Braga R in literature [9] choose a custom 6-tuple as the features of a flow to be detected, which consists of average of packets number per flow, average of bytes per flow, average of duration per flow, percentage of pair-flows, flows growth and ports growth. However, the implied features of the distribution of IP addresses and distribution of port numbers are not considered in this method. In order to describe the differences between normal traffic and abnormal traffic comprehensively, the flow features are described as a 8-tuple in this paper, which composed of number of packets per flow, average bytes per packet, duration time per flow, protocol type, source IP address, destination IP address, source port number and destination port number.

(1) Number of packets and average bytes per packet in one flow: Common abnormal traffic usually contains small amounts of packet. For example, about 3 packets constitute a flow during the DDoS attacks [13]. While normal traffic tends to transmit a large number of packets to complete the data communication task. Therefore, the number of packets of each flow can represent as one feature of abnormal traffic. Besides, normal traffic often carries large number of valuable data, so normal traffic contains a lot of bytes in each packet. On the contrary, abnormal traffic aim at sniffer or attacking, using only several bytes without real content, even its value is determined. Hence the average of bytes per packet can be chosen as another feature of the abnormal traffic and it can be calculated by formula (1):

$$\text{Average bytes in packet(Bytes)} = \frac{\text{number of transmitted bytes}}{\text{number of successfully matched packets}} \quad (1)$$

(2) Duration time and protocol type of a flow: In the data center network architecture of SDN, the features of a flow are obvious. Jouet [14], Noormohammadpour [15], Sasaki [16] showed in their researches most flows have shorter lifetimes and TCP protocol is widely utilized in data center network. On the contrast, anomaly flows

last longer time to keep generating harmful effects to network and use different protocols. For example, the Death of Ping and Worm Welchia use ICMP protocol, while DDoS attack usually use TCP protocol. In consideration of the flow features in data center network under the SDN architecture, duration time and protocol type of flow are used as parts of the features of a flow in this paper.

(3) Distribution of IP addresses and port numbers of flows: the distribution of IP addresses and port numbers are the important features to understand flows in network. In data center network, the distribution of IP addresses and port numbers of normal traffic are scattered, but abnormal traffic tends to focus traffic on one or more specific targets. For instance, DDoS attack will infect multiple puppet machines and launch concentrated attacks on a specific target. During an anomaly attack, several flow-table entries which consist of different source IP addresses and same destination IP address will emerge in a flow-table. Thus, the distribution of destination IP addresses tends to be centralized and the distribution of source IP address becomes more scattered.

4.2 Quantization Processing on Nonnumeric Flow Features

Since the distribution of IP addresses and port numbers of a flow cannot be described by numeric data directly and take part in the calculation, the theory of entropy is employed in this paper to process the distribution of IP addresses and port numbers into numeric data.

A flow F can be denoted as $\{F = A_{srcip}, A_{srcport}, A_{dstip...}\}$. The entropy of some features in a flow F is defined as $H(X) = -\sum_{i=1}^{n} P_i(x_i) log_2 P_i(x_i)$. Among them, $\mathrm{P_i(x_i)}$ indicates the happening probability of event x_i.

For example, to calculate the entropy of source IP address, the formula is $H(SrcIP) = -\sum_{i=1}^{N} P_i(SrcIP) log_2 P_i(SrcIP)$. N denotes the total flow number of different source IP addresses, $\mathrm{P_i(SrcIP)}$ denotes the ratio of the number of flows which contains ith source IP address to the number of total flows. It can be expressed as

$$\mathrm{P_i(SrcIP)} = \frac{\text{dataflow number that contains SrcIP}}{\text{total dataflow number}} \tag{2}$$

Because the scale of dataset will affect the calculation of entropy, we normalize the entropy by dividing it with the maximum entropy value of the dataset, so the entropy can be described as:

$$\mathrm{H(SrcIP)} = -\frac{\sum_{i=1}^{N} P_i(SrcIP) log_2 P_i(SrcIP)}{log_2 N} \tag{3}$$

So, the range of the entropy value can be normalized to the interval of (0, 1). For the destination IP address, the source port number and the destination port number can be obtained by the same calculation method for each entropy value.

By employing the concept of information entropy, the nonnumeric flow feature can be transmitted into numeric type. So the distribution of IP addresses and port numbers

can be directly expressed and easily be calculated in abnormal traffic detection algorithm.

4.3 Quantization Processing on Nonnumeric Flow Features

In the process of selecting the features of abnormal flows, there may be the situation of which multiple features are related. For example, if feature A and feature B are related each other, then the feature vector of flows will make its importance strengthening in specific aspects because of the correlation of features A and B. Thus, it will weaken the importance of other features in the feature vector. Therefore, we must carry out consistency test before determining the feature vector of flows.

In order to improve the rationality and fairness of the model calculation, we utilize the feature redundant coefficient proposed by Wang [17] to estimate the redundancy among the features. The feature redundant coefficient can be calculated as:

$$t_{AB} = min\left(abs\left(\frac{\overline{A_S} - \overline{B_S}}{\overline{A_S} + \overline{B_S}} \right), abs\left(\overline{A_S} - \overline{B_S} \right) \right) \tag{4}$$

$\overline{A_S}$, $\overline{B_S}$ respectively denotes the average entropy per second of feature A and feature B in a period of time. The more close to 1 t_{AB} is, the more irrelevant of feature A and feature B is. Otherwise, the two features are more relevant. In this paper, we set the threshold as 0.1. If $t_{AB} < 0.1$, these two features are considered as redundant. Then, one feature with smaller entropy should be deleted in this condition. For our selected features, the consistency test results indicate that the consistency of all the features does not exist.

5 Abnormal Traffic Detection in SDN

At present, BP Artificial Neural Network (BPANN) and Support Vector Machine (SVM) are two main machine learning models for abnormal traffic detection. Contrast experiments are needed in order to choose a better model.

SVM is a kind of machine learning method that based on the principle of structural risk minimization. It utilizes dataset to train the classification model, which maps the feature vector into high demission space and determines one maximum margin hyper plane to identify the classification of data. The bigger the margin is, the more accurate the results of classification are.

Supposing that the training dataset can be expressed as $(y_1, x_1), (y_2, x_2), \ldots, (y_i, x_i), \ldots, (y_l, x_l)$. Among them, $y_i = \{0, 1\}^l$ denotes the class label, 0 is the normal class and 1 is the abnormal class. $x_i \in R^n, i = 1, \ldots, l$, which denotes an n-dimensional feature vector. In the feature space, the linear equation of $\omega \cdot x + b = 0$ is used to make the margin distance maximum between hyper plane and the two classes. In this equation, ω presents the weight of vector and b is defined as an offset. The process of seeking the optimal separating hyper plane is the process of machine learning and the core problem is to solve the minimal solution of formula (5):

$$f(x) = sgn[(\omega \cdot x + b)] = sgn\left[\sum_{i=1}^{L}\left(\alpha_i^* y_i (x \cdot x_i) + b^*\right)\right] \tag{5}$$

In this formula α_i^*, b^* are parameters of this optimal separating hyperplane. Different core function can be employed in formula (5) to construct different SVM model. Common used core functions are polynomial function, RBF core function and sigmoid core function. In this paper, we employ the RBF core function for its excellent overall performance.

BPANN is one kind of the artificial neural networks and has been applied in many fields. It is composed of multiple layer interconnected neuron. The neural model can be expressed as:

$$u_k = \sum_{i=1}^{n} w_{ik} x_i \tag{6}$$

$$y_k = f(u_k + b_k) \tag{7}$$

$x_i (i = 1, \ldots, n)$ is the input vector and $w_{ik} (i = 1, \ldots, n)$ is the weight of neuron k. The number of input vector is denoted by n. u_k is the linear combination output of input vector. b_k is denoted as the threshold value of a neuron. $f(x)$ is the activation function and y_k is the output of neuron.

The biggest advantage of BP Neural Network is that can adjust the artificial network model gradually so as to optimize the classification result through its self-learning ability by instant feedback process. An 8-tuple vector is chosen as the flow features in this paper and the output result is defined as flow classification. The corresponding BPANN construction is shown in Fig. 2.

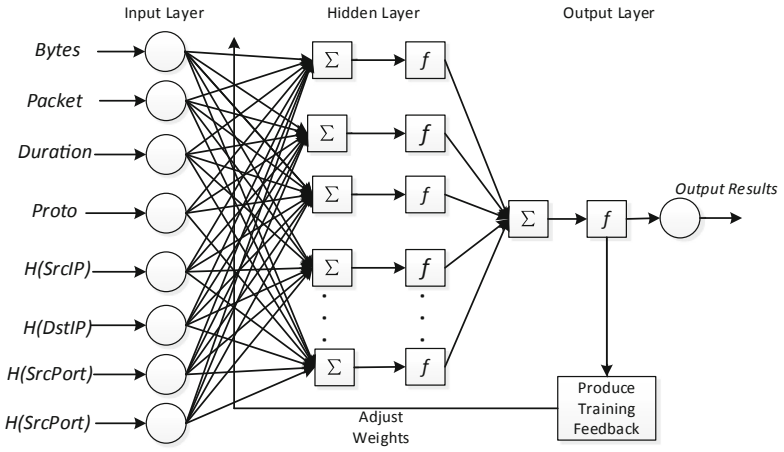

Fig. 2. BPANN structure for anomaly flow detection

6 Experiment and Evaluation

In order to evaluate the detection method of abnormal traffic in this paper, BPANN and SVM methods are used to detect the anomaly flows in dataset and the effects are compared. The dataset which we used is DARPA evaluating dataset from Lincoln Laboratory [18]. It is currently the most comprehensive attack test dataset at present. The DARPA dataset contains the simulating data from 5 week. The data of the first two weeks are provided as training data, and the data of the last two weeks are used as test data. The data from the first and the third week is normal flows without any attack. While several attack data are involved in the data from the second week.

In this paper, we use the precision (rate), recall (rate) and F-Measure as the evaluation criterions in contrast experiments. Formula (8) and (9) depict the formula to calculate precision rate and recall rate.

$$precision = \frac{tp}{tp + fp} \tag{8}$$

$$recall = \frac{tp}{tp + fn} \tag{9}$$

Among those formulas, *tp* denotes the number of attack flows that labeled as attack, fp is the legitimate flows that classified as attack, and *fn* denotes the attack flows that classified as legitimate flows. Because the restrictive relation exists between precision rate and recall rate, the detection effect of anomaly flows cannot be fully reflected by just using these two criterions. Therefore, F-measure is selected as a comprehensive evaluation criterion. F-measure can be calculated as:

$$F - measure = \frac{precision \times recall \times 2}{precision + recall} \tag{10}$$

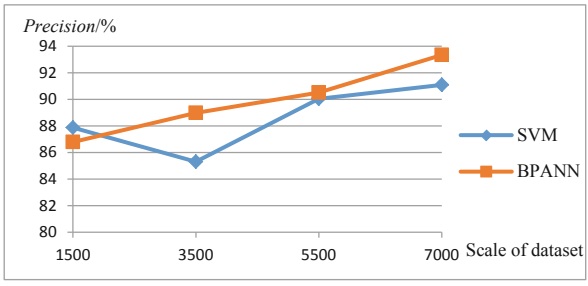

Fig. 3. Precision rate comparison between SVM and BPANN

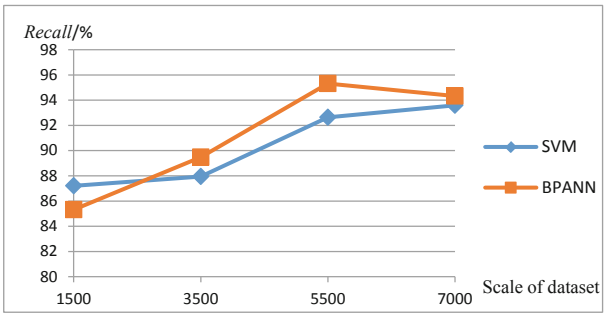

Fig. 4. Recall rate comparison between SVM and BPANN

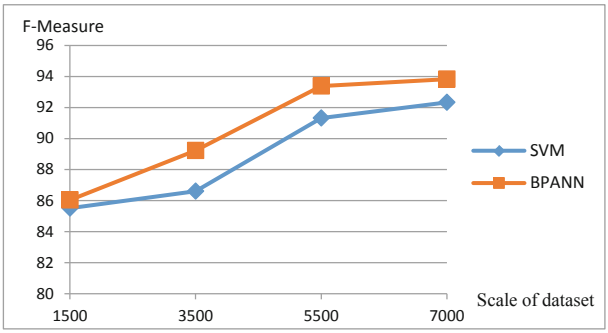

Fig. 5. F-measure comparison between SVM and BPANN

Contrast experimental results about the two kinds of machine learning methods are shown in Figs. 3, 4 and 5.

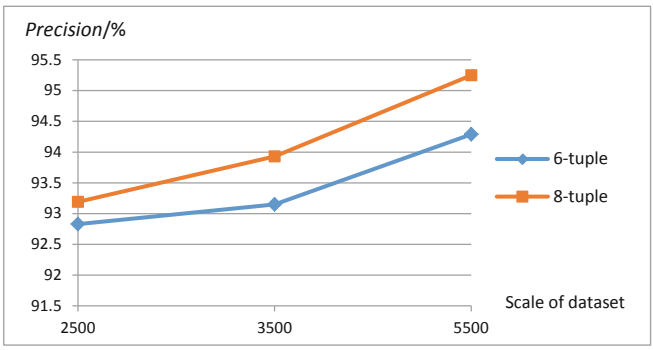

Fig. 6. Precision rate comparison between two methods

Figures 3, 4 and 5 show the comparison results of precision, recall and F-measure under the different scale of dataset. From the comparison figures we can see that, the detection model grow more rational and the performances of two model keep improving along with the expansion of the scale of dataset. Generally speaking, the precision rate of BPANN is higher than that of SVM. With the expansion of the scale of dataset, detection effect of BPANN gains a steady improvement. In recall rate, BPANN fluctuates in a relative acceptable range. The comparison results of F-measure between these two methods are shown in Fig. 5. As the size of dataset increases, the value of F-measure of both methods improves. But, in general, the value of F-measure of BPANN method is higher than that of SVM, which indicates that BPANN detection method perform better than SVM detection method. So BPANN model is selected as the detection method of abnormal traffic.

Since the detection methods of abnormal traffic in SDN are mostly aimed at DDoS attack at present, we take the BPANN method to detect DDoS attacks and compare the detection results with SOM method proposed by Braga R [9]. A DDoS attack generator called LOIC (LOIC) [19] is utilized in this contrast experiments. LOIC simulates the DDoS attack and generates the DDoS flows. During the whole experiment, 500 DDoS attack flows mixed with 5000 normal flows are evenly distributed. The contrast experiment results about precision rate, recall rate and F-measure are shown in Figs. 6, 7 and 8. In these figures, 6-tuple and 8-tuple respectively represents the method proposed by Braga R [9] and in this paper.

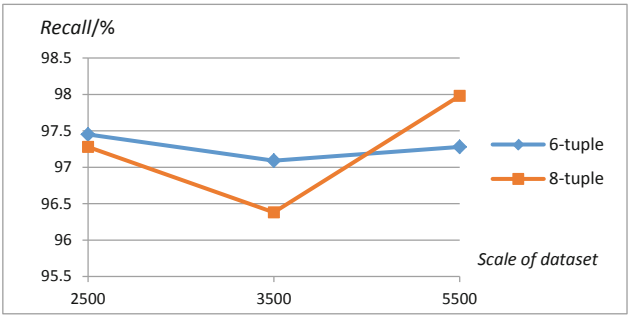

Fig. 7. Recall rate comparison between two methods

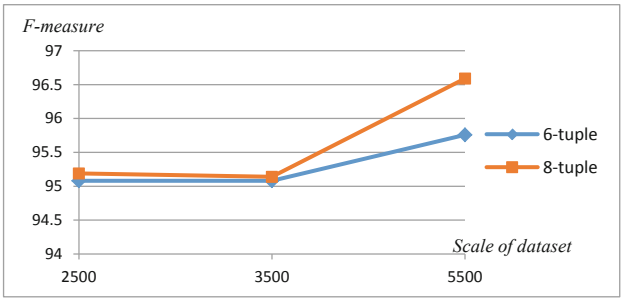

Fig. 8. F-measure comparison between two methods

From these figures it can be seen that precision rate and F-measure of these two methods are both improve along with the expansion of the scale of data flows. Although recall rate declines in the beginning, it shows a rapid upward trend when experimental data flows increase to a certain extent. Overall, every kinds of performance from the detection method based on 8-tuple proposed in this paper are much better than the detection method based on 6-tuple proposed by Braga R [9].

7 Conclusion

Taking advantage of forwarding data based on flows in SDN and obtaining information from flow-table in OpenFlow switches, this paper proposes a detection method of abnormal traffic based on a 8-tuple of flow features in SDN and BPANN. This method extracts data flow features from OpenFlow switches and uses BP neural network classification method to detect anomalies. As for the selection of the flow features, we utilize the 6-tuple features proposed by Braga R [9], and add the distribution of network IP addresses and port numbers. And we also use the information entropy theory to reflect the network traffic distribution in SDN with numerical data, so as to form more comprehensive and rational data flow characteristics. This paper uses the DARPA data set as the test data for the simulation experiment. The results of experiments shows, the detection model and features of abnormal flows proposed and used in this paper can effectively detect anomaly flows in SDN.

Acknowledgement. This work was supported by the National Natural Science Foundation of China Nos. 61672101, the Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (ICDDXN004)* and Key Lab of Information Network Security, Ministry of Public Security, No. C18601.

References

1. McKeown, N., Anderson, T., Balakrishnan, H., et al.: OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69–74 (2008)
2. Huang, H., Deng, H., Chen, J., et al.: Automatic multi-task learning system for abnormal network traffic detection. Int. J. Emerg. Technol. Learn. (iJET) **13**(4), 4–20 (2018)
3. Zhu, M.-J., Guo, N.-W.: Abnormal network traffic detection based on semi-supervised machine learning. DEStech Trans. Eng. Technol. Res. (2017). (ecame)
4. Kong, L., Huang, G., Wu, K.: Identification of abnormal network traffic using support vector machine. In: International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 288–292. IEEE Computer Society (2017)
5. Cheng, R., Xu, R., Tang, X., Sheng, V.S., Cai, C.: An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. CMC: Comput. Mater. Continua **55**(1), 095–119 (2018)
6. Chang, S., Qiu, X., Gao, Z., et al.: A flow-based anomaly detection method using sketch and combinations of traffic features. In: International Conference on Network and Service Management, pp. 302–305. IEEE (2011)

7. Wan, M., Yao, J., Jing, Y., Jin, X.: Event-based anomaly detection for non-public industrial communication protocols in SDN-based control systems. CMC: Comput. Mater. Continua **55**(3), 447–463 (2018)
8. Zhang, Y.: An adaptive flow counting method for anomaly detection in SDN. In: Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies, pp. 25–30. ACM (2013)
9. Braga, R., Mota, E., Passito, A.: Lightweight DDoS flooding attack detection using NOX/OpenFlow. In: IEEE Local Computer Network Conference, pp. 408–415. IEEE Computer Society (2010)
10. Giotis, K., Argyropoulos, C., Androulidakis, G., et al.: Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. Comput. Netw. **62**(5), 122–136 (2014)
11. Mehdi, S.A., Khalid, J., Khayam, S.A.: Revisiting traffic anomaly detection using software defined networking. In: Sommer, R., Balzarotti, D., Maier, G. (eds.) RAID 2011. LNCS, vol. 6961, pp. 161–180. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23644-0_9
12. Zuo, Q., Chen, M., Wang, X., et al.: Online traffic anomaly detection method for SDN. J. Xidian Univ. (Nat. Sci.) **42**(1), 155–160 (2015). (in Chinese)
13. Chi, S., Zhou, S.: Research on defend against DDoS attacks. Netinfo Secur. (5), 27–31 (2012). (in Chinese)
14. Jouet, S., Perkins, C., Pezaros, D.: OTCP: SDN-managed congestion control for data center networks. In: Network Operations and Management Symposium, pp. 171–179. IEEE (2016)
15. Noormohammadpour, M., Raghavendra, C.S.: Datacenter traffic control: understanding techniques and trade-offs. IEEE Commun. Surv. Tutor. **20**(2), 1492–1525 (2017)
16. Sasaki, T., Pappas, C., Lee, T., et al.: SDNsec: forwarding accountability for the SDN data plane. In: International Conference on Computer Communication and Networks, pp. 1–10. IEEE (2016)
17. Wang, X., Shang, Z., Chen, L.: Feature selection algorithm toward abnormal traffic detection. Comput. Eng. Appl. **46**(28), 125–127 (2010). (in Chinese)
18. DARPA Intrusion Detection Data Sets. http://www.ll.mit.edu/ideval/data/index.html
19. LOIC: Low Orbit Ion Cannon. http://sourceforge.net/projects/loic/