An Improved DDoS Attack Detection Model Based on Unsupervised Learning in Smart Grid

Zhili Ma[1,2], Hongzhong Ma[1], Xiang Gao[2], Jiyang Gai[3], Xuejun Zhang[3(✉)], Fucun He[3], and Jinxiong Zhao[1]

[1] State Grid Gansu Electric Power Research Institute, Lanzhou 730070, Gansu, China
[2] State Grid Gansu Electric Power Company, Lanzhou 730000, Gansu, China
[3] School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China
xuejunzhang@mail.lzjtu.cn

Abstract. The bidirectional communication system in smart grid is vulnerable to distributed denial of service (DDoS) attacks, due to its characteristics of complex system structure and difficult to control. The multiple nodes in the smart grid system will be compromised when the DDoS attack happen, thus resulting in the denial of legitimate services to users and disruption of the normal operation in power grid system. In order to defense such attack, some detection methods have been proposed in recent years. However, most of the existing detection methods have the characteristics of low detection accuracy and high false positive rate. In this paper, we proposed a novel DDoS attack detection method which only uses unlabeled abnormal network traffic data to build the detection model. Our method firstly uses Balanced Iterative Reducing and Clustering Using Hierarchies algorithm (BIRCH) to pre-cluster the abnormal network traffic data, and then explores autoencoder (AE) to build the detection model in an unsupervised manner based on the clustering subsets. In order to verify the performance of our method, we perform experiments on KDDCUP99 dataset and compare our method with existing classical anomaly detection methods. Results show that the proposed method has higher detection accuracy for abnormal traffic detection.

Keywords: Smart grid · DDoS attack detection · Autoencoder · BIRCH algorithm · Unsupervised learning

Supported by the NSFC project (grant no. 61762058, and no. 61861024), the Science and Technology project of Gansu Province (grant no. 20JR5RA404) and the Science and Technology project of State Grid Gansu Electric Power Research Institute (grant no. 52272219100P).

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 550–562, 2021.
https://doi.org/10.1007/978-3-030-89814-4_40

1 Introduction

As the most widely distributed and complex power Internet of Things (IoT) system, smart grid combines the existing power network and the modern information network to provide users with more novel services, such as bidirectional communication system, remote controlling of smart home appliances, updating of consumer behavior and stability tracking of the power grid. These services effectively respond to the energy demand and distribution of power services. However, more and more heterogeneous terminals are generated while improving the efficiency of the smart grid, which results in a wider exposure of the grid systems. So the vulnerability of smart grids to security threats has also increased. In particular, the bidirectional communication and software-oriented nature of the smart grid makes it highly vulnerable to cyberattacks, which not only affect the normal operation of the grid system, but also disturb social stability and cause property damage. Distributed Denial of Services (DDoS) attack is one of the major security threats to smart grid, in which attackers always use multiple puppet machines to attack targets at the same time. The DDoS attack seriously affects the continuity and availability of smart grid. Thus, proactive defense, one of the most promising methods to enhance power system network security [22], is becoming more and more important, which calculates the ultimate benefits of hackers and defenders under different conditions based on the constructed model, then predicts possible attack behaviors and evaluate the best defense strategy for the power system. Therefore, it is critical to proactively and accurately detect multiple nodes in smart grid system at the same time for guaranteeing the safe operation of the grid system.

Recently, the anomaly-based detection methods are the most common ways to solve problems in intrusion detection systems, such as the machine learning-based detection methods [12] which have made ideal achievements in network security, privacy protection and so on. Machine learning-based detection methods usually include supervised learning-based methods and unsupervised learning-based methods. The attack detection methods based on supervised learning is essentially a kind of classification method, which can achieve expected detection accuracy when being provided well-labeled datasets. These methods require a large number of labeled samples which are difficult to obtain or high cost of acquisition.

Based on the existing research, this paper proposes a DDoS attack detection method based on unsupervised learning to detect abnormal traffic of the communication network in smart grid. In this paper, we firstly use BIRCH algorithm [11] to pre-cluster abnormal network traffic in an unsupervised way to obtain different patterns of clustering subsets. Then, The resulted clustering subsets are respectively input into the AE for model training, through the process of "encoding-decoding" to reconstruct the input data and obtain the average reconstruction error (the training loss of the model). Then the obtained reconstruction error is used as detection threshold to detect the normal traffic and attack traffic in the network.

The main contributions of our work are summarized as follows:

1) In this paper, we use autoencoder to train the normal traffic and abnormal traffic, and build the threshold-based abnormal traffic detection models respectively. Experimental results show that the detection accuracy using abnormal traffic to build detection model is higher than that of using normal traffic to build detection model. Therefore, it is innovative to build the anomaly detection model only using anomaly traffic data.

2) Because the training data plays an important role on the detection performance of the threshold-based anomaly detection method. Thus, we use clustering algorithm BIRCH to pre-cluster the original abnormal traffic data in a way of unsupervised learning, and the data with similar patterns can be clustered to get different clustering subsets. Experiments in this paper show that the detection model built with pre-clustered datasets can achieve higher detection accuracy.

The rest of our paper is organized as follows. Section 2 mainly introduces the machine learning based methods for DDoS attacks detection in the communication network of smart grid. Section 3 introduces the structure of the architecture of the proposed abnormal traffic detection model in this paper. Section 4 mainly presents the experimental results and analysis of the proposed method on different datasets and the comparison with previous methods. Section 5 concludes our work and discussion of future research.

2 Related Work

In this section, we make a brief introduction of some existing DDoS attack detection methods in smart grid, such as the mainstream machine learning-based detection methods and threshold-based network anomaly detection methods. Most of the previous machine learning-based methods used shallow learning or the combination of linear and nonlinear to achieve ideal detection results. For example, Rohan et al. [5] built detection models based on machine learning algorithm K-nearest neighbors (KNN), Random Forests (RF) and Decision tree (DT) to detect abnormal network traffic in IoT environments. Aamir et al. [2] used supervised machine learning techniques to classify DDoS attacks, including random forests (RF), k-neighbors (KNN) and support vector machines (SVM). Zekriet al. [21] classified the DDoS attacks using the Naive-Bayes and Decision-Tree in cloud computing environments in real time. The threshold-based network traffic detection method achieves the data training by discovering the correlation between the data features, and reconstructs the data using the method of mapping the data to the optimal subspace. In the process of forming a subspace, the data with large reconstruction error are identified as abnormal samples which are different from normal samples. The classical Principal Component Analysis (PCA) method is a threshold-based anomaly detection algorithm. For example, Anukool Lakhina et al. [9] used PCA algorithms to separate the high-dimensional space of network traffic data into non-intersecting subspaces corresponding to normal and abnormal network conditions, and performed anomaly detection on that subspaces. Paffenroth et al. [13] used robust principal component analysis (RPCA) to detect anomalies, the method achieved low FPR on individual packets and further supported

the hypothesis that the low dimensional subspace computed by RPCA is more representative of normal data. However, in the actual network environment, the majority of features of traffic data are nonlinear, and the PCA algorithm which has a process of linear transformation cannot capture the nonlinear relationship between features. Yang et al. [19] and Zhu et al. [23] have used reconstruction error in different ways to detect anomalies in network traffic. The Particle swarm optimization (PSO-BP) algorithm [20] based on BP and the OS-ELM [10] algorithm based on online extreme learning machine use reconstruction error as threshold for anomaly detection. Wei et al. [6] propose a novel DDoS attack real-time defense mechanism based on deep Q-learning network (DQN), which dynamically adjusts the service resource according to the current operating state of the system and ensures the response rate of normal service requests. PARK et al. [14] use stack noise reduction auto-encoder technology to achieve feature dimensional reduction and capture nonlinear correlation between features, and the results show that the anomaly detection method based on auto-encoder is superior than other traditional methods. Yang et al. [18] input normal traffic into AE, using minimal reconstruction error as the detection threshold to detect whether the traffic data is normal, but which cannot provide the specific type of attack, which is very important for defenders to adopt defense measures. Chen et al. [3] introduce a detection method based on unsupervised outliers, which shows that it achieves good detection accuracy. But their method integrates multi-autoencoders to a single model to detect anomaly samples, hence, the process of model training and anomaly detection has a large time overhead.

Through the analysis of the existing DDoS attack detection methods based on unsupervised learning in communication network, it can be found that the AE has been used for network anomaly detection research. But most studies used AE to learn the high-dimensional characteristics of training data and get its low-dimensional features, then using supervised learning algorithms to achieve the detection of abnormal traffic or normal network traffic data to build an anomaly detection model based on AE. In this paper, we perform experiments and find that the detection model based on abnormal network traffic has better detection performance than that obtained based on normal traffic data. Thus we propose an anomaly detection method built based on abnormal network traffic, and perform experiments on benchmark network intrusion detection dataset KDDCUP99. Results show that the model trained with the clustering subsets can achieve better performance in terms of detection accuracy.

3 Proposed Method

In this section, we present the architecture of our DDoS attack detection method as shown in Fig. 1.

As can be seen from Fig. 1, our method is an unsupervised learning-based anomaly detection framework and mainly consists of data preprocess, detection model training and detection phase. Firstly, the abnormal network traffic datasets need to be normalized and standardized, then the processed data will be input into the BIRCH algorithm and achieve the unsupervised pre-clustering

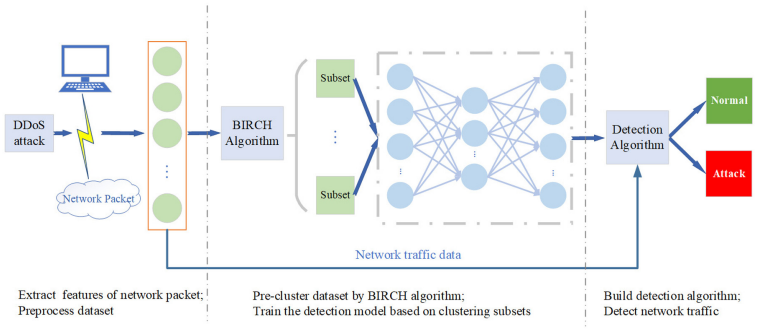

Fig. 1. Architecture of the proposed detection model in this paper.

of data using cluster feature tree (CFT). In the phase of model training, the clustering subsets are input into AE for the model training, and the minimized reconstruction error of the train data is used as the detection threshold for subsequent anomaly detection. In the detection phase, we input the test data into the trained model to get its output, then we calculate the reconstruction error between the output and its input. If the reconstruction error is higher than the preset detection threshold, the test data is marked as normal traffic data. Otherwise, it is marked as abnormal traffic data.

3.1 Data Pre-processing

In this paper, we use the open benchmark intrusion detection datasets KDD-CUP99 to perform experiments. Because the original dataset contains multiple types of data features, it directly affects the calculation of clustering features and the generation of clustering feature trees. Therefore, in our proposed method, we firstly use Min-Max techniques [8] to normalize the numerical data in the dataset and transform the raw data linearly, then we map the normalized data to [0, 1]. The conversion function is expressed as equation (1):

$$x\prime = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

3.2 Pre-classify Dataset Using BIRCH Algorithm

As we know, the network traffic data generated in different environments will produce different distribution features. Thus, the inherent change of the features of data will inevitably affect the performance of the detection model. That is, the threshold obtained by using unbalanced network traffic data with complex data patterns and distribution features could affect detection accuracy rate for detecting attack traffic. Therefore, we use the BIRCH algorithm to pre-cluster the original dataset firstly, by which the data with similar patterns will be clustered into clustering subsets. As a result, the clustering subsets are used as train

data to build detection models, which can avoid the influence of unbalanced training data on the performance of the detection model.

The important characteristic of the BIRCH algorithm is that it can accomplish high-quality clustering of large datasets with limited memory resources. In addition, BIRCH algorithm uses cluster features (CF) to summarize a cluster and cluster feature trees (CF-trees) to represent clustered hierarchies, which enables clustering methods achieve higher speed and greater scalability for operating to large databases. At the same time, this method is also very effective for incremental dynamic clustering. Therefore, it is always used in many fields in recent years.

Clustering features [16] can not only effectively reduce the storage space of data, but also efficiently calculate all the indicators in the BIRCH algorithm that form clustering decisions.

Given a cluster that contains N d-dimensional data: $\{X_i\}$, in which i = 1 ,2, . . . , N, then the following indicators can be defined:

$$D = \sqrt{\frac{\sum_i^N \sum_j^N \left(X_i - X_0\right)^2}{N\left(N-1\right)}} \tag{2}$$

In which X_0 is the center of the cluster, and the D is the average distance between the two clusters.

The BIRCH algorithm consists of four stages [17]. At the stage 1, all samples are read in turn, and a CF tree is created in memory to contain as much attribute information as possible. Stage 2 is an optional process in which memory occupancy is compressed to the desired range by scanning leaf nodes to reconstruct a smaller cluster feature tree. At the stage 3, all CF tuples are clustered using hierarchical agglomerative clustering algorithm, and a better clustering tree is obtained. In this part, formula (2) is used to calculate an accurate distance measure. Stage 4 is also an optional process which uses the center point generated by stage 3 as the seed, then scanning the original dataset again and mapping the data points to its nearest seed for better clustering results.

Based on the above theory, this paper uses BIRCH algorithm to pre-classify the dataset. By calculating the number of clusters using Davies-Bouldin, the pre-processed dataset is clustered into subsets. And the subsets are labeled for subsequent model training.

3.3 Determine Detection Threshold and Build Detection Model

Autoencoder [15] is an artificial neural network algorithm. The method proposed in this paper mainly utilizes the AE algorithm to learn the data features of the clustering subsets, and reconstructs the input data. And uses the processes of "encoding" and "decoding" to get the minimized reconstruction error of the input data. As a result, we can distinguish the normal traffic and abnormal traffic by using the obtained reconstruction error as the detection threshold.

As shown in Fig. 2. The AE model consists of encoding layer, hidden layer and decoding layer. The function of the encoding layer aims to encode the input

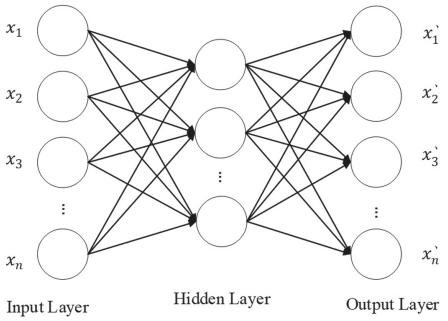

Fig. 2. The structure of the autoencoder.

data. The hidden layer aims to learn each feature of the input data. And the decoding layer aims to reconstruct the input data.

The encoding process is used to reduce the dimension of a high-dimensional features, compressing a given input data into a specified dimension which is equal to the number of neurons in the hidden layer, and the mapping of the input data x to the hidden layer is represented by h, which is expressed as (3):

$$h_i = f(x) = s(\sum_{j=1}^{d} W_{ij}^{input} x_j + b_i^{input}) \tag{3}$$

where x is the input vector, W is the weight matrix of the coding layer, b is the bias matrix, and s is a nonlinear activation function, which generally are sigmoid or Relu function, by contrast, the rule function has a better effect [7], so we use Relu function as the activation function in this experiment. It is expressed as (4):

$$f(x) = max(0, \quad x) \tag{4}$$

decoding layer mainly aims to reconstruct the input data, and decode the low-dimensional data of the hidden layer to the size of the original vector space, which is regarded as an inverse process of encoding. The mapping function is (5):

$$h_i\prime = g(h) = s(\sum_{j=1}^{d} W_{ij}^{hiddern} h_j + b_i^{hidden}) \tag{5}$$

where s is the activation function of the decoder, and we use Relu as the activation function in this experiment.

The process of training the model is to calculate the minimized average reconstruction error between input data $\{x_1, x_2, ..., x_n\}$ and the reconstructed data $\{x_1\prime, x_2\prime, ..., x_n\prime\}$. Therefore, we use mean square error to calculate the error between the test sample and its reconstructed output in this paper. It is defined as (6):

$$\theta = L(x_i, x_i\prime) = \frac{1}{n} \sum_{i=1}^{n} (x_i - x_i\prime)^2 \tag{6}$$

The proposed method uses AE to learn the most important features of input data, and the reconstruction error is obtained through the process of "encoding-decoding". Then the reconstruction error is used as the detection threshold for subsequent anomaly detection. In the test process, the test data will be input into the trained model. Then we calculate the minimized reconstruction error between the data and its original input by mean square error function. If the reconstruction error is higher than the predetermined detection threshold, the test data is marked as normal traffic data. Otherwise, it is marked as abnormal traffic data.

4 Experimental Analysis and Results

In this section, we perform experiment on benchmark intrusion detection datasets KDDCUP99 to verify the performance of our method, and compare the results with the recently machine learning-based DDoS detection methods [2,5] including RF (Random Forest), NB (Naive-Bayes), DT (Decision-Tree), and LR (Logistic-Regression). The follows are results and details about all experiments.

4.1 Introduction of the Dataset

The KDDCUP99 dataset [1], a published security audit dataset from Columbia University's IDS Labs, contains 488,734 training data for 23 different types of attacks, each data has 38 features, including source IP, source port, destination IP, destination port, transaction protocol, status, duration and attack category. Since the experiments in this paper are based on normal traffic and abnormal traffic. Therefore, we select part of data labeled with normal and Dos from KDDCUP99 dataset as datasets for follow experiments.

4.2 The Results of the Experiment

In order to verify that the detection model built by abnormal traffic data can achieve better detect performance. We use the un-clustered traffic dataset labeled with Dos and the dataset labeled with Normal as the training data to build the detection model. As shown in Fig. 3, through experiment statistics we found that the minimum reconstruction error of normal traffic data is higher than the preset detection threshold, which means that if the reconstruction error of the test sample is higher than the preset detection threshold, the traffic data is marked as normal traffic data. Otherwise, it is marked as abnormal traffic.

At the same time, in order to verify the overall performance of the model, the Accuracy, Precision, Recall and F-score are used as evaluation indicators, and these values can be obtained by the follows: $Accuracy = \frac{TP+FN}{TP+TN+FN+FP}$, $Precision = \frac{TP}{FP+TP}$, $Recall = \frac{TP}{TP+FN}$, $F-score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$, in which $TN(TrueNegative)$is used for normal traffic to be detected as normal traffic. FN (False Negative) is used for normal traffic to be detected as abnormal traffic. TP (True Positive) is used for abnormal traffic to be detected as abnormal traffic.

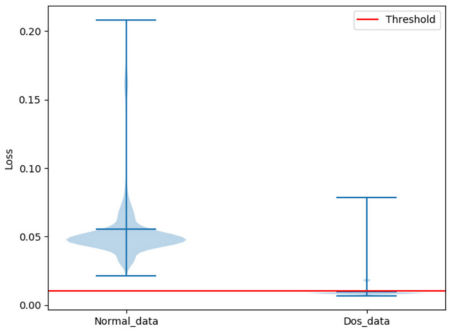

Fig. 3. Distribution of reconstruction errors for normal and Dos data

Table 1. Performance of detection models based on normal traffic and abnormal traffic.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Dos	97.128	100	98.218	99.212
Normal	83.846	97.602	98.925	95.318

FP (False Positive) is used for abnormal traffic to be detected as normal traffic. In order to ensure the accuracy of the experimental results, for each method, we carry out ten experiments and take the average results of the ten experimental as the final results. As shown in Table 1, the results show that the detection model trained based on abnormal traffic data is more efficient, and the detection Accuracy, Precision, Recall and F-score are 97.128%, 100%, 98.218% and 99.212% respectively, the corresponding value obtained by the model based on normal traffic data are 83.846%, 97.602%, 98.925% and 95.318%. Especially the detection accuracy was 13.282% higher than that of obtained by the detection model based on normal traffic data. Therefore, it is obvious that the detection model obtained by the abnormal traffic data has higher detection performance.

Therefore, the detection model in this paper is built based on abnormal traffic data. In order to get the optimal clustering subsets, we determine the best number of cluster using DBI (Davies-Bouldin index). The index is also known as classification fitness indicator which is used to evaluate the merits of clustering algorithms [4]. Typically, the smaller the indicator, the better the clustering effect. The indicator is defined as:

$$DBI = \frac{1}{N}\sum_{i=1}^{N} max(\frac{s_i + s_j}{d_{ij}}),\ i \neq j \tag{7}$$

where s_i is the average distance from each point in cluster i to the cluster center, d_{ij} is the distance between the center of cluster i and the center of cluster j, and N is the number of clusters.

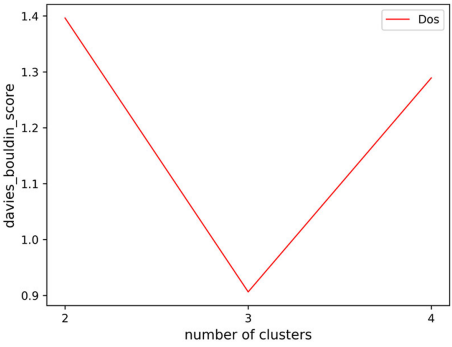

Fig. 4. Davies-Bouldin-Score

As shown in Fig. 4, we take different number of clusters to get the Davies-Bouldin score corresponding to them. We find that when the number of clusters was 3, the Davies-Bouldin score was the smallest (0.906), which brought the best clustering effect.

Table 2. Clustering results for the Dos dataset in KDDCUP99.

Dataset (Dos)	Data-size
Dos	102309
ID 0	100057
ID 1	2036
ID 2	216

The results are shown in Table 2. We use BIRCH algorithm to pre-cluster the original Dos traffic dataset in KDDCUP99, which categorizes original dataset from different data patterns and gets three subsets of clusters with different sample sizes, and the subsets were labeled ID 0 (100057 samples), ID 1 (2036 samples) and ID 2 (216 samples) respectively. Considering the comprehensive consideration, our method uses ID 0 as the train data to build the detection model.

In order to verify the overall performance of this method, this paper compared the proposed method with the recently machine learning-based DDoS attack detection methods on the KDDCUP99 dataset. In order to ensure the accuracy of the experimental results, based on each method, we carry out ten experiments and take the average results of the ten experimental as the final results. Table 3 shows the comparison of the indicators for each method, we can find the Accuracy, Precision and F-score of the proposed method are the highest, at 99.360%, 100% and 99.679% respectively. And it is the NB that obtain the highest Recall value of 99.994%, which is 0.666% higher than that of our

Table 3. The experimental results compared with the methods based on supervised learning.

Indicators	NB	RF	DT	LR	Proposed
Accuracy (%)	93.188	94.647	92.775	96.766	99.360
Precision (%)	93.007	94.942	92.953	95.949	100
Recall (%)	99.994	98.587	95.619	99.894	99.328
F-score (%)	95.776	96.648	94.018	97.671	99.679

method. Overall, our method is superior to the comparison method in terms of detection Accuracy, Precision and F-score.

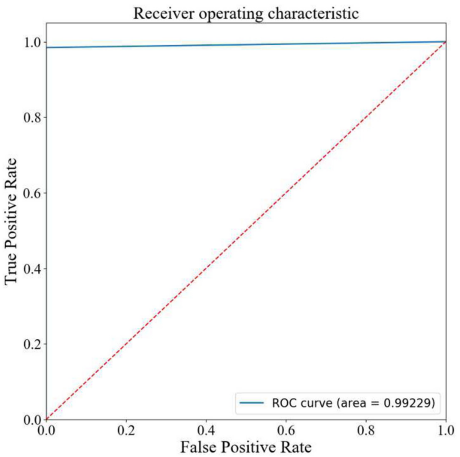

Fig. 5. ROC curve and AUC area

In order to visually evaluate the detection accuracy of the proposed method, we evaluate the classification effect of the method by ROC curve and AUC indicator. Based on the prediction results of the proposed algorithm, we calculate the values of TPR (True Positive Rate) and FPR (False Positive Rate) to form a ROC graph. The AUC is defined as the area under the ROC curve, with a range of values between 0.5 and 1. The reason to use AUC as the evaluation criterion is because in many cases the ROC curve cannot clearly explain which classifier is better. As a value, the larger of the corresponding AUC value, the better of the classifier effect. From Fig. 5, it can be known that the AUC of the proposed method in this paper reached to 0.99229, which shows that the method has a better classification effect.

Experiments show that the detection accuracy of abnormal traffic is obviously better than that of comparison method. This is because we use BIRCH cluster algorithm to pre-classify network traffic data with similar distribution features and obtain good clustered subsets. Based on the clustering subsets, we design a stacked autoencoder to train the detection model, which can learn the high-dimensional features of the pre-clustered subsets and get more suitable detection threshold. Thus, we achieve better detection performance.

5 Summary

This paper proposed an improved DDoS attack detection model based on unsupervised learning, the goal of which is to achieve more accurate and efficient DDoS attack detection in communication network of smart grid. Our method uses BIRCH algorithm to cluster network traffic data with similar features and utilizes greater cluster subset as training data. Then, we train the detection model using stacked AE in an unsupervised way, and use the average reconstruction error of train data (the training loss of the model) detection threshold. We perform experiments based on KDDCUP99 dataset, and compare the proposed method with recently machine learning-based DDoS detection methods, the results show that the proposed method is superior to the comparison method in detection performance.

In addition, the method proposed in this paper need to be further improved to adapt more extensive and complex smart grid environment. 1) In order to improve the generalization capability of the model, we will consider whether the BIRCH-AE based method is suitable for DDoS attack detection with various rates in different environment. 2) In order to enhance the robustness of the model, we consider building a real-time and adaptive attack detection framework to achieve efficient real-time monitoring of DDoS attacks in communication network of smart grid.

References

1. Dataset (1999). http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
2. Aamir, M., Zaidi, S.: Clustering based semi-supervised machine learning for DDoS attack classification. J. King Saud Univ. Comput. Inf. Sci. **33**(4), 436–446 (2019)
3. Chen, J., Sathe, S., Aggarwal, C., Turaga, D.: Outlier detection with autoencoder ensembles. In: Proceedings of the 2017 SIAM International Conference on Data Mining (2017)
4. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-*1*(2), 224–227 (1979)
5. Doshi, R., Apthorpe, N., Feamster, N.: Machine learning DDoS detection for consumer Internet of Things devices, pp. 29–35 (2018)
6. Feng, W., Wu, Y.: DDoS attack real-time defense mechanism using deep q-learning network. Int. J. Performability Eng. **16**(9), 1362–1373 (2020)
7. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. J. Mach. Learn. Res. **15**, 315–323 (2011)

8. Ihsan, Z., Idris, M.Y., Abdullah, A.H.: Attribute normalization techniques and performance of intrusion classifiers: a comparative analysis. Life Sci. J. **10**(4), 2568–2576 (2013)
9. Lakhina, A., Crovella, M., Diot, C.: Diagnosing network-wide traffic anomalies. Comput. Commun. Rev. **34**(4), 219–230 (2004)
10. Li, Y., Qiu, R., Jing, S., Li, D.: Intrusion detection system using online sequence extreme learning machine (OS-ELM) in advanced metering infrastructure of smart grid. PloS ONE **13**(2), e0192216 (2018)
11. Minxuan, Y.: Research and implementation of data mining system based on improved clustering algorithm. Ph.D. thesis, University of Electronic Science and Technology (2012)
12. Moustafa, N., Hu, J., Slay, J.: A holistic review of network anomaly detection systems: a comprehensive survey. J. Netw. Comput. Appl. **128**, 33–55 (2019)
13. Paffenroth, R., Kay, K., Servi, L.: Robust PCA for anomaly detection in cyber networks (2018)
14. Park, S., Seo, S., Kim, J.: Network intrusion detection using stacked denoising autoencoder. Adv. Sci. Lett. **23**(10), 9907–9911 (2017)
15. Spatiotemporal, E., Related, S., Context, A.I.: Unsupervised feature learning for audio classification using convolutional deep belief networks (2009)
16. Tao, Cui, X.: The research of high efficient data mining algorithms for massive data sets. Appl. Mech. Mater. **556–562**, 3901–3904 (2014)
17. Tian, Z., Ramakrishnan, R., Livny, M.: Birch: an efficient data clustering method for very large. ACM Sigmod Rec. **25**(2), 103–114 (1996)
18. Yang, K., Zhang, J., Xu, Y., Chao, J.: DDoS attacks detection with autoencoder. In: NOMS 2020–2020 IEEE/IFIP Network Operations and Management Symposium (2020)
19. Yang, S., Zhang, R., Nie, F., Li, X.: Unsupervised feature selection based on reconstruction error minimization. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019)
20. Yu, S., Xu, L.: Research of intrusion detection based on PSO-BP algorithm. J. Shazhou Prof. Inst. Technol. (2018)
21. Zekri, M., Kafhali, S.E., Aboutabit, N., Saadi, Y.: DDoS attack detection using machine learning techniques in cloud computing environments. In: 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech) (2017)
22. Zhao, J., Zhang, X., Di, F., Guo, S., Mu, D.: Exploring the optimum proactive defense strategy for the power systems from an attack perspective. Secur. Commun. Netw. **2021**(6), 1–14 (2021)
23. Zhu, Q.H., Yang, Y.B.: Subspace clustering via seeking neighbors with minimum reconstruction error. Pattern Recogn. Lett. **115**(NOV.1), 66–73 (2017)