Acta Technica *62* No. 1B/2017, 489–498

Research on application of DDos attack detection technology based on software defined network[1]

Guo Jing[2]

Abstract. Distributed denial of service (DDoS) attack is one of the main threats to network security. Although the principle and mechanism of DDoS attack has been widely understood and researched, the detection and prevention of the attack behavior is still very arduous, because this form of attack has high degree of concealment and the dynamic distributivity. In this paper, the characteristics of software defined network and the characteristics of DDoS attacks are analyzed and researched. An improved DDoS attack detection method is proposed, which is different from the traditional method that only detects the single link and network for the victims. Based on the software defined network (SDN), classification learning algorithm is used to detect DDoS attacks online through the construction of an efficient global network flow table feature values, and combined with the feature selection algorithm that can optimize the flow table feature sequence. The experimental results show that for the DARPA 99 and CAIDA DDoS 2007 two typical real data sets, the method can improve the detection rate of the DDoS attacks and significantly reduce the false alarm rate, and it has good comprehensive performance.

Key words. Software defined network, DDoS attack detection, feature selection algorithm.

1. Introduction

The Internet has become an important information infrastructure in modern society. The rapid popularization and wide application of the Internet have profoundly changed the way of human life, and the Internet has become an indispensable part of daily affairs. A short service interruption may jeopardize people's normal work and life, the financial order, economic stability, and even national security, bringing immeasurable loss. Therefore, it is of great significance to ensure the continued applications of network for the maintenance of network security. Distributed denial of service attack (DDoS) [1] sends a large number of data packets to the target host by

[1]The authors acknowledge the Special research project of Shaanxi Provincial Department of Education(17JK2038).

[2]School of Information & Engineering, Shaanxi Institute of international Trade & Commerce, Xi'an, 712046, China

multiple hosts, making the target host resources unable to provide normal services because of excessive consumption. DDoS shows the characteristics of strong attack intensity, strong source of attack, attack range and strong concealment on attack means, characteristics and effects [2]. Compared with other attacks, DDoS attack is simple, the destructiveness is strong, and it is difficult to detect and defend.

2. Related technology

2.1. OpenFlow protocol

The OpenFlow protocol is the communication interface standard between the controller and the switch controller [3]. The configuration and management of controller to switch is through the regulation message types of protocols, including symmetric message, the controller-to-switch and asynchronous message, and each kind of news has many types of sub-news. Among them, the symmetric message can be initiated by either the switch or controller, and is used to establish or maintain the connection between controller and switch; the controller-to-switch messages is initiated by the controller, and is used to acquire or manage the state of the switch for the controller. The asynchronous message is initiated by the switch, and is used for the exchange to notify the state changes or network event to the controller.

According to the OpenFlow protocol, the flow table in the switch is the basis for transmitting of the data packets. Each of which consists of a plurality of flow table entries [4]. The flow table represents the data forwarding rules, including the matching domain, operation, counters. The matching domain is used for matching packets, so as to use rich rules in forwarding a packet, including most of the key identifications of the link layer, network layer and transport layer. Each address identification can be a certain value or any value, but also can be used to achieve a more accurate matching. The operation shows the actions to the data packet that the matching is successful, such as forwarding to a port, packet loss or modify the packet header information. For the data package that does not match the flow table, the switch will encapsulate it to sent to the controller through a secure channel, and the controller will decide what kind of operations on it. The counter is used for the the statistical of the basic information of data stream, including the number of packets and the number of bits that match the flow table.

2.2. DDoS detection technology

According to the DDoS attack detection strategy, the detection technology can be divided into two kinds: anomaly detection and misuse detection. Misuse detection can match the known attack characteristics with the collected and observed user behavior characteristics [5]. The system can quickly judge the attack behavior. At present, the most representative misuse detection methods include expert system, pattern matching, state transition analysis and so on. Different from other network attacks, the content of the DDoS attack is legal without any vicious data code, so it is difficult to extract and detect the inherent characteristics of the attacks.

Therefore, the detection efficiency based on misuse DDoS attack detection method is limited, and the missing report rate is higher. Anomaly detection method is based on the establishment of the normal behavior model of the target user and the system to compare the deviation between the measured behavior characteristics and the normal behavior model to identify the attack [6]. This method not only can detect DDoS attacks effectively, but also can detect new attacks that is similar to attack features. This paper mainly studies the anomaly detection for DDoS

3. Attack detection

In this chapter, we design and implement a DDoS attack detection method based on the characteristics of flow table. This method is applicable to software defined network environment [7]. Through the analysis of the characteristics of DDoS and Open flow protocol, combined with mutual information to select the optimal flow surface features, the classification algorithm of attack detection is used to realize the comprehensive and effective attack behavior judgment.

3.1. Attack detection method based on the characteristics of flow table

The basic idea of attack detection method based on flow table feature is that the flow table information in the software defined network switching equipment is extracted to convert into feature vectors, and the optimal feature is extracted to build the attack detection classifier [8], and finally the new network flow is classified. In this section, we introduce an attack detection method which can be applied to software defined network from the flow table feature matrix and feature selection algorithm.

3.2. OpenFlow flow table features

The characteristics can be used to characterize some known attacks, is a description for the attack behavior [9]. For each kind of network attack, we should extract the characteristics of the attack, in the ideal state, we should always be able to detect and identify malicious attacks through these features. The feature vector of the flow table is a high-level abstraction of the attack and normal data, which is the basis for the identification of the attack.

4. Experiment and analysis

4.1. Experimental environment

In this paper, DARPA 99 is used as normal data set Trace normal, CAIDA DDoS2007 data set is used as a base of the generation of abnormal flow Tracel, and the attack in DARPA 99 data set ia taken as Trace2 abnormal flow, the ab-

normal flow of Tracel and Trace2 are mixed as the attack data sets Trace DARPA 99 is attack attack. DARPA 99 is used for the laboratory evaluation of intrusion detection system by MIT Lincoln laboratory and the U.S. Air Force Research Laboratory, and each data of the data set has recorded the detailed information of data packet. CAIDADDoS 2007 contains approximately one-hour DDoS attack flow data, in which the flow data contains only the attack flow rate and the response of the destination host, which is stored in pcap format.

Firstly, the normal data set and the attack data set are sampled periodically, and the sampling periods are recorded as t, and T, respectively. At the same time, the flow set is converted to a set of flow table items. The average packet number, average number of bits, flow table rate, single stream item rate, request flow ratio, source address entropy, source port entropy, and destination port entropy 8 kinds of attributes are obtained to generate the experimental samples and the signature sequences are signed after the flow table feature of the generated flow table item set are counted. The characteristic attributes are 8 categories, which are labeled as 1–8, and the classification attributes are classified into 2 categories, namely: 1 and -1.

KNN, SOM and SVM algorithms are used to learn and test data sets. The analysis tool used by SVM classifier is the LIBSVM software package. KNN and SOM use the analysis tools in Matlab. In this paper, the detection rate, false alarm rate and the total error rate these 3 evaluation indicators are used to assess the effectiveness of the test, which are expressed as DR, FR, ER, respectively, as shown in the following formulae:

$$DR = \frac{TN}{TN + FN}\,, \tag{1}$$

$$FR = \frac{FP}{TP + FP}\,, \tag{2}$$

$$ER = \frac{FN + FP}{TP + FP + TN + FN}\,. \tag{3}$$

Here, TN is the number of attack samples labeled in the attack samples to be tested, FN is the normal samples number labeled in the attack samples to be tested, TP is the normal samples number labeled in the normal samples to be tested, FP is the number of attack samples labeled in the normal samples to be tested.

5. Experimental results

5.1. Feature selection experiment

5.1.1. Influence of selected feature sets on different classifiers. In this paper, we propose a feature selection method based on mutual information, which can stop the search of the feature subset after getting the characteristic attributes of the setting number, thus generating 8 different feature subsets. KNN, SOM and SVM three kinds of classification algorithm can test the different obtained characteristics, in which the SVM function (RBF) is the kernel function. The classification effect

test includes detection rate, false alarm rate, and index. The distribution of feature selection sample set is shown in Table 1.

Table 1. The number distribution of feature selection samples

		Training set	Test set
Normal sample Trace normal		500	2000
Attack samples Trace attack	Trace1 attack	200	1400
	Trace2 attack	200	400

Figures 1 and 2 show the detection effect of classifier on k ($1 < k < 8$) kinds of sets, which are detection rate, false alarm rate and error rate. When $k \geq 4$, the classification detection rate of SVM algorithm is higher than the KNN and SOM algorithm. Only when the feature number is 3, the error rate of SVM algorithm is lower than the SOM algorithm, and in other cases, it is higher than the other two algorithms. When the classification error of SVM algorithm K is 4,5, it is lower than the other two algorithms. Through the analysis of test results of the classifier, it can be seen that the SVM algorithm is better than KNN. SOM algorithm in the detection rate, but the false alarm rate and error rate are relatively low in some cases, but higher than the other two algorithms in other cases.

The KNN algorithm achieves the highest detection rate when k=6, and has the lowest false alarm rate and error rate. When $k = 3$ and $k = 4$, the SOM algorithm and the SVM algorithm achieve good detection results, respectively. The characteristic attributes and the time consumption corresponding to the three kinds of classification algorithms are shown in Table 1. The characteristic attributes have 8 categories, respectively, labeled as a 1–8.

It can be seen from Table 1, the optimal characteristics number of the SOM algorithm is the least, but the training time is the most. On the other hand, the KNN algorithm has the largest number of characteristics and the shortest classification time. The SVM algorithm is between the two, and its classification time is far lower than that of SOM algorithm, but by about 0.0156 seconds higher than KNN algorithm, so it has better classification efficiency.

This paper is mainly based on feature selection algorithm, the classifier of the classification efficiency of the optimal characteristics sample on DDoS attack detection effect in the number of different characteristics options is considered. The selected feature number $k = 4$ and SVM algorithm are taken as the classification algorithm in subsequent tests, which has a higher detection rate and lower error rate.

5.1.2. Distribution of the selected feature attributes. Table 2 shows that when $k = 4$, the selected feature categories of the SVM algorithm are 3, 4, 6.8, respectively representing the feature types: flows rate, single flow rate, source address entropy and destination port entropy, which can be expressed as FR, SFR, H (sip) and H (deport). In order to better reflect the distribution of the characteristics value, each column of characteristics in the sample set are normalized, at the same time, the

feature sequences after standardized have not units.

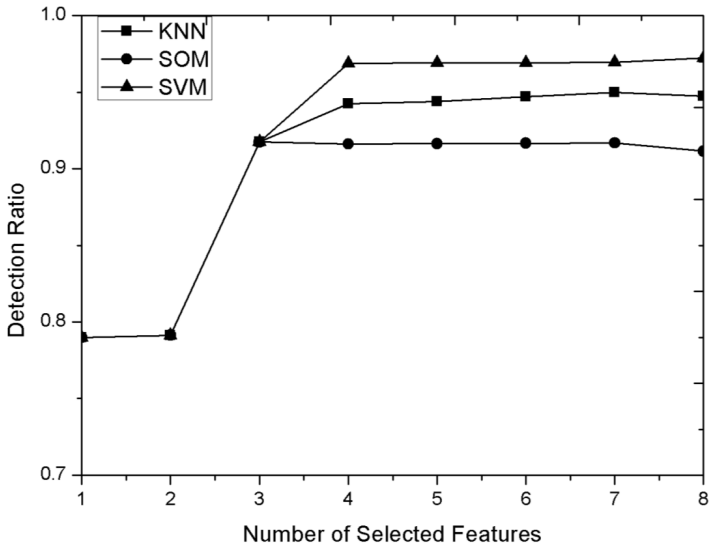

Fig. 1. Detection effect of classifier on K subset–Detection Ratio

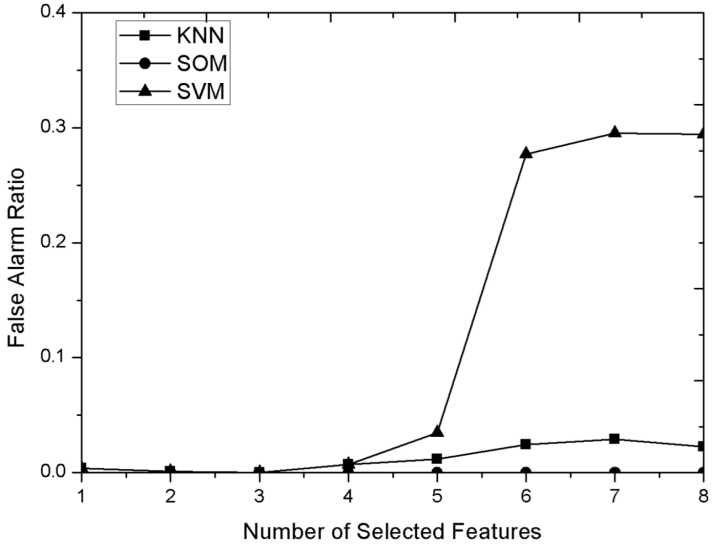

Fig. 2. Detection effect of classifier on K subset–False Alarm Ratio

These two kinds of feature samples have a clear interface when the sampling frequency of the attack characteristics is 100 times of the normal characteristics. It can be explained that the feature attribute of the optimal feature subset obtained according to the characteristics selected algorithm can better reflect the difference

between the normal flow and attack flow, and enhance the ability to recognize attack.

Table 2. Classification efficiency of best feature selection samples

Classification algorithm	Characteristic attribute	Classification time (seconds)
KNN ($k = 6$)	2, 3, 4, 5, 6, 8	0.0468
SOM ($k = 3$)	4, 6, 8	38.8910
SVM ($k = 4$)	3, 4, 6, 8	0.0624

5.2. Attack detection experiment

According to the feature selected experiment, we can know that the classification algorithm of attack detection is the support vector machine, and the optimal flow table number is $k = 4$. First of all, the obtained attack data set is converted to the flow table set with normal data set, and the flow table set is sampled periodically, while the statistics of its flow table characteristics is completed, including flows rate, single flow rate, source address entropy and destination port entropy. The number distribution of characteristic samples in this experiment is shown in Table 3.

Table 3. Classification performance of SVM sample data with different kernel functions

	Kernel functions	Detection rate
Liner kernel function	$K(x_i, y_j) = x_i^{\mathrm{T}} x_j$	99.76 %
Polynomial kernel function	$K(x_i, y_j) = (y x_i^{\mathrm{T}} xj + b)^d$ $y = 0.001, b = 0, d = 3$	99.73 %
RBFkernel function	$K(x_i, y_j) = \exp(-y \parallel x_i - x_j \parallel^2)$ $y = 0.001$	93.76 %
Sigmoid kernel function	$K(x_i, x_j) = \tanh(y \mathrm{x}_i^{\mathrm{T}} x_j + \mathrm{b})$ $y = 0.1, b = 2.1$	92.49 %

After determining the normal and attack samples, the detection rate of the classifier is taken as the evaluation index, and the kernel function of the SVM is determined to be $K(x_i, y_j) = h(x_i) \cdot h(x_j)$, realizing the mapping from input space and feature space. As shown in table 3-S, the kernel function with higher detection rate is selected for further analysis by comparing the classification performance of SVM classifier with different kernel functions on sample data.

By comparing the classification results of four kinds of kernel functions in table 4, we finally choose the linear kernel function with higher detection rate as the $k(x_i, y_j)$ of SVM, and then proceed to the next classification experiment. The selected flow table feature training set is used as the input data of the classifier, and the classification model is constructed. The test set is used to test the classification model.

Table 4. Classification effect before and after feature selection

	Detection results		Classification time (s)
The flow table feature after selection ($k = 4$)	DR	99.76 %	0.0368
	FR	0.3 %	
	ER	0.27 %	
The flow table feature after selection ($k = 8$)	DR	95.56 %	0.0368
	FR	0.6 %	
	ER	2.54 %	

Table 4 is the comparison of the classification performance of the SVM classifier to the feature set after selected and the feature set before selected. It can be seen that the detection rate of the former is higher than that of the latter, and has lower false alarm rate and error rate. At the same time, the input test feature sequence has different dimensions, which makes the flow table after selected feature set have faster classification efficiency.

5.3. Comparative analysis

The detection rate, the false alarm rate and total error rate are selected as the evaluation index, and the detection after selected feature is taken as the contrast method, as shown in Table 5.

It can be seen in table 5 that compared with the traditional network attack detection method, the detection algorithm in this paper has higher detection rate and lower error rate. Compared with OpenTAD method of software defined network, this method has high detection rate and low false alarm rate. It can be concluded that the attack detection method based on the flow table features in this paper has good comprehensive performance, and can effectively identify DDos attacks.

6. Conclusion

This paper mainly introduces the method and experiment of DDos attack detection based on the characteristics of flow table. First of all, the characteristics of OpenFlow flow table and attack flow are analyzed, and the characteristic matrix of flow table for all communication IP is constructed. Secondly, the correlation between the self and the categories and the features is considered, the feature selection algorithm is designed based on the comprehensive correlation which is taken as the evaluation function of feature selection. Next, the DDos attack detection method and its implementation based on the flow table characteristics is described. The feature selection algorithm is used to preprocess the flow table features, and the optimal characteristic attribute is chosen to build the classification model on the basis of the treated samples so as to identify the attack. Finally, the realization of attack detection method is based on the Matlab, and the experimental results of this

method are verified and analyzed. The data shows: the classification model after the feature selection has a good classification effect and shorter classification time. At the same time, compared with the previous attack detection method, it has good comprehensive performance.

Table 5. Detection results of different algorithm in SDN networks and traditional networks on DDoS attacks

	Detection results	
Software defined network (the flow table feature after selection, $k = 4$)	DR	99.76 %
	FR	0.3 %
	ER	0.27 %
Software defined network (Open TAD)	DR	91.7 %
	FR	4.2 %
	ER	...
Traditional network (TFCE)	DR	97.4 %
	FR	0.1 %
	ER	1.4 %

References

[1] Q. Yan, F. R. Yu, Q. Gong, J. Li: *Software-defined networking (SDN) and distributed ddenial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges.* IEEE Communications Surveys & Tutorials *18* (2016), No. 1, 602–622.

[2] S. Scott-Hayward, S. Natarajan, S. Sezer: *A survey of security in software defined networks.* IEEE Communications Surveys & Tutorials *18* (2016), No. 1, 623–654.

[3] K. Giotis, G. Androulidakis, V. Maglaris: *A scalable anomaly detection and mitigation architecture for legacy networks via an openflow middlebox.* Security & Communication Networks *9* (2016), No. 13, 1958–1970.

[4] R. Latif, H. Abbas, S. Assar: *Distributed denial of service (DDoS) attack in cloud-assisted wireless body area networks: A systematic literature review.* Journal of Medical Systems *38* (2014), No. 11, 1–10.

[5] R. Yan, Q. Zheng, H. Li: *Combining adaptive filtering and IF flows to detect DDoS attacks within a router.* KSII Transactions on Internet and Information Systems *4* (2010), No. 3, 428–451.

[6] A. Manimaran: *An extemporized confidence based filtering technique to mitigate Ddos attack in cloud environment.* International Journal of Control Theory and Applications *8* (2015), No. 5, 2405–2413.

[7] P. S. Mann, D. Kumar: *A reactive defense mechanism based on an analytical approach to mitigate DDoS attacks and improve network performance.* International Journal of Computer Applications *12* (2011), No. 12, 43–46.

[8] N. A. Suryawanshi, S. R. Todmal: *DDoS attacks detection of application layer for web services using information based metrics.* International Journal of Computer Applications *117* (2015), No. 9, 22–30.

[9] N. Jeyanthi, N. C. S. N. Iyengar: *An entropy based approach to detect and distinguish DDoS attacks from flash crowds in VoIP networks*. International Journal of Network Security *14* (2012), No. 5, 257–269.

Received July 12, 2017