Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

A DDoS Attack Detection Method Based on Machine Learning

To cite this article: Jiangtao Pei *et al* 2019 *J. Phys.: Conf. Ser.* **1237** 032040

View the article online for updates and enhancements.

You may also like

- A passive DDoS attack detection approach based on abnormal analysis in SDN environment
Shimin Sun, Xinchao Zhang, Wentian Huang et al.
- A Comprehensive Analysis of DDoS attacks based on DNS
Lei Fang, Hongbin Wu, Kexiang Qian et al.
- DDoS Detection and Protection Based on Cloud Computing Platform
Tianwen Jili and Nanfeng Xiao

This content was downloaded from IP address 49.207.245.96 on 19/07/2023 at 19:23

IOP Conf. Series: Journal of Physics: Conf. Series **1237** (2019) 032040 doi:10.1088/1742-6596/1237/3/032040

A DDoS Attack Detection Method Based on Machine Learning

Jiangtao Pei[1], Yunli Chen[1*], Wei Ji[1*]

Beijing University of technology Chaoyang District, Beijing(100124) , China

534893313@qq.com

Abstract. Distributed denial-of-service attack, also known as DDoS attack, is one of the most common network attacks at present. With the rapid development of computer and communication technology, the harm of DDoS attack is becoming more and more serious. Therefore, the research on DDoS attack detection becomes more important. Nowadays, some related research work has been done and some progress has been made. However, due to the diversity of DDoS attack modes and the variable size of attack traffic, there has not yet been a detection method with satisfactory detection accuracy at present. In view of this, this paper proposes a DDoS attack detection method based on machine learning, which includes two steps: feature extraction and model detection. In the feature extraction stage, the DDoS attack traffic characteristics with a large proportion are extracted by comparing the data packages classified according to rules. In the model detection stage, the extracted features are used as input features of machine learning, and the random forest algorithm is used to train the attack detection model. The experimental results show that the proposed DDoS attack detection method based on machine learning has a good detection rate for the current popular DDoS attack.

1. Introduction

Denial-of-service(DDoS) attack refers to the use of client/server technology to combine multiple computers as an attack platform to launch attacks on one or more targets to increase the power of the attack[1]. Distributed denial-of-service attack has changed the traditional peer-to-peer attack mode, so there is no statistical rule for attack behavior, in addition, common protocols and services are used in the attack. It is difficult to distinguish attack or normal behavior only through the types of protocols and services. The distributed denial-of-service attack is not easy to detect[2]. At present, the research on defense technology against DDoS attack at home and abroad is mostly based on the method of network intrusion detection. According to the characteristics of many-to-one attack in the process of DDoS attack, three characteristics[3-5] including the number of source IP addresses, the number of destination ports and the flow density were used to describe the characteristics of attack. These methods can distinguish whether most of the attack flows are rational, but only use less message information, most of which only use the source IP address and destination port information, and can not determine the specific attack type, so the detection rate is not high. Machine learning plays an important role in prediction. DDoS attack detection based on machine learning also has made some progress. The machine learning algorithms used for DDoS attack detection mainly include naive Bayesian algorithm, hidden Markov model and support vector machine[6]. Tama's team[7] used the method of anomaly detection to model the network data stream according to the header attribute, and used the naive Bayesian algorithm to score each arriving data stream to evaluate the rationality of the

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

message. The methods in the above literature improve the detection accuracy to a certain extent, but do not make full use of the context of the data stream[8].

This paper proposes a DDoS attack detection method based on machine learning. Based on the previous research, through the analysis of the principle of DDoS attack, the three common attack packets obtained by operating the DDoS attack tool are grouped in the feature extraction stage. Through the analysis of normal flow data, the characteristics of attack flow are obtained. The characteristics of the attack traffic obtained in the model detection phase are trained in the training model based on the random forest algorithm. Finally, the test model is validated by the DDoS attack, and the SVM method in the machine learning is compared in terms of detection accuracy. The results show that the DDoS attack detection method based on machine learning proposed in this paper has a good detection rate for the current popular DDoS attack.

2. DDoS Attack Detection Method Based on Machine Learning

This paper uses the common DDoS attack tool to conduct local attacks. The packet capture tool compares the captured attack packets with the normal data packets, finds the rules of the attack data, and converts them into the characteristics of the attack data, thus serving as machine learning. Model input for training.

I learned through the query that there are many open source DDoS attack tools, such as Synk4, Hyenac, LetDown, Hping, PenTBox and TFN2K. This paper selects the commonly used TFN2K as an attack tool to obtain DDoS attack traffic. TFN2K (Tribe Flood Network 2000) is an upgraded version of TFN written by the famous German hacker Mixter. It can cooperate with one or more attack targets by using a large number of agents to break the machine's resources. It can support multiple DDoS attacks, such as Common TCP, UDP and ICMP flood attacks, etc., and the tool is highly portable.

The packet capture tool will use the powerful network data analysis tool TcpDump. As one of the Internet's classic system administrators, TcpDump can analyze the data packets acquired by the network according to the user definition, and the operation is simple. For example, the command "tcpdump -i en1 -v tcp" can be used to receive the network card en1. The packet of the tcp protocol is crawled.

After the three attack packets of TCP, UDP and ICMP flood are grouped, the normal TCP, UDP and ICMP traffic are compared and found to be transformed into the characteristics of the respective attack modes:

Table 1. Normal TCP data is compared with TCP flood attack packets.

type of data	Normal TCP data	TCP flood attack data
Difference point 1	Packet sequence number is regular	Packet sequence number is random
Difference point 2	The source IP is regular, and the destination IP is more than one	Source IP is confusing, only one destination IP
Difference point 3	Packet identification bits are different	The packet identifier is the same as the packet sent by the TCP flood attack

Table 2. Normal UDP data is compared with UDP flood attack packets.

type of data	Normal UDP data	UDP flood attack data
Difference point 1	Specific port number	Random port number
Difference point 2	Source IP is regular	Source IP confusion
Difference point 3	Irregular packet length	The packet length is the same

Table 3. Normal ICMP data is compared with ICMP flood attack packets.

type of data	Normal ICMP data	ICMP flood attack data
Difference point 1	Packet sequence increment	Packet sequence confusion
Difference point 2	Invariant identifier	Random identifier
Difference point 3	One to one or one to many	IP addresses are rarely duplicated

In summary, according to the comparative analysis of normal protocol data and protocol attack data, the characteristics of common TCP, UDP, and ICMP flood attacks can be summarized as:

TCP_FEA=（TCP_NUM，TCP_LEN，TCP_TIM，TCP_IDEN）

UDP_FEA=（UDP_NUM，UDP_LEN，UDP_PLEN，UDP_TIM）

ICMP_FEA=（ICMP_NUM，ICMP_LEN，ICMP_TIM，ICMP_IDEN，ICMP_ORD）

Random forest is an important integrated learning method based on Bagging, which is usually used to solve the classification regression problem. The decision tree is used as a model for bagging. The random forest algorithm has the advantages of easy parallelization and improved prediction accuracy without significantly increasing the amount of computation. The construction process of random forests is roughly as follows:

1. From the original training set, the Bootstraping method is used to randomly select and sample m samples, and a total of n samples are generated, and n training sets are generated, and trained as n decision tree models respectively.

2. For a single decision tree model, s variables are randomly obtained from the nodes of each tree in the n classification trees, and the most representative variables are selected from these variables. The threshold of the classification is determined by multiple classification points.

3. There is no need for pruning in the decision tree splitting process, and each tree is split until all samples of the node belong to the same class stop.

4. The resulting multiple classification trees together form a random forest. The new samples are divided by the constructed random forest, classified and voted by the classifier.

According to the classification of attack protocols, the attack detection models are classified into three categories, namely, the TCP attack detection model, the UDP attack detection model, and the ICMP attack detection model. The specific training model steps are as follows:

1. Perform feature extraction, format conversion and dimensional reconstruction as effective data sets by attacking the attack data obtained above according to the feature values to be retained.

2. Divide the training set into K shares of the same size, select K-1 of them for model training, and the remaining one to do the cross-validation set.

3. Repeat the model by using different K values, and then select the number of decision trees corresponding to the highest average accuracy under different Ks as the number of decision trees in the random forest algorithm.

3. Test Results and Discussions

Whether the detection data is a DDOS attack data belongs to the classification problem, so the evaluation index uses the false positive rate, the detection rate and the total detection rate to analyze the experimental results. among them:

The false positive rate refers to the attack data detection as the normal behavior ratio, that is, FR=FP/(FP+TP)

The detection rate refers to the attack data detection as the proportion of attack behavior, that is, DR=TN/(TN+FN)

The total detection rate refers to the normal data detection as normal data, and the attack data detection is the proportion of attack data, that is, AR=(TP+TN)/(TP+TN+FP+FN)

TP refers to a positive sample that is predicted to be positive, and in this context is normal data predicted to be normal behavior.

TN refers to a negative sample that is predicted to be negative, and in this paper is the attack data predicted to be aggressive.

FP refers to a negative sample that is predicted to be positive, and in this paper is attack data predicted to be normal behavior.

FN refers to a positive sample that is predicted to be negative, and in this paper is normal data predicted to be aggressive.

After training the random forest model with the training data set, the remaining set of attack data packets are mixed with the normal traffic as the test set to detect the model. Cross-sampling normal traffic and attack traffic, calculating the classification behavior of each sample, and controlling the sampling flow period to control the ratio of normal traffic to attack traffic. At the same time, the LIBSVM library is used to detect the data of the SVM algorithm, and compared with the random forest model detection results. The detection results of the DDOS attack data for the three protocol types are as follows:

Table 4. TCP flood attack detection result.

Algorithm model	The sampling period (T)/s	2	4	6	8
Random forest	FR	0.14	0.15	0.15	0.16
	DR	99.15	98.69	98.50	98.10
	AR	99.93	99.67	99.57	99.49
SVM	FR	0.25	0.50	0.43	0.68
	DR	98.15	97.25	96.14	94.48
	AR	98.93	98.5	98.38	98.2

Table 5. UDP flood attack detection result.

Algorithm model	The sampling period (T)/s	2	4	6	8
Random forest	FR	0.24	0.30	0.53	0.42
	DR	97.75	96.83	95.23	93.13
	AR	99.93	99.67	99.57	99.49
SVM	FR	0.25	0.48	0.49	0.51
	DR	99.16	98.95	98.43	98.05
	AR	98.93	98.5	98.38	98.2

Table 6. ICMP flood attack detection result.

Algorithm model	The sampling period (T)/s	2	4	6	8
Random forest	FR	0.12	0.28	0.75	1.06
	DR	99.14	98.63	98.42	97.91
	AR	99.87	99.67	99.14	98.56
SVM	FR	0.91	1.12	2.75	3.33
	DR	98.22	97.22	96.34	94.41
	AR	98.87	97.79	96.90	95.49

It can be seen from the above three tables that the detection model of this paper still has a higher detection rate for the DDOS attack detection results of the three protocols as the background traffic increases, and is superior to the SVM algorithm model.

4. Conclusion

This paper proposes a new DDOS attack detection method, which is a random forest algorithm model based on machine learning. By extracting the three protocol attack packets of the DDOS attack tool, feature extraction and format conversion are performed to extract the DDoS attack traffic

characteristics with a large proportion. Then the extracted features are used as input features of machine learning, and the random forest algorithm is used to train and obtain the DDoS attack detection model. Then the normal traffic data is mixed with the attack data for model test. The experimental results show that the proposed DDoS attack detection method based on machine learning has a good detection rate for the current popular DDoS attacks.

References

[1] Zargar S T，Joshi J，Tipper D．A survey of defense mechanisms against distributed denial of service(DDoS)flooding attacks[J]．IEEE Corn— munications Surveys＆Tutorials。2013，15(4)：2046—2069．

[2] Wang Bing，Zheng Yao，Lou Wenjing，et a1．DDoS attack protection in the era of cloud computing and software—defined networking[J]． Computer Networks，2015，81(4)：308—319．

[3] Yu Shui，Tian Yonghong，Cuo Song，et a1．Can we beat DDoS attacks in clouds?[J]．IEEE Trans on Parallel and Distributed Systems，2014，25(9)：2245—2254．

[4] Kotenko I，Ulanov A．Agent—based simulation of DDOS attacks and de— fense mechanisms[J]．International Joumal of Computing，2014， 4(2)：113—123.

[5] Gupta B B，Joshi R C，Misra M．ANN based scheme to predict number of zombies in a DDoS attack f J]．IntemationaI JoumaI of Network Security，2012，14(2)：61-70．

[6] Yu Penchen, Qi Yong, Li Qianmu. DDoS attack detection method based on random forest classification model [J]. Application Research of Computers, 2017, 34(10):3068-3072(in Chinese).

[7] Tama B A，Rhee K H．Data mining techniques in DoS／DDoS attack detection：a literature review[C]//Proc of the 3rd International Con— ference on Computer Applications and Information Processing Techno— logy．2015：23-26．

[8] Tan Miao. Research and Implementation of DDoS Attack Detection Based on Machine Learning in Distributed Environment [D],2018(in Chinese).