A Small Sample DDoS Attack Detection Method Based on Deep Transfer Learning

Jiawei He
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
gavin.x.h@foxmail.com

Yejin Tan
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
yjtan0118@foxmail.com

Wangshu Guo*
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
* Corresponding author: cqguowangshu@163.com

Ming Xian
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
qwertmingx@sina.com

Abstract—**When using deep learning for DDoS attack detection, there is a general degradation in detection performance due to small sample size. This paper proposes a small-sample DDoS attack detection method based on deep transfer learning. First, deep learning techniques are used to train several neural networks that can be used for transfer in DDoS attacks with sufficient samples. Then we design a transferability metric to compare the transfer performance of different networks. With this metric, the network with the best transfer performance can be selected among the four networks. Then for a small sample of DDoS attacks, this paper demonstrates that the deep learning detection technique brings deterioration in performance, with the detection performance dropping from 99.28% to 67%. Finally, we end up with a 20.8% improvement in detection performance by deep transfer of the 8LANN network in the target domain. The experiment shows that the detection method based on deep transfer learning proposed in this paper can well improve the performance deterioration of deep learning techniques for small sample DDoS attack detection.**

Keywords-deep learning; transfer learning; small sample DDoS attack;

I. INTRODUCTION

DDoS attacks use the same attack technique as DoS, but the attack is implemented by a large number of zombie devices under control [1]. Distributed zombie devices make frequent requests to the attacked device, which depletes its resources and eventually crashes the server. Research [2] shows that although the number of means used to maintain network security has increased, DDoS attacks are gradually evolving, and their destructiveness to the network is gradually increasing. On the one hand, the traditional congestion-based DDoS attack method has seen its peak traffic increase year by year during attacks. Tencent, a leading Chinese Internet company, released a research report [3] stating that the peak traffic of DDoS attacks from 2013 to 2018 was growing linearly; the peak DDoS attack launched against a game company in March 2018 has reached 1.7 Tbps.

On the other hand, new DDoS attacks are no longer satisfied with high-cost, low-return flood attacks [4]. Attackers evade traditional detection techniques, including deep packet inspection (DPI), by frequently changing the characteristics of DDoS attacks or attacking small samples. For example, SYN Flood has been used as the primary technique for DDoS attacks. However, with the online black market platform, the launch of SYN Floods has changed from massive bot machines to packet-sending machines; this has also changed the characteristics of SYN Flood attacks. In addition to flooding attacks, DDoS attacks also appear in a large number of protocol-based attack methods, such as Http Flood, UDP Flood, TCP Flood.

Traditional DDoS attack detection techniques are generally divided into misuse-based detection techniques and anomaly-based detection techniques [5]. Misuse-based detection is also known as a rule-based detection technique. The advantage of this method is that it has a high detection accuracy and low false detection rate against known attacks. However, the disadvantage of this method is that it cannot effectively detect zero-day attacks, and the rule building depends on humans. The second anomaly-based detection technique is a means to detect unknown DDoS attacks. The advantage of this method is that it can detect zero-day attacks; the disadvantage is that it has a specific false alarm rate and relies too much on the features extracted from expert experience. In recent years, due to the good end-to-end characteristics of deep learning techniques [6], researchers have used them to detect DDoS attacks with increasing amounts of data. However, new types of DDoS attacks are also generally characterized by small samples of labeled data, which can significantly deteriorate the detection performance of deep learning techniques that rely on labeled data. Therefore, research into techniques that can take advantage of deep learning end-to-end and detection of small sample DDoS attacks is valuable.

978-1-7281-4349-1/20/$31.00 ©2020 IEEE
DOI 10.1109/CCNS50731.2020.00019

II. RALATED WORK

DDoS attack detection belongs to a kind of traffic classification. The traditional methods of traffic classification can be divided into methods based on misuse and methods based on anomalies [5]. [7] proposes a DDoS attack detection method based on the random forest algorithm. The advantage of this document is that it can detect some zero-day attacks; however, when it is insufficient, it depends on the features extracted by experts. When the traffic characteristics change, the feature extraction process needs to be repeated, which can be cumbersome.

Moreover, with the advent of the era of big data, traffic classification methods based on expert feature extraction are facing a huge amount of labor. The adaptability and good end-to-end characteristics of deep learning technology in big data have led researchers to use this method to classify traffic. The literature [8] proposes an end-to-end encryption traffic classification method, which uses a common fully-connected neural network to classify and detect encrypted traffic packets. It then classifies whether the traffic packets are encrypted or not. The end features, avoiding the tedious steps of manually extracting features. But the disadvantage is that deep learning technology generally relies on a large number of labeled samples for training. Therefore, this method is not suitable for the small sample DDoS attack detection problem faced here. The literature [9] proposes a DDoS attack detection method based on ensemble learning through integrated neural network, support vector machine, random forest, and other methods; its experiments show that this detection method has good detection performance against most DDoS attacks. Similarly, the literature does not analyze the problem of DDoS attack detection in a small sample.

This paper proposes a small sample DDoS attack detection method based on deep transfer learning. The article [10] suggests transfer learning, hoping to transfer the source domain knowledge better to make the learning effect in the target domain better. Due to the excellent end-to-end characteristics of deep learning technology, the detection performance of DDoS attacks is significantly deteriorated in the face of a few labeled samples. Therefore, this article combines deep learning and transfer learning to detect small samples of DDoS attacks. The final experiment proved that performance had been improved very well.

III. PROPOSED METHOD

We define the source domain as D_S and the target domain as D_T. The goal of transfer learning is to apply the knowledge learned in the source domain D_S to the target domain D_T, to achieve better performance in the target domain. In the field of DDoS detection in this article, it is to transfer the knowledge of the old DDoS attack detection field to help improve the performance of small sample DDoS attack detection.

A. Implementation Steps

First, we use deep learning technology to train neural network clusters with up-to-standard performance on classic DDoS attacks with sufficient labeled samples. The source domain D_S selects representative DDoS attacks with sufficient labeled samples. On this source domain, train m basic neural network clusters with up to standard performance $\mathbb{N} = \{\mathrm{N}_1, \mathrm{N}_2, ..., \mathrm{N}_m\}$. For each network $\mathrm{N}_i, i \in [1, m]$, the prediction function training can be calculated using the following formula:

$$f_{\mathrm{N}_i}(x) = softmax(W^{(L+1)} h^L) \tag{1}$$

$$h^l = \sigma(W^l h^{(l-1)}), \forall l = 1, \dots, L; h^0 = x \tag{2}$$

In the formula (1,2), $f_{\mathrm{N}_i}(\mathrm{x})$ represents the prediction function of the network; h^l represents the hidden layer of the network; L represents the total number of layers of the neural network; x represents the training data input to the neural network; $\sigma(\cdot)$ represents the activation function used by the neural network.

Then, we perform transfer performance comparison, which transfer the neural network cluster $\mathbb{N}$ trained on the source domain D_S to other types of DDoS attack target domain D_T. The transfer performance comparison experiment is to fix all the parameters of the network $\mathrm{N}_i, i \in [1, m]$ except the output layer, that is, $fix(parameters_k^{\mathrm{N}_i}), k \in [1, L_{\mathrm{N}_i} - 1]$. To compare the transfer performance of neural network clusters, this paper designs a transferability metric that quantifies the transfer performance between networks:

$$S_{N_i}^{D_T} = w_1 P_1 + w_2 P_2 + \ldots + w_E P_E, i \in [1, m] \tag{3}$$

$$P_j = \frac{2 \times \mathrm{Pr}_j \times \mathrm{Re}_j}{\mathrm{Pr}_j + \mathrm{Re}_j} \tag{4}$$

In the transfer performance comparison experiment, the network $\mathrm{N}_r, \mathrm{r} \in [1, m]$ with the best transfer performance value $S_{N_i}^{D_T}$ is selected. Use network N_r to perform network transfer on the target domain. In the network transfer, the parameters contained in the first l layer of the network N_r are transferred to the first l layer of the transferred network N_T, where $1 \le l \le L_{N_T}, L_{N_r}$. The transfer formula is:

$$parameters_k^{N_T} = parameters_k^{N_r}, k \in [1, l] \tag{5}$$

Among them, $parameters_k^N$ represents the k-th layer parameters of the network $\mathbb{N}$.

Use fine-tuning technology to fine-tune the parameters of the first l layers of the transferred network N_T , that is $finetune(parameters_k^{(N_T)}), k \in [1, l]$. Train the transferred network N_T on the target domain D_T , and obtain the network performance value $P_i^{(N_T)}, i \in [1, E]$ each time.

$$P_i^{N_T} = \frac{2 \times \Pr_i^{N_T} \times \mathrm{Re}_i^{N_T}}{\Pr_i^{N_T} + \mathrm{Re}_i^{N_T}} \tag{6}$$

IV. Experimental Results and Discussion

The primary parameter setting of the simulation experiment in this paper is: the operating system is Ubuntu 16.04 64-bit OS with 64GB of memory. The software framework is Pytorch. The GPU accelerator is Nvidia RTX 2080Ti. The training batch size is 500, and the loss function is the cross-entropy loss function. The optimization function uses the built-in stochastic gradient descent optimizer in Pytorch. The training learning rate is set to 0.001. The source domain in the data set selects classic SYN-type DDoS attacks with sufficient labeled samples, and the target domain is an LDAP-type DDoS attack. Each time the neural network is trained, 80% of the data is the training data set, and the rest is the verification data set. In the final deep transfer learning DDoS attack detection experiment on small samples, we reduced the number of LDAP-type DDoS attacks labeled samples by ten times to simulate the small sample situation in the real case.

A. Transfer Comparison Experiment

In order to compare the transfer performance of multiple neural networks in the source domain, this paper designs 4 different neural networks.

TABLE I. Neural network structure table

Network Name	6LANN	7LANN	8LANN	9LANN

Layer1	L(100,500)
	BN(500)
	ReLU()
		L(100,500)
		BN(500)
		ReLU()
			L(100,500)
			BN(500)
			ReLU()
				L(100,500)
				BN(500)
				ReLU()

Layer2	L(500,1000)
	BN(1000)
	ReLU()
		L(500,1000)
		BN(1000)
		ReLU()
			L(500,1000)
			BN(1000)
			ReLU()
				L(500,1000)
				BN(1000)
				ReLU()

Layer3	L(1000,2000)
	BN(2000)
	ReLU()
		L(1000,2000)
		BN(2000)
		ReLU()
			L(1000,2000)
			BN(2000)
			ReLU()
				L(1000,2000)
				BN(2000)
				ReLU()

Layer4	L(2000,3000)
	BN(3000)
	ReLU()
		L(2000,3000)
		BN(3000)
		ReLU()
			L(2000,3000)
			BN(3000)
			ReLU()
				L(2000,3000)
				BN(3000)
				ReLU()

Layer5	L(3000,2000)
	BN(2000)
	ReLU()
		L(3000,2000)
		BN(2000)
		ReLU()
			L(3000,2000)
			BN(2000)
			ReLU()
				L(3000,2000)
				BN(2000)
				ReLU()

Layer6	L(2000,2)	L(2000,1000)
		BN(1000)
		ReLU()
			L(2000,1000)
			BN(1000)
			ReLU()
				L(2000,1000)
				BN(1000)
				ReLU()

Layer7	none	L(1000,2)	L(1000,500)
			BN(500)
			ReLU()
				L(1000,500)
				BN(500)
				ReLU()

Layer8	none	none	L(500,2)	L(500,800)
				BN(800)
				ReLU()

Layer9	none	none	none	L(800,2)

Table 1 shows the network structure of 4 neural networks, where $L(m,n)$ represents the fully connected layer, the input is m dimensional features, and the output is n dimensional. BN(c) is a layer of neural network for batch normalization [11]. The input is c-dimension and the output is c-dimension. ReLU() means a rectified Linear Unit.

After the above four neural networks with different structures are trained under the same conditions in the source domain SYN-type DDoS attack, we will apply them to the target domain LDAP-type DDoS attack for transfer comparison experiments. Finally, we can get the transfer performance values of these four networks. As shown in the following table:

TABLE II. Transferability Comparison Experiment Result

Network Name	Target domain detection performance value	Transferability value
6LANN	98.67%	19.47
7LANN	99.18%	19.63
8LANN	99.28%	19.65
9LANN	99.08%	19.60

From Table 2, we can see that network 8LANN has the best transfer performance value. Compared with randomly selecting a network as the transferred network, the transfer performance value of the network 8LANN is improved by 0.02-0.18. Next, we will directly use the network 8LANN trained in the source domain SYN-type DDoS attack detection to conduct a small sample DDoS attack detection transfer experiment.

B. Small Sample DDoS Target Domain Transfer Experiment

To simulate the small sample DDoS attack scenario in the real situation, we reduced the sample number of LDAP-type DDoS attacks by ten times.

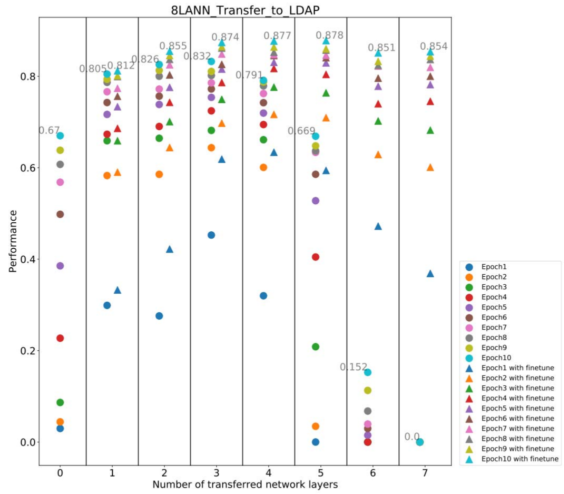

Figure1. Transfer Experiment Result

Figure 1 shows the detection performance of a small sample of LDAP-type DDoS attacks, and N_r represents the training effect of transfer different layers with or without fine-tuning. The horizontal axis x represents the first l layers of the transfer network N_r . The circular dots on the graph represent the performance value of ten epoch training without fine-tuning. The ten triangle points are the detection performance after adding fine-tuning technology. For clarity of visualization, this article shifts the dots and triangles left and right to avoid overlapping. Comparing the detection performance values of 8LANN in Figure 1 and Table 2, we can see that when a small sample for DDoS attacks, all networks' detection performance has decreased to varying degrees, which is consistent with the fact that the DDoS detection accuracy rate is low when there are few attack samples. For the network where the parameters of all layers are initialized randomly, the final detection performance drops from 99.28% to 67%. It can be seen from the triangle points in Figure 1 that by combining the fine-tuning technology, the network performance on the target domain is better than the transferred network without fine-tuning. Moreover, when the transfer depth is 5, the highest detection performance is 87.8%, and the performance when the transfer depth is 6 is 85.1% higher than that before fine-tuning. Therefore, in the present invention, the deep transfer network method combined with fine-tuning technology improves the deterioration of detection performance caused by an insufficient sample amount of new attacks.

V. Conclusions

DDoS attacks increasingly threaten today's network environment. Classical detection methods can effectively detect DDoS attacks with obvious characteristics. However, when facing a small sample of DDoS attacks, the classic method's detection performance will be greatly deteriorated due to the small amount of sample data. In response to this problem, this paper designs a small sample DDoS attack detection technology based on deep transfer learning. By learning the knowledge of old DDoS attack detection, it can improve the performance of small sample DDoS attack detection. problem. This paper first designs a transfer performance index that quantifies network transfer capabilities, and compares the transfer performance values of multiple network structures in the source domain. The final transfer comparison experiment shows that the 8LANN network's transfer performance value is 10-20% higher than that of other networks. Next, we use the 8LANN network trained in the source domain SYN-type DDoS attack to transfer to a small sample of LDAP-type DDoS attacks in the target domain. The transfer experiment proves that the detection method based on deep transfer learning can well improve the problem of the deterioration of the detection performance of small samples of DDoS attacks. The improved performance results can be increased by 20%.

References

[1] Chaudhari R S , Talmale G R . A Review on Detection Approaches for Distributed Denial of Service Attacks[C]// 2019 International Conference on Intelligent Sustainable Systems (ICISS). 2019.

[2] Jaafar G A , Abdullah S M , Ismail S . Review of Recent Detection Methods for HTTP DDoS Attack[J]. Journal of Computer Networks & Communications, 2019, 2019:1-10.

[3] Tencent-Cloud:https://www.cnblogs.com/qcloud1001/p/9177111.html

[4] Zhang C , Cai Z , Chen W , et al. Flow level detection and filtering of low-rate DDoS[J]. Computer Networks the International Journal of Computer & Telecommunications Networking, 2012, 56(15):3417-3431.

[5] Sharafaldin I , Lashkari A H , Hakak S , et al. Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy[C]// 2019 International Carnahan Conference on Security Technology (ICCST). 2019.

[6] Wang W , Zhu M , Wang J , et al. [IEEE 2017 IEEE International Conference on Intelligence and Security Informatics (ISI) - Beijing, China (2017.7.22-2017.7.24)] 2017 IEEE International Conference on Intelligence and Security Informatics (ISI) - End-to-end encrypted traffic classification with one-dimensional convolution neural networks[J]. 2017:43-48.

[7] Qian Jijun, Li Junhua, Chen Chen, etc. Network intrusion detection method based on random forest[J]. Computer Engineering and Applications, 2020.

[8] Lotfollahi M, Siavoshani M J, Zade R S, et al. Deep Packet: A Novel Approach For Encrypted Traffic Classification Using Deep Learning[C]. soft computing, 2020, 24(3): 1999-2012.

[9] Yao Honglei. DDoS attack detection based on integrated learning [D].

[10] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[11] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. arXiv: Learning, 2015.