Anomaly Detection Based on PMF Encoding and Adversarially Learned Inference

Lin Zhang¹, Wentai Yang², Hua Gan¹, Meng Li¹, Xiaoming Wang¹, Gang Liang³

¹ Chengdu City Electric Power Engineering Design Company, Chengdu, Sichuan, 610065, China
² College of Computer Science, Sichuan University, Chengdu, Sichuan, 610065, China
³ College of Cyber-security, Sichuan University, Chengdu, Sichuan, 610065, China

*Corresponding author’s e-mail: 2274504924@qq.com

Abstract. In order to solve the problem of increasing the dimension and sparse feature space caused by the categorization coding method in the existing abnormal traffic detection problem, a coding method based on Probability Mass Function (PMF) is proposed. Secondly, in order to improve the ability of abnormal traffic detection algorithms to identify unknown attack type data and improve detection efficiency, we use Adversarially Learned Inference as the basic detection algorithm. The comparison experiments on the standard dataset show that the proposed method has improved the accuracy and detection efficiency greatly compared with the existing anomaly detection methods.

1. Introduction
The network has now gone deep into life. The internet connects the world, and the intranet connects whole organization, both of which have greatly improved the efficiency of work and brought convenience to people's lives from time to time. This is undoubtedly the welfare brought by the advancement of information technology. However, it should not be overlooked that cyberspace is not a pure land. In fact, cyber security issues are becoming more and more serious. In recent years, malicious network intrusion events, cyber fraud incidents, and network information disclosure incidents with the network as the carrier and implementation path have emerged in an endless stream, which make individuals, enterprises and countries suffer economic and reputation losses. Frequent network security incidents have made network users pay attention to network security, and people are gradually realizing the importance of network security to the network age. The frequent cyber security incidents, and subjective demand of cyber security, give great importance to network security related works.

Network security field can be generally divided into sub-fields such as intrusion detection, intrusion prevention, information confidentiality, security auditing and etc. Among many network security works, network abnormal traffic detection belonging to intrusion detection is of great significance. The main reason is that it is often used as a means of finding problems. For network security workers, the ability to find problems is very important. If it is impossible to accurately determine whether an intrusion event occurs on the network, it is impossible to make a timely response, so that subsequent operations such as intrusion event implementation response and system state recovery cannot be performed. Network abnormal traffic detection is a means to discover abnormal traffic based on the normal use state of the network, and thus provide more warning information for intrusion detection.
According to the adopted detection ideas, the abnormal traffic detection methods can be divided into four classes, which are respectively based on statistics, information theory, classification and clustering [1]. 1) Probability and statistics theory has an abnormal point detection algorithm. For example, in the case of one-dimensional data and assumption of normal distribution, if the difference between the mean and the value to be detected is greater than 3 times of standard deviations, it can be marked as an abnormal point with high probability. For multidimensional data, in addition to the extended one-dimensional anomaly detection method, there are correspondingly multivariate Gaussian distribution method, chi-square method, PCA method, etc; 2) Information Entropy in information theory is often used for abnormal traffic detection, which monitors the overall traffic rate. Information Entropy-based method assumes that Information Entropy will change correspondingly when anomaly occurs. So that abnormal traffic detection depends on monitoring of Information Entropy of whole traffic. Such methods have strong applicability to attack types with significant changes of traffic information entropy (eg DDOS); 3) cluster-based methods use unsupervised machine learning clustering algorithms as model algorithms, such as K-Means. Clustering-based detection methods have innate advantage of detection in the context where attack types are unknown; 4) classification-based method is mainly based on machine learning classification method to detect abnormal traffic, involved algorithms are: Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Neural Network (NN), and etc. Most of these methods use supervised machine learning where annotated data is necessary, whereas their detection efficiency and accuracy are relatively high.

The main work of this paper is as follows: The first part introduces the significance and background of abnormal traffic detection. The second part analyses and summarizes related network abnormal traffic detection works. The third part introduces our contributions, including Probabilistic Quality Function (PMF) based encoding method and abnormal traffic detection algorithm based on the Adversarially Learned Inference (ALI). The fourth part describes experiment on the NSL-KDD dataset, and analysis and discussion are given. At last, we summarize our contribution.

2. Related works
In this section, we first introduce related existing works of four types mentioned above in section 1 briefly, and then we summarize the problems in them.

2.1. Anomaly Traffic Detection Methods
Statistical-based methods appear earlier, its basic idea of anomaly detection is to determine whether observations fall within some confidence interval. Applied methods include Chi-Square Distribution, Wavelet Transform, Smooth Regression, Principal Component Analysis (PCA), and etc. Qian YK [2] proposed an anomaly detection algorithm that anomaly was detected by comparing Chi-Square statistic of both test data and normal data. W LU [3] applied the Wavelet Transform to anomaly detection and verified its algorithm performance on DARPA dataset, where dataset was organized in a time series style. HZ Moayedi [4] applied Autoregressive Integrated Moving Average model (ARIMA) to model time series data, comparing with model of normal data. M-L Shyu [5] detected abnormal traffic based on PCA, which was used to extract better features for classifier.

Main idea of Information Entropy-based methods is that Information Entropy of abnormal traffic is special, so such methods detect anomaly by monitoring Information Entropy of stream data. Information Entropy based methods perform well in the case of attack type are unknown, this is because that no specific label is applied. Lakhina [6] applied the Information Entropy based on feature distribution to abnormal traffic detection task for the first time. Zheng LM [7] studied the Information Entropy based traffic classification problem in the multi-dimensional case.

The principle of cluster-based abnormal traffic detection methods is that similar traffic is classified into one class according to certain traffic similarity threshold. While traffic that is not classified into any cluster is considered to be abnormal. Similarly, cluster-based methods are also applicable to the case where attack type is unknown. L PORTNOY [8] used clustering algorithm to realize abnormal traffic detection for the first time, which gave abnormal traffic detection the ability to detect unknown attack...
types. Zuo J [9] improved selection method of the initial clustering centers in K-means clustering algorithm, avoiding the situation where outlier points were selected as the initial clustering centers, thereby reducing the number of iterations, and applying K-means clustering algorithm more efficiently to abnormal traffic detection task.

Classification-based anomaly traffic detection methods use annotated data. Conventional machine learning algorithms have good classification performance and are widely used in abnormal traffic detection tasks. Generally, detection efficiency and accuracy of these methods are better. Zhu Yingwu [10] regarded abnormal traffic detection problem as binary classification problem based on Information Entropy and SVM. H SAXENA [11] used Information Gain to pre-selected features that were beneficial to classification, then the model was validated on the KDD-99 dataset and SVM algorithm. Li Q [12] used C4.5 algorithm to detect abnormal traffic, and optimal features were found while model was constructing. In addition to conventional machine learning algorithms, Neural Network (NN) based anomaly traffic detection methods have also been studied. In abnormal traffic detection task, NN is usually used as dimension reduction method as well as detection algorithm. 1) Because of its strong representation of dataset, NN is often used as a pre-training part for constructing hybrid classification model, which plays the role of dimension reduction and feature extraction. J YANG [13] combined Restricted Boltzmann Machine (RBM) with SVM to construct a hybrid abnormal traffic detection model, when Spark was leveraged to accelerate training. F LIU [14] used Deep Belief Networks (DBN) as feature extraction and dimension reduction method to construct a hybrid APT detection model with Support Vector Data Description (SVDD). 2) NN can be also used as detection algorithm. U FIORE [15] proposed a discriminative restricted Boltzmann machine (DRBM), which gives RBM the ability for classification, and finally achieving semi-supervised abnormal traffic detection. D WULSIN [16] used DBN as semi-supervised classification method to verify model performance on clinical medical image dataset. For the first time, J AN [17] imported Variational Auto-Encoder (VAE) into anomaly detection task, and then compared the model with Auto-Encoder(AE) and PCA on MNIST dataset and KDD-99 dataset. For the first time, T SCHLEGL [18] applied Generative Adversarial Network (GAN) in anomaly detection, and proposed a new measurement for classifier based on Feature Matching (FM) beyond existing Cross entropy based measurement. The work trained the model based on both generator loss and classifier loss, and performance of the model was tested on image dataset.

2.2. Challenge of Existing Methods

The above related works involve nearly all aspects of the field of abnormal traffic detection. Except non-negligible contributions, there are also challenges they are faced with. We conclude as bellow:

1) Existing methods can hardly satisfy the requirements of both identifying unknown attack and high detection efficiency. Among existing methods, there are three kinds of algorithms for identifying unknown attack types: one class classification-based methods, clustering-based methods and generation model based methods. Among them, the accuracy of clustering-based methods and one class classification-based methods are low relatively. While generative model based methods have higher accuracy, whereas their detection efficiency is too low to meet practical demands.

2) Existing encoding methods for categorical feature bring about dimensions expansion problem. These encoding methods are One-Hot Encoding, Dummy Encoding, and Label Encoding. The first two are similar with each other in principle and form. They both are based on all possible values of categorical features. Due to the expansion of original data dimension after encoding, the feature space becomes very sparse and harmful to model training. Therefore, One-Hot Encoding or Dummy Encoding is often applied together with dimensionality reduction technology such as PCA. Label Encoding assigns specific number to each possible value of categorical features. Although this encoding method can avoid dimension expansion, unnecessary bias may be drawn into model because that encoding numbers have little correlation with true distribution of features.

3. Our model
Abnormal Traffic Detection model work on a series of input data, and specify prediction label of current input. Actual network traffic data usually are multi-dimensional. The \( i \)th input data can be denoted as vector form \( T_i = \{ f_1, f_2, \ldots, f_n \} \). The abnormal traffic detection model \( g(T_i) \) outputs the label of \( T_i \), so the detection process can be illustrated as formula (1):

\[
C_i = g(T_i)
\]  

(1)

If the true label of current input is denoted as \( G_i = \{ T, F \} \), then the training goal of the model is to minimize error rate of the model on the training dataset. And minimizing the model error rate is equivalent to minimizing the following loss function:

\[
S(D) = \sum_{i=1}^{m} \left\{ 0, G_i = C_i \right\} + \left\{ 1, G_i \neq C_i \right\}
\]  

(2)

Where \( D \) denotes the input dataset, \( m \) denotes the size of this dataset.

There are many factors that affect the performance of abnormal traffic detection model, such as data preprocessing methods, data dimensions and classification algorithms. This paper mainly works on encoding methods for categorical features and anomaly detection algorithms adopted in abnormal traffic detection. We propose a new abnormal traffic detection method based on Probabilistic Quality Function (PMF) and Adversarially Learned Inference (ALI).

3.1. PMF encoding

Most of existing related works rely on vectorization and normalization for input features. For non-continuous features, also known as categorical features, specific encoding is needed to convert them to continuous features firstly. Conventional encoding methods for categorical features include Label Encoding, One-Hot Encoding, Dummy Encoding.

Label Encoding designates encoding value for categorical features based on an encoding dictionary built manually. For example, for feature “gender”, the encoding result may be 1 for male and 0 for female. The problem with Label Encoding is that the encoding value is meaningless and can hardly reflect true distribution. In the above example, it is difficult to explain why the value of male is larger than the value of female, and actually 0 for male and 1 for female is also okay in Label Encoding. However, different encoding results lead to different classification performance, and the detection performance of classifier may be hurt when encoding result has nothing to do with true data distribution.

One-Hot Encoding method overcome the shortcomings of Label Encoding by generating encoding value from feature distribution instead of manual labor. It expands dimension according to number of whole possible values of categorical features, and every dimension represents a possible value of current feature. Due to that one feature can only have one value for current input data, there is only one dimension getting non-zero encoding value while the other dimensions get zero value. Dummy Encoding is very similar to One-Hot Encoding method. The only difference is that Dummy Encoding represent a value of the feature by zero for all encoding dimensions. As a consequence, Dummy Encoding could save one encoding dimension. The problem with One-Hot Encoding and Dummy Encoding is dimension expansion, which reduces model efficiency. And dimension expansion also causes sparse and high-dimensional feature space. In extreme cases, it may cause Curse of Dimensionality, which makes model hard to train.

In this paper, we proposed a new encoding method for categorical feature based on PMF. PMF denotes probability of Discrete Type Random Variable in statistics. It has the following three important characteristics: 1) each value is between 0 and 1; 2) the sum of the values of equals 1; 3) each value denotes probability of a value of current feature. If the value of the PMF in a certain category is large, it indicates that the sample of the category has a large proportion in the dataset. Thus PMF indeed reflects the true distribution of feature.

The aforementioned three characteristics of PMF make PMF not only satisfy the requirements of vectorization and normalization but also be meaningful to reflect origin information of dataset. Compared with Label Encoding, PMF based encoding wouldn’t disturb classifier. Moreover, PMF is a
one to one encoding method, which means that the dimension stays constant and there is no dimension expansion problem like in One-Hot Encoding and Dummy Encoding.

Figure 1 illustrates PMF based encoding by the “weather” example. Suppose all possible values of feature “weather” are rainy, sunny, cloudy, and the values above bars are PMF values of corresponding weather in current dataset. For example, 0.3 of “rainy” means there are 30 percent rainy data of all data. And we can also find the total PMF of all possible weather (including rainy, sunny and cloudy) equals 1, which means PMF based encoding normalizes data at the same time.

For an actual dataset $D$, the probability is represent by frequency. So the PMF encoding value $p_{c,k}$ for feature $c$ and its categorical value $k$ can be computed by formula (3):

$$p_{c,k} = \frac{\sum_{i=1}^{n} I(i,c,k)}{|D|}$$  

(3)

$$I(i,c,k) = \begin{cases} 0, & i_c \neq k \\ 1, & i_c = k \end{cases}$$  

(4)

$|D|$ represents the size of the dataset, and $i_c$ represents the value of feature $c$ of sample data $i$.

Compared with One-Hot Encoding and Label Encoding, PMF based encoding shows obvious dimensionality reduction effect. Giving the number of features $N$ and the number of categorical features $C$, the ratio of dimensionality reduction of the PMF based encoding can be calculated by formula (5):

$$R_{c \rightarrow \infty} = \frac{N + \sum_{i=1}^{C} x_i}{N + C} \geq 1 + \frac{C}{N + C} \approx 2$$  

(5)

$x_i$ denotes the number of value of feature $i$. Because that $x_i$ is greater than 2 at least, so the lower bound of $R$ is always greater than 2 when $C$ approaches infinity. Namely, we can at least get 2 times of dimensionality reduction of One-Hot Encoding and Dummy Encoding when categorical features are the main part of all features.

3.2. Anomaly Detection Based on ALI

According to Section 2.2, existing abnormal traffic detection related works cannot guarantee both of high detection performance and the detection ability for unknown attack types. To solve this problem, we investigate the latest research progress about abnormal traffic detection and deep generative model, and we propose a new detection method based on Adversarially Learned Inference (ALI).
Adversarially Learned Inference (ALI) was first proposed by V DUMOULIN [20] based on both of Variational Auto-Encoder (VAE) and Generative Adversarial Network (GAN) to overcome existing problems with which VAE and GAN were faced. While the model was evaluated on MNIST dataset. In this paper, we apply ALI to abnormal traffic detection task.

GAN, VAE and ALI are deep generative models, which means that the methods based on these algorithms have ability to identify unknown attack because of their unsupervised characteristic. Moreover, because of strong representation of deep networks, these methods may get higher accuracy than clustering based methods and classification based methods such as One-class SVM

3.2.1. Methods based on VAE and GAN

VAE is a deep generative network composed of an encoder and a decoder. The principle of generating data is as follows: firstly, the hidden variable $z$ is obtained by sampling from the conditional distribution $p(z \mid x, \theta)$, and then new data is generated by sampling from the conditional distribution $p(z \mid x, \theta)$. Encoder is used to train and get $p(z \mid x, \theta)$ and decoder is used to train and to get $p(x \mid z, \theta)$. When training is successful, generated data will be similar to origin input data. The basic detection principle of VAE based methods is to compare differences between generated data and origin data. For test data, if the difference is small, the test data and training data may belong to the same class with a high probability, which means the test data is normal. If the difference is large, the test data may belong to the opposite class of training data, in this case, the test data is likely to be anomaly data. Figure 2 shows the structure of the three-layer VAE.

The loss function of VAE is based on Cross Entropy. Equation (6) shows how to compute Cross Entropy. Equation (7) shows the loss function of VAE.

$$H(p, q) = -\sum_x p(x) \log q(x)$$

$$\zeta(q) = E_{z \sim q(z \mid x)} \log p_{\text{model}}(z \mid x) + H(q(z \mid x))$$

$$H(q(z \mid x)) = E_{z \sim q(z \mid x)} \log p_{\text{model}}(x \mid z)$$

$$-D_{KL}(q(z \mid x) \parallel p_{\text{model}}(z))$$

GAN is a deep generative model based on game theory. It consists of generative network $G$ and discriminant network $D$. $G$ tries to forge fake data that can be judged as true by $D$, and $D$ tries to distinguish the generated forged data from $G$. $G$ and $D$ are trained by turn. According to the zero-sum rule in game theory, training ends up when it is unable to obtain more benefits for either $G$ and $D$. Therefore, the model of GAN is trained by formula (9).

$$g^* = \arg \min_g \max_{d} v(g,d)$$

$$v(g,d) = E_{x \sim p_{\text{data}}} \log d(x) + E_{x \sim p_{\text{model}}} \log(1-d(x))$$
$p_{data}$ represents the real distribution of input data, and $p_{model}$ represents the distribution of generated data. The training process of GAN is shown in Figure 3.

![Figure 3 Training process of GAN](image)

Compared with VAE, GAN has lower error[20], but since GAN doesn’t have inference network which maps the input data to the hidden layer variables, it takes a lot of time to recover the hidden layer representation in the test phase in order to calculate the overall loss, so the test efficiency of GAN based methods is much lower than the VAE based method.

3.2.2. Anomaly detection based on ALI

ALI is a deep generative network that have complementary advantages of VAE and GAN. Specifically, ALI adds an inference network to GAN (similar to the generator in VAE), and ALI includes generative network, discriminant network and inference network, thus ALI needs only a little time to restore hidden distribution of test data, which makes calculation of overall test error more efficient.

Unlike GAN, the discriminant network in ALI receives a vector pair at the same time, and there are two cases for this vector pair: 1) real data and its encoding; 2) generated data and generated hidden variables in the network. The goal of the discriminator is to determine if they match with each other. ALI is trained by equation (11).

$$g^* = \arg\min_{g} \max_{d} V(d, g)$$

$$V(d, g) = E_{q(x)}(\log(d(x, g(x)))) + E_{p(z)}(\log(1 - d(g(z), z)))$$

$q(x)$ represents the mapping of the actual feature vector to the hidden layer vector, and $p(x)$ represents the mapping of the hidden layer vector to the actual feature vector. Figure 4 shows the training process of ALI.

The principle of ALI for anomaly detection is similar to VAE and GAN. Firstly, normal data is used to train ALI, and then the test data including the abnormal data is used to test the model. Namely, the abnormal traffic is detected by monitoring the difference of generated data and test data. If the difference of generated data and test data is found to be big, the current test data can be determined as abnormal data.
The abnormal traffic detection task has the requirements of algorithm detection accuracy, detection efficiency and the ability to recognize unknown attacks. Compared with VAE and GAN, ALI algorithm has higher detection accuracy and higher detection efficiency[20]. ALI also belongs to unsupervised generation model, which means that it has the ability to identify unknown attack. Therefore, we apply ALI into abnormal traffic detection task.

4. Experiment

4.1. Dataset

Table 1 Categorical features in KDD-99 dataset

| Features          | Number of possible values |
|-------------------|---------------------------|
| protocol type     | 3                         |
| service           | 70                        |
| flag              | 11                        |
| land              | 2                         |
| logged_in         | 2                         |
| is_host_login     | 2                         |
| is_guest_login    | 2                         |

We evaluate our model based on PMF encoding and ALI with others on real abnormal traffic detection dataset KDD-99. This article used 10% of the KDD-99 dataset as experimental dataset. According to the PMF based encoding method proposed in section 3.1, there are seven categorical features are encoded. Table 2 describes these seven categorical features. The feature dimension based on PMF
encoding is 41, and the feature dimension based on One-Hot encoding is 121. Therefore, our encoding method for categorical features reduces the feature dimension by 66%.

In this paper, the features except above seven features are processed by the maximum and minimum normalization method (Maxmin), and the Maxmin value of feature $x$ for sample $i$ can be calculated by formula (13).

$$f(x_i) = (x_i - \min(x)) / (\max(x) - \min(x))$$

(13)

4.2. Environment and evaluation indicators

4.2.1. Experimental environment.
Ubuntu 16.04 LTS operating system, Inter(R) Core(TM) i7-6700 CPU @3.40GHz 3.41GHz processor, 12.0GB RAM.

4.2.2. Evaluation indicator
a) Detection accuracy $F_1$-measure, the higher value indicates the higher accuracy;
b) Train time $\text{train-t}$ and test time $\text{test-t}$, the shorter time indicates the higher efficiency.

4.3. Training
The experiment dataset is divided randomly into training datasets and test datasets in a ratio of 1:1. For the training dataset, only the normal data is retained, and the test dataset is not further processed.

The setting of hyper-parameters in ALI model refers to existing related work[21]. In addition, batch size is 50, initial learning rate and initial weight are 0.1. On our dataset, it takes 16 minutes and 20 seconds to finish model training.

4.4. Results and Discussion
We compare our methods with SVM based on binary classification [11] (SVM), standard One-Class SVM (OC-SVM), One-Class SVM with 3 features selected in advance (OC-SVM-3), and GAN based method [18] (GAN).

| Methods     | $F_1$-measure | train-t | test-t |
|-------------|---------------|---------|--------|
| SVM         | 0.8318        | 37.9    | 1.4    |
| OC-SVM      | 0.0078        | 940.5   | 257.5  |
| OC-SVM-3    | 0.9691        | 9.9     | 1.8    |
| GAN         | 0.9247        | 1223    | 7291   |
| ALI         | 0.9602        | 980.3   | 6.8    |

The settings of SVM based method refers to the work of H SAXENA [11], the setting of GAN based method refers to the work of T SCHLEGL [18]. The only difference between OC-SVM and OC-SVM-3 is that OC-SCM-3 use 3 features selected in advance based on experience instead of all origin features.

We implement the above models with TensorFlow and Scikit-Learn.

Table 2 gives $F_1$-measure, train-t, and test-t (in seconds) of these five methods on our dataset.

Figure 5 shows the confusion matrix of the detection results of ALI method in the KDD test dataset. Among the total of 24,701 test data, 243,107 data is classified correctly.
We can draw from Table 2 that our method gets higher $F_1$-measure, $train-t$ and $test-t$ than the other 4 methods. It is worth noting that the test time of our method is 1072 times shorter than GAN based method. The accuracy and efficiency of SVM based method are high relatively, but the fatal disadvantage of this method is its supervised characteristic, which makes it hard to recognize unknown attacks. The accuracy of OC-SVM-3 is higher than OC-SVM, but the pre-selected features depend on manual labor, which makes its generalization ability is limited. Based on above analysis, our method has better performance on accuracy and efficiency than existing methods, with the ability to detect unknown attacks and without manual feature selection, and our method shows better practical application ability.

5. Conclusions
Based on the investigation of background, significance and related works of abnormal traffic detection, we propose a new encoding method based on PMF towards categorical features, we also propose a new detection method based on ALI. We evaluate our method on KDD-99 dataset. The experiment results show that our method has the advantages of high detection accuracy and high training efficiency, and has the ability to detect unknown attacks. In the future, we will continue to pay attention to latest generative models and to study their application in the cyberspace security.

Acknowledgments
We would like to express our gratitude to the corresponding author of this paper, Yang Wentai, a respectable, responsible and resourceful scholar, who has provided us with valuable advice in every stage of the writing of this article.

This work is partially supported by Sichuan Science and Technology Department Application Foundation Project (2018JY0193,2018GZDZX0010), the Sichuan Provincial Department of Education focuses on funding research projects (61370065, 61502040), Chengdu Chengdian Power Engineering Design Co., Ltd. and Sichuan University Cooperation Project (17H0781).

References
[1] Tang LJ, Li HY. Survey of Network Anomaly Traffic Detection Based on Large Data[J]. Computer Knowledge and Technology, 2017,13:19-21
[2] Qian YK, Chen M, Ye LX, Liu FR, Zhu SW, Zhang H. Network-Wide anomaly detection method based on multiscale principal component analysis. Journal of Software, 2012,23(2):361–377

[3] LU W, GHORBANI A A. Network anomaly detection based on wavelet analysis[J]. EURASIP J. Adv. Signal Process, 2009, 2009: 1-16.

[4] MOAYEDI H Z, MASNADI-SHIRAZI M. Arima model for network traffic prediction and anomaly detection[C]. Information Technology, 2008. ITSim 2008. International Symposium on, 2008 : 1-6.

[5] SHYU M-L, CHEN S-C, SARINNAPAKORN K, et al. A novel anomaly detection scheme based on principal component classifier[R]. MIAMI UNIV CORAL GABLES FL DEPT OF ELECTRICAL AND COMPUTER ENGINEERING, 2003.

[6] LAKHINA A, CROVELLA M, DIOT C. Mining anomalies using traffic feature distributions[C]. ACM SIGCOMM Computer Communication Review, 2005 : 217-228.

[7] Zheng LM, Zou P, Han WH, Li AP, J Y. Traffic anomaly detection using multi-dimensional entropy classification in backbone network[J]. Journal of Computer Research and Development, 2012,49(9):1972-1981

[8] PORTNOY L, ESKIN E, STOLFO S. Intrusion detection with unlabeled data using clustering[C]. In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001, 2001 : .

[9] Zuo J, Chen ZM. Anomaly detection algorithm based on improved K-means clustering[J]. Computer Science, 2016,43(8):258-261

[10] Zhu YW, Yang JM, Zhang JX. Anomaly detection based on traffic information structure[J]. Journal of Software, 2010,21(10):2573-2583

[11] SAXENA H, RICHIARIYA V. Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain[J]. International Journal of Computer Applications, 2014, 98(6).

[12] Li Q, Yan CH, Zhu Y. Analysis and detection of network traffic anomaly based on decision tree[J]. Computer Engineering, 2012,38(5):92-95

[13] YANG J, DENG J, LI S, et al. Improved traffic detection with support vector machine based on restricted Boltzmann machine[J]. Soft Computing, 2017, 21(11): 3101-3112.

[14] LIU F, LI Y, XIA F, et al. A Method of APT Attack Detection Based on DBN-SVDD[J], 2017, : .

[15] FIORE U, PALMIERI F, CASTIGLIONE A, et al. Network anomaly detection with the restricted Boltzmann machine[J]. Neurocomputing, 2013, 122: 13-23.

[16] WULSIN D, BLANCO J, MANI R, et al. Semi-supervised anomaly detection for EEG waveforms using deep belief network[J]. IEEE Transactions on Medical Imaging, 2017 : 1-18.

[17] SCHLEG T, SEEBOCK P, WALDSTEIN S M, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery[J]. Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on, 2010 : 436-441.

[18] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special Lecture on IE, 2015, 2: 1-18.

[19] FRIEDMAN I. On bias, variance, 0/1—loss, and the curse of dimensionality[J]. Data mining and knowledge discovery, 1997, 1(1): 55-77.

[20] DUMOULIN V, BELGHAZI I, POOLE B, et al. Adversarially learned inference[J]. arXiv preprint arXiv:1606.00704, 2016, : .

[21] ZENATTI H, FOO C S, LECOUAT B, et al. Efficient GAN-based anomaly detection[J]. arXiv preprint arXiv:1802.06222, 2018, : .