Mutual Information-based Intrusion Detection Model for Industrial Internet

Rui-Hong Dong, Dong-Fang Wu, Qiu-Yu Zhang and Hong-Xiang Duan
(Corresponding author: Qiu-Yu Zhang)

School of Computer and Communication, Lanzhou University of Technology
No. 287, Lan-Gong-Ping Road, Lanzhou 730050, China
(Email: zhangqylz@163.com)
(Received May. 25, 2017; revised and accepted August 26, 2017)

Abstract

High dimension, redundancy attributes and high computing cost issues usually exist in the industrial Internet intrusion detection field. For solving these problems, the mutual information-based intrusion detection model for industrial Internet was proposed. Firstly, by using features selection method based on mutual information, the attributes set was reduced and traffic characteristics vector was established. Secondly, the normal and abnormal traffic characteristics maps were obtained via the traffic characteristics map technology based on multi correlation analysis. Finally, with the using of discrete cosine transform and nonnegative matrix factorization, we can produce normal and abnormal hash digest, which were used to produce intrusion detection rules. To verify the effectiveness of this model, we adopt NSL-KDD data as the experimental data. The experimental results show that, by using the features selection approach based on mutual information, the proposed model has good classification accuracy and gets good detection performance.

Keywords: Intrusion Detection; Mutual Information; NSL-KDD Data; Perceptual Hash; Traffic Characteristics Map

1 Introduction

With the appearances of Industrial Internet and Industrial 4.0, industrial control area attaches more and more attentions from researchers [25]. Nowadays, big data, cloud computing and Internet of things have been the kernel techniques for the national critical infrastructures [13]. Furthermore, industrial control network security gradually transformed into the Industrial Internet security [26]. Looking through these major security issues happened in USA, Israel, German and Poland in 2016, it is conclude that network intrusion in industrial control area is powerful destructive and complex, which can launch attacks widely [4].

Aiming at the security problems in Industrial Internet, the existing solutions mainly have two kinds. One is to build a passive defense line via firewall, information encryption and user authentication. Another is to establish an active defense line by the intrusion detection method or system. In the research of Industrial Internet intrusion detection problems, researchers adopt classification, clustering, information theories and mathematical statistics four classes of approaches to deal with intrusion detection problems [12, 19]. The recent approaches include Naive Bayes [6, 15], Bayes Tree [7], Hidden Bayes [7], SVM [16, 17], least square SVM [2], artificial neural network [10], neural network with random weight [3], accelerated deep neural networks [18], artificial immune [8].

Meanwhile, many researchers made great contributions in the major of standard experimental data set and preprocessing works. Many researches choose NSL-KDD [24] standard experimental data to validate the detection performance. What's more, features selection method, as the pre-processing operations, can be used in features selection and dimension reduction of data. By this way, the computing cost and time cost of intrusion detection can be reduced. And the detection performance can also be improved obviously. The recent features selection methods include correlation based on features selection method [6], linear discriminant analysis [15], fast correlation based filter [7], mutual information [2], rough set, decision-theoretic rough set [8], information gain [14], gain ratio [14].

In the Industrial Internet intrusion detection issues, there exist high dimension of data, redundancy attributes, high computing cost problems. Aiming at these above problems, the research works about intrusion detection in Industrial Internet and features selection methods were finished in this paper. Comparing the classification accuracy among information entropy, information gain, decision-theoretic rough set and mutual information four methods, the features selection method based on mutual information got the highest accuracy. In the perceptual hash intrusion detection for Industrial Internet, the dy-

namic feedback mechanism was added. And the NSL-KDD standard experimental data was used in the validation experiments.

The rest of the paper is organized as follows. In Section 2, the features selection method, standard data set and image perceptual hash features extraction approach three aspects are introduced in the related works. The problem statement and preliminaries parts including the research problems and related theories are illustrated in Section 3. In Section 4, we present the intrusion detection model based on mutual information for industrial Internet. The results and performances of the proposed model are analyzed in Section 5. Finally, we conclude our paper in Section 6.

2 Related Works

The related works are mainly about features selection method, standard experimental data and image perceptual hash features extraction approaches. The amount of experimental data is continually increasing, which directly affect the detection performance, efficiency, and accuracy of IDS. Considering these issues, features selection method is adopted in the pre-processing operations.

In [6], the supervised classification algorithm based on Naive Bayes was used to establish intrusion detection system. By using correlation based on features selection method (CFS), the correlation between different attributes and the relation between attributes and class were analyzed. Therefore, the redundant attributes and irrelevant attributes were removed. In [15], two-tier network intrusion detection model based on machine learning was introduced. Adopting Naive Bayes and K nearest neighbor method, the intrusion detection classifier was established. The linear discriminant analysis (LDA) was utilized to achieve features selection missions. This method got high detection rate for U2R and R2L two kinds of attacks. In [7], intrusion detection classifiers were established by machine learning algorithms and preprocessing operations. By using fast correlation based filter (FCBF), the vital attributes were selected. The research utilized Naive Bayes, Hidden Naive Bayes and Naive Bayes Tree to classify the normal and abnormal traffic records. The high computing payload of FCBF decreased the efficiency of intrusion detection. In [16], the intrusion detection method based on SVM was researched. The redundant attributes were deleted via filter method. And the vital attributes were selected to build attributes set, which obviously decreased the computing cost of IDS. However, the setting of threshold in filter and classification accuracy still needs optimization. In [2], the least square support vector machine (LSSVM) was used to establish intrusion detection system. The experimental results show that, this method obtained high detection rate and low computing cost. In [14], adopting embedded filters to build intrusion detection model in cloud platform, information gain, gain ratio, Chi-square and feature weighting algorithm were used to select vital attributes. Yet, the structure of features selection method is complex and with high computing payload. In [21], the correlation-based feature selection for intrusion detection system was presented. Introducing the correlation-based feature selection matrices and symmetrical uncertain matrices, the correlation between attributes and classes were analyzed. And different classification approaches were used to validate the accuracy of features selection method. In [11], the embedded SVM and non-linear projection techniques were used to achieve the classification and detection of abnormal intrusions. With the help of linear and non-linear dimension reduction methods, 5 kinds classifiers were produced. The research chose NSL-KDD data to test the detection performance of the proposed method.

The NSL-KDD data is an improvement of KDD Cup 99 data. In many recent researches, the NSL-KDD data is used to validate the performance of the proposed method or model. And it is widely used in the research of Industrial Internet intrusion detection problems.

In [10], the artificial neural network (ANN) was used to analyze the performance of NSL-KDD data. And the dimension reduction is obvious after the information entropy, information gain and correlation analysis features selection operations. In [3], the neural network with random weights (NNRw), a semi-supervised learning algorithm, was proposed. The non-iterative neural network model was trained by the randomize method. By using fuzzy variable, the unsigned record can be classified. This method had better learning ability and computing efficiency, but the classification accuracy need to be improved. In [18], the intrusion detection system was established via deep neural network (DNN). The DNN has forward transmission and back forward transmission features. By this way, large amount of data characteristics obtained from training data were used to establish intrusion detection model and classifiers. However, the proposed approach highly depended on the hardware in the platform and the computing cost is high. In [5], a varying chaos particle swarm optimization approach (TVCPSO) was presented. And the intrusion detection model based on SVM was proposed. The chaos particle, time varying inertia coefficient and time varying acceleration coefficient three conceptions were introduced. The local optimum solutions problems of particle swarm algorithm were effectively resolved. And, the weight object function was utilized to balance the max true positive rate and the min false positive rate. In [20], utilizing four frequently used classification methods in the NSL-KDD data, the imbalance problem of the data was analyzed. The experimental results show that the NSL-KDD data, as the standard test data, can be used to validate the detection performance of intrusion detection for Industrial Internet. In [22], an intrusion detection system based on neural network was proposed. Adopting the feed forward and reverse methods, combining with optimal technique, the computing payload of intrusion detection method was decreased. The NSL-KDD data was chosen as the exper-

imental data. In [1], using a colony optimization method, the effective and key features were selected, which improved the performance of intrusion detection system.

The intrusion detection method for Industrial Internet based on perceptual hash is a new and effective method from the point of image [8]. The recent image perceptual hash features extraction method includes discrete cosine transform (DCT), discrete wavelet transform (DWT), singular value decomposition (SVD), non-negative matrix factorization (NMF) and local binary pattern (LBP). The intrusion detection method based on perceptual hash has robustness and discrimination, which maintains the detection performance. It is proofed that the proposed method is short time consuming. The NMF method needs less storage and it sensitive to the local features in the traffic characteristics map. Therefore, by using DCT and NMF methods, hash digest are produced.

3 Problem Statement and Preliminaries

The proposed model includes three parts: features selection method based on mutual information, traffic characteristics map technique and improved image perceptual hash intrusion detection method.

3.1 Features Selection Method Based on Mutual Information

The attributes redundancy and attributes irrelevant issues existing in intrusion detection lead to the low classification accuracy, high computing cost and time consuming problems. Taking into account of these problems, the feature selection method based on mutual information [9] is adopted to select vital attributes set and reduce the high dimension of data. The produced attributes set is the input data to the traffic characteristics map technique. And the information entropy and decision-theoretic rough set are forward features selection approaches. These methods have low computing cost but without considering the relationships between each attribute. In the feature selection method based on mutual information, the correlation and redundancy concepts are used to describe the correlations between attributes.

Assume that the total number of the experimental data is N. Every traffic record includes m attributes described as $\{f_1, f_2, \ldots, f_m\}$. $P(f_t)$ is the corresponding probability when f_t get different values. And the information entropy can be defined as

$$H(f_t) = -\sum_{f_t} P(f_t) \log P(f_t),$$

where $H(f_t)$ is the information entropy of attribute f_t, $0 \leq P(f_t) < 1$. When the value of f_t is known, the uncertainty of attribute f_t can be described with condition information entropy, which can be defined as

$$H(\frac{f_t}{f_i}) = -\sum_{f_i} P(f_i) \sum_{f_t} P(\frac{f_t}{f_i}) \log P(\frac{f_t}{f_i}),$$

where $H(f_t/f_i)$ expresses the condition entropy of f_t with the condition of f_i. $P(f_t/f_i)$ is the condition probability of the corresponding attribute, $0 \leq P(f_t/f_i) < 1$. According to information entropy and condition entropy concepts, the mutual information can be defined as

$$I(f_t; f_i) = I(f_i; f_t) = H(f_t) - H(\frac{f_t}{f_i}),$$

where $I(f_t; f_i)$ is the mutual information value. And $I(f_t; f_i)$ equal to $I(f_i; f_t)$. The average mutual information is the mean value of mutual information between attribute f_i and every possible attribute f_t, $t \in [1, 41]$. The average mutual information can be defined as

$$ave\_MI(f_i) = \frac{1}{m} \sum_{t=1}^{m} I(f_i; f_t), \tag{1}$$

where $ave\_MI(f_i)$ is the average mutual information of attribute f_i.

Definition 1. *(Correlation Degree) The correlation degree of attribute f_i is the average mutual information of f_i. The correlation degree can be defined as*

$$Rel(f_i) = \frac{1}{m} \sum_{t=1}^{m} I(f_i; f_t),$$

where $Rel(f_i)$ is the correlation degree of f_i, it is average mutual information, $m = 41$.

Definition 2. *(Condition Correlation Degrees) In the condition of attribute f_i, condition correlation degrees of attribute f_t can be defined as*

$$Rel(\frac{f_t}{f_i}) = \frac{H(\frac{f_t}{f_i})}{H(f_t)} Rel(f_t),$$

where $Rel(f_t/f_i)$ is the condition correlation degrees of attribute f_t in condition of f_i. $Rel(f_t)$ is the correlation degree of attribute f_t.

Definition 3. *(Redundancy Degrees) The redundancy degrees between attribute f_i and f_t can be expressed as*

$$Red(f_i; f_t) = Rel(f_t) - Rel(\frac{f_t}{f_i}),$$

where $Red(f_i; f_t)$ is the redundancy degrees between attribute f_i and attribute f_t. $Rel(f_t)$ is the average mutual information of f_t, and $Rel(f_t/f_i)$ is the condition correlation degrees of f_t in condition of f_i.

In the features selection method based on mutual information, the significance of attribute can express the importance of the waiting selecting attributes in attributes set $\boldsymbol{U}$. Meanwhile, the most significant attribute can be

added into selected attributes set $\boldsymbol{S}$. The significance of attributes can be expressed as

$$UmRMR(f_i) = Rel(f_i) - \max_{f_t \in S_m - 1}\{Red(f_i; f_t)\}, \quad (2)$$

where $Rel(f_i)$ is the correlation degree of attribute f_i. $Red(f_i; f_t)$ is the redundancy degree between attribute f_i and f_t. And attributed f_t belongs to selected attributes set $\boldsymbol{S}$. Every time, the max significance of attribute f_i can be added into selected set $\boldsymbol{S}$.

Algorithm 1 Features Selection method based on Mutual Information

1: Input: Experimental data train-set and the number of the selecting features $\boldsymbol{K}$
2: Output: The selected features set $\boldsymbol{S}$
3: Initialize the features selected set $\boldsymbol{S} = \emptyset$, store selected attributes.
4: Initialize the features selecting set $\boldsymbol{U} = \{f_1, f_2, \ldots, f_m\}$, $m \in [1, 41]$.
5: According to Equation (1), computing average mutual information of every attribute.
6: Choose the max average mutual information of attribute f_i, add attribute f_i to the selected set $\boldsymbol{S}$, and delete attribute f_i in the selecting set $\boldsymbol{U}$.
7: **while** the number of selected features $< \boldsymbol{K}$ **do**
8: According to Equation (2), compute the significance of every selecting attribute f_i
9: Choose the most vital attribute f_i, add f_i to set $\boldsymbol{S}$ and delete f_i in the set $\boldsymbol{U}$
10: **end while**
11: Output the selected attributes $\boldsymbol{S}$

After the features selection method, the selected features set is the input for the traffic characteristics map technique.

3.2 Traffic Characteristics Map Technique

The traffic characteristics map technique is based on multi correlation analysis (MCA) [23]. By computing the triangle area mapping, the correlation information between attributes in normal and abnormal network traffics. By this way, the text network traffic records with $1 \times m$ vector format can be transformed into $m \times m$ network traffic matrices. And $m = 14$ is the feature selection results. The traffic characteristics map includes correlation between each attributes.

The experimental data is $X = \{x_1, x_2, \ldots, x_n\}$, and according to the selected features set, the $i-$th traffic record is $x_i = [f_1^i, f_2^i, \ldots, f_m^i]$, $(1 \le k \le m)$. The correlation between $j-$th attribute and $k-$th attribute can be computed by triangle area.

The vector x_i can map into the $(j-k)$ two-dimension Euclidean subspace, $y_{i,j,k} = [\varepsilon_j \varepsilon_k]^T = [f_j^i f_k^i]^T$, $(1 \le i \le n, 1 \le j \le m, 1 \le k \le m, j \ne k)$. Variable $\varepsilon_j = [\varepsilon_{j,1}, \varepsilon_{j,2}, \ldots, \varepsilon_{j,n},]^T$, where $e_{j,j} = 1$, $e_{k,k} = 1$, other elements is 0. $y_{i,j,k}$ is two-dimension column vector, which is the point $(f_j^i f_k^i)$ of $(j-k)$ two-dimension Euclidean subspace in Descartes coordinate system. Then, in Descartes coordinate system, connecting the origin with the point f_j^i mapping in j coordinate axis and point f_k^i mapping in k coordinate axis, the triangle area is obtained, named $\Delta f_j^i O f_k^i$. The triangle area is marked as $Tr_{j,k}^i$.

$$Tr_{j,k}^i = (\parallel (f_j^i, 0) - (0,0) \parallel \times \parallel (0, f_k^i) - (0,0) \parallel)/2,$$

where $1 \le i \le n$, $1 \le j \le m$, $1 \le k \le m$, and $j \ne k$. The complete triangle area mapping of a network traffic record includes the triangle area of every pairs of attributes. $Tr_{j,k}^i$ is $j-$th row and $k-$th column. When $j = k$, $Tr_{j,k}^i = 0$. The correlation between different attributes is the key point. The symmetric matrix $\boldsymbol{TAM}$ can be got. For example, the $4-$dimension $\boldsymbol{TAM}$ is shown.

$$TAM_x^i = \begin{bmatrix} 0 & Tr_{1,2}^i & Tr_{1,3}^i & Tr_{1,4}^i \\ Tr_{2,1}^i & 0 & Tr_{2,3}^i & Tr_{2,4}^i \\ Tr_{3,1}^i & Tr_{3,2}^i & 0 & Tr_{3,4}^i \\ Tr_{4,1}^i & Tr_{4,2}^i & Tr_{4,3}^i & 0 \end{bmatrix}$$

3.3 Image Perceptual Hash Intrusion Detection Method

The image perceptual hash features extraction based on DCT and NMF is utilized to produce normal and abnormal hash digests. Meanwhile, after the features selection operation, the selected features is $m = 14$. Therefore, the 14×14 traffic characteristics map is established.

In the preparing stage of perceptual hash features extraction method, DCT coefficient is computed.

$$F(u,v) = \frac{C(u)C(v)}{4} \sum_{x=0}^{m} \sum_{y=0}^{m} f(x,y) \bullet \cos(\frac{\pi(2x+1)u}{m^2}) \cos(\frac{\pi(2y+1)v}{m^2}) \quad (3)$$

where f is $m \times m$ image pixel point and F is $m \times m$ DCT coefficient matrix. $\boldsymbol{C}$ is cosine coefficient matrix.
Algorithm descriptions:

1) After the features selection method, 14×14 pixel traffic characteristics map is obtained by using traffic characteristics map technique.

2) According to Equation (3), 14×14 coefficient matrix can be computed by DCT.

3) In order to extract the local features in the map and get the discriminative perceptual hash digest, the low-frequency region of coefficient matrix is rebuilt by NMF. And the local saliency information of map is extracted. By NMF, the DCT coefficient matrix

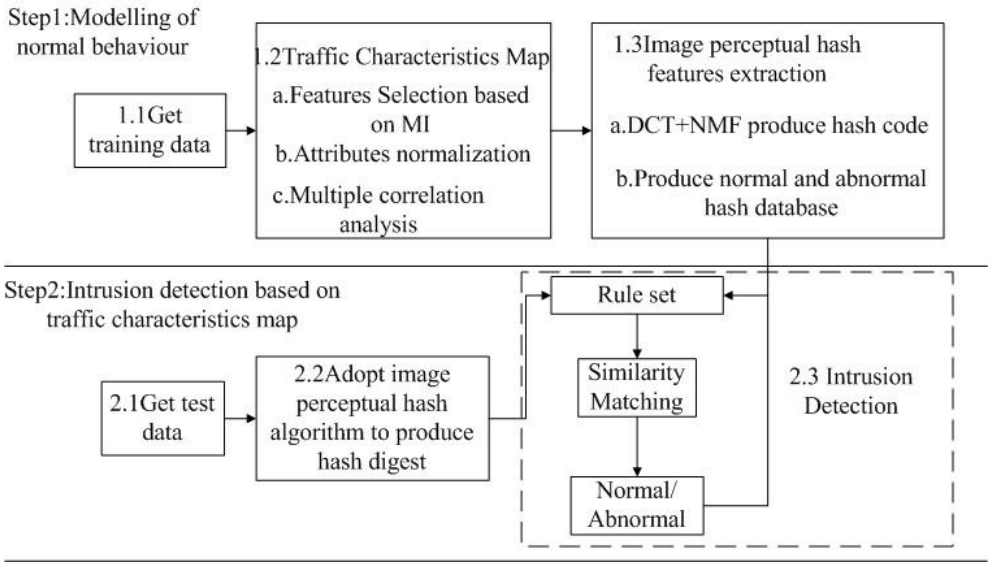

Figure 1: Intrusion Detection model based on mutual information for Industrial Internet

can factorize into basis matrix $\boldsymbol{W}$ and weights coefficient matrix $\boldsymbol{H}$.

$$DCT\_Coefficient = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \begin{bmatrix} h_1 & h_2 & \cdots & h_m \end{bmatrix},$$

where $\boldsymbol{DCT\_Coefficient}$ is the DCT coefficient matrix. $\boldsymbol{W}$ is the basis matrix and $\boldsymbol{H}$ is the weights coefficient matrix. Each column vector sum in $\boldsymbol{W}$ is the column vector of DCT coefficient matrix. The matrix factorization vector $[W'H]$ is established and the mean value of $[W'H]$ is computed.

$$NMF\_matrix = [W'Hmean].$$

4) According to the hash rules, the hash digest can be produced. The hash rule is defined as

$$h(x_{i+1}) = \begin{cases} 1 & x_{i+1} > x_i \\ 0 & x_{i+1} \leq x_i \end{cases}$$

where the $(i+1)$th x_{i+1} is got. When $x_{i+1} > x_i$, the hash code of x_{i+1} is 1, or 0, $1 \leq i \leq 29$. The length of hash code is 28 bit with the binary form. The normal and abnormal hash digest base is established and the corresponding intrusion detection rules are extracted.

In the hash matching phase, the normalization Hamming distance is used to measure the similarity between different hash digest. The normalization Hamming distance is defined as

$$D_H(H_{s1}, H_{s2}) = \frac{1}{L} \sum_{w=1}^{L} \mid H_{s1}(w) - H_{s2}(w) \mid, \qquad (4)$$

where H_{s1} and H_{s2} are two hash digest, whose length are 28 bit. w is one bit in the hash digest. When the hash matching threshold is set, if the hash similarity is above threshold, it is abnormal or normal. The joint threshold is 0.15.

4 The Proposed Method

The flow works of the intrusion detection model based on mutual information for Industrial Internet (IDM-MI) is shown in Figure 1. The method is divided into preparing phase and intrusion detection phase based on Industrial Internet.

As the Figure 1 shown, in the preparing phase, the numerical operation is achieved. The features are reduced via features selection method based on mutual information. Then the normalization operation is finished. By using MCA, the traffic characteristics map is produced. With the using of DCT and NMF, the hash digests are extracted to establish normal and abnormal hash digest base. And the intrusion detection rule is also produced.

In the intrusion detection phase, the numerical operation is finished and the vital features set are extracted. By MCA, the traffic characteristics map of test record is produced. And the hash code is produced via DCT and NMF. Adopting normalization Hamming distance, the similarity between test hash and hash digest base is measured. If the similarity is lower than threshold, it's normal or abnormal.

The scale of the training data is N_1, and the scale of the test data is N_2. The original number of attributes is M. After features selection operation, the number features is M_1. The number abnormal hash digest is t_1 and the abnormal hash digest is t_2. The length of hash digest is L. According to the analysis of the algorithm flows, the

Algorithm 2 IDM-MI

```
Input: The standard NSL-KDD data train-data and test-data
Output: Intrusion detection result
Obtain the train-data
while the number of train-data > 0 do
    Adopt MCA method to produce traffic characteristics map
    Utilize DCT and NMF methods to extract hash digest of normal and abnormal network traffic
    Establish the intrusion detection rule set
    Number −−
end while
Obtain test-data
while the number of test-data > 0 do
    Adopt MCA method to produce traffic characteristics amp
    Using DCT and NMF method to produce hash code
    According to Equation (4), compute the similarity between hash code
    if similarity < threshold then
        The record is normal
    else
        The record is abnormal
    end if
    Number −−
end while
```

time complexity is $O((N_1 + N_2)(M_1^2 + 5M_1))$.

5 Experimental Results and Analysis

5.1 Preparing for the Simulation Experiment Environment

The experiments were carried out by a ThinkPad computer with 2.5 GHz quad-core i5-3210M and 8GB of RAM. The operate system is windows 7, 64bits and the simulation platform is Matlab R2013a.

The NSL-KDD [24] standard data set was chose to validate the performance of the proposed model, which is improved from the KDD Cup 99 data set. To keep the effective evaluation for the intrusion detection methods, the percentage of each kinds of data are same with the KDD Cup 99. Some intrusion detection methods only detect part of the repeated data effectively. Therefore, the redundant records in the training data set and the repeated records in the test data set were removed. In order to decrease the running payload, the number of the data is reasonable. So, the NSL-KDDTrain+_20Persent and NSL-KDDTest-21 were selected to take part in the experiments, which are named as Data1 and Data2 respectively. The details of the experimental data are shown as Table 1.

Firstly, the numerical operations of the training data and test data were finished. The protocol-type, service, flag and attack four attributes experienced numerical operations. Secondly, according to the label of the attack, the training data were classified into normal and abnormal. Label $\{1\}$ is normal and Label $\{2, 3, 4, 5\}$ are abnormal. Finally, the normalization of the training data and test data were achieved. The data is normalized into $[0, 255]$.

Table 1: The details of the experimental data

Name	Total	Normal	Dos	Probe	U2R	R2L
Data1	25192	13449	9234	2289	11	209
Data2	22544	9711	7458	2421	200	2754

$$f(x) = \begin{cases} 0 & x \in [0, \min) \\ \frac{255x}{\max - \min} & x \in [\min, \max] \\ 255 & x \in (\max, \infty) \end{cases},$$

where *max* is the max value and *min* is the min value. $f(x)$ is the normalized value.

5.2 Feature Selection Methods

According to the recent research results, the preprocessing selections of the experimental data were finished. Attributes $\{9, 20, 21\}$ take no effect on the classification. Attributes $\{15, 17, 19, 32, 40\}$ have little influence on the classification. The values of attributes $\{7, 8, 11, 14\}$ are mostly 0. The above mentioned attributes were removed. Then, by feature selection method based on mutual information, the dimension of the data was reduced.

According to the conceptions of the information entropy, condition entropy and mutual information, the average mutual information was computed. By correlation degree, condition correlation degrees and redundancy degree, the significance of the selecting attributes, named max correlation-min redundancy, was obtained. Considering the significance of the selecting attributes, the candidate attributes can be added into the selected attributes set $\boldsymbol{S}$. The number of the selected attributes is set $K = 14$. Therefore, the result of the feature selection is 14 vital attributes.

The result of the feature selection operation is $set = \{3, 5, 6, 8, 12, 23, 24, 32, 33, 34, 35, 37, 38, 39\}$. Table 2 displays several recent features selection approaches and compares their accuracy of the classification.

From Table 2, when the number of the attributes is 14, the classification accuracy of mutual information is 0.9940, which is the max value. So, the mutual information features selection method is used to reduce the dimension of data. However, the time consuming and computing cost of this method are still need to be optimal. It's difficult to balance the time cost and classification accuracy, which attach more attentions from the researchers. In the next work, this problem is still the research emphasis.

Table 2: The details of the experimental data

Method	Features selection results	Accuracy
Information Entropy	3,5,6,23,24,29,30,31,32,33,34,35,36,37	0.9587
Information Gain	3,4,5,6,12,23,25,29,30,33,34,35,38,39	0.9744
Decision-Theoretic Rough Set	2,3,4,5,6,8,12,17,27,28,30,31,36,37	0.9865
Mutual Information	3,5,6,8,12,23,24,32,33,34,35,37,38,39	0.9940

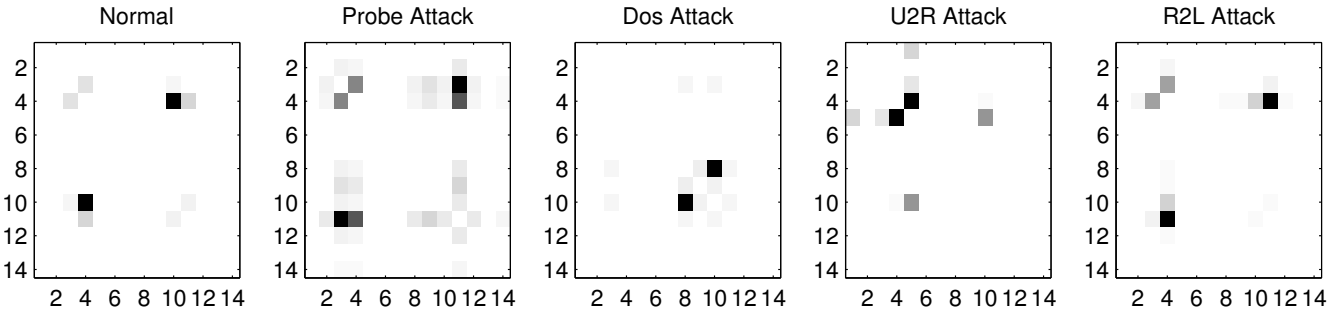

Figure 2: The traffic characteristics map of NSL-KDD training data set

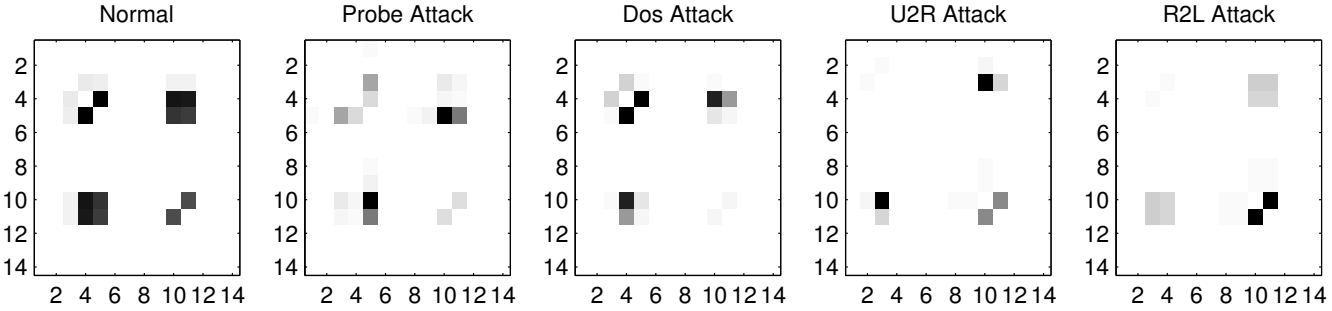

Figure 3: The traffic characteristics map of the NSL-KDD test data set

By utilizing image perceptual hash features extraction method, the hash code of the normal and abnormal network traffic records are produced. The length of the hash code is 28 bits. The number of the normal rule is 471 and the number of the abnormal rule is 535.

5.3 Traffic Characteristics Map

After the features selection operations, the features space is reduced. Then, the traffic characteristics map is produced via traffic characteristics technique, as the Figure 2 and 3 shown. The size of the selected features is 14, and the traffic characteristics map is 14×14 matrix.

In Figure 2, 5 kinds of records in the training data are shown. According to the results of the features selection method, the size of the map is 14×14. The difference between every map is obvious. These maps are the input data for the next operation.

As the Figure 3 shown, 5 kinds of record in test data are produced. Comparing Figure 2 with Figure 3, the features of the image is obvious. The discrimination is good. In the 14×14 area, the almost same place exist some pixel blocks which have different grey value.

5.4 Discriminative Experiments

The intrusion detection method based on mutual information for industrial Internet has robustness and discrimination which keep the performance and efficiency of the intrusion detection. The robustness ensures that the same normal and abnormal records produce same hash code. And the discrimination ensures that the different records produce different hash code. With the help of robustness and discrimination, the proposed model can detect the existing records or new records. Therefore, this model has adaptability. The false accept ratio (FAR) is selected in the discriminative experiments. The FAR can be defined as

$$FAR = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\tau} \exp[\frac{-(x-\mu)_2}{2\sigma^2}],$$

where μ is the exception mean of the normal distribution. σ is the standard deviation and τ is the matching threshold. Figure 4 is the normal distribution figure of this mode.

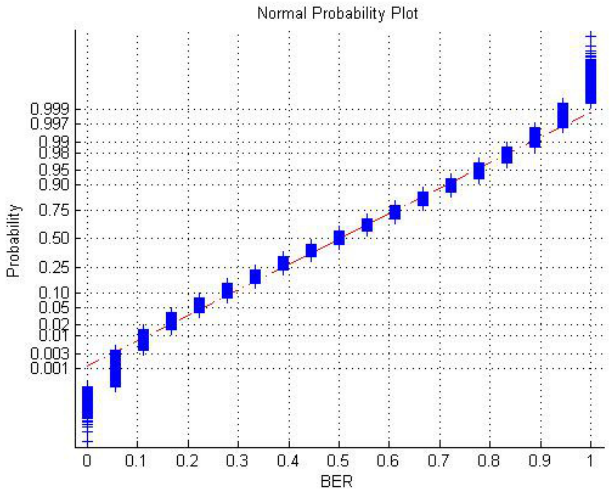

Figure 4: Normplot figure of the proposed model

The total number of the hash code in training data is 1006. So, 505515 bit error ratio (BER) can be obtained. Figure 4 is the normal distribution curve of the BER. In Figure 4, the curve is almost overlapping with the straight line of mean value. But the fluctuations also exist in the two sides of the curve. The mean value is 0.4951 and the standard deviation is 0.0945. The real standard deviation is 0.1724.

When $\tau = 0$, $FAR = 0.0020$. That is to say that, when the matching threshold is $\tau = 0$, in 1000 traffic records, there are 2 false detections, which meet the requirements of detection. The threshold of FAR is shown in Table 3.

Table 3: FAR comparing table

Threshold τ	0	0.005	0.01	0.015
FAR	0.0020	0.0022	0.0024	0.0027

As the Figure 5 shown, the hash matching of normalization Hamming distance obey to Gaussian distribution, the mean value $\mu = 0.5$, standard deviation $\mu = 0.5/\sqrt{N}$. And N is the length of hash code, $N = 28$. Figure 5 is the BER histogram obtained from the discriminative experiments. The center of the histogram is 0.4951 which is close to 0.5. The standard deviation of distribution is 0.1724.

5.5 Experiments Results and Analysis

We choose TP, FP, TN, FN and Acc as the evaluation indexes. The TP is the percentage of abnormal records correct classification in all abnormal records. The number

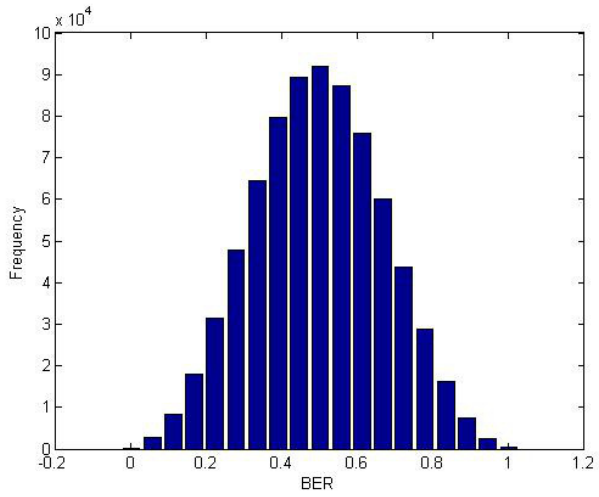

Figure 5: The BER predict histogram of image perceptual hash features extraction method

of abnormal record correct classification is $num1$, and the total number of abnormal records is N. The TP can be defined as

$$TP = \frac{num1}{N},$$

where the FN is the percentage of abnormal records false classification in all abnormal records. The number of abnormal record false classification is $num2$. The FN can be expressed as

$$FN = \frac{num2}{N},$$

where $TP + FN = 1$. The FP is the percentage of normal records false classification in all normal records. The number of normal record false classification is $num3$, and the total number of normal records is M. The FP can be defined as

$$FP = \frac{num3}{M},$$

where the TN is the percentage of normal records correct classification in all normal records. The number of normal record correct classification is $num4$. The TN can be expressed as

$$TN = \frac{num4}{M},$$

where $FP + TN = 1$. The Acc express the average detection ratio of normal and abnormal records. It is also the ratio of correct predicted records to the entire records. The Acc can be defined as

$$Acc = \frac{TP + TN}{TP + FN + FP + TN}.$$

Table 4 displays the difference with different methods in these indexes.

As Table 4 shown, the *Acc* of the proposed method is 0.9940, which is less than 0.9985 of NB Tree method. And the *FP* of the proposed method is 0.0012 which also less than 0.0020 of NB Tree approach. The *TP* is 0.9893 which is less than 0.9990 of NB Tree and 0.9916 of ANN. But, the *FP* of our method is less than 0.0036 of ANN. The *Acc* of our method is equal to ANN. The features selection method based on mutual information has better performance and good results. The perceptual hash method has a better detection efficiency and short time consuming. Therefore, the IDM-MI has a better detection performance.

Table 4: The details of the experimental data

Method	Reference	TP	FP	Acc
LSSVM-IDS	Ref. [2]	0.9893	0.0028	0.9932
ANN	Ref. [22]	0.9916	0.0036	0.9940
TVCPSO-SVM	Ref. [5]	0.9703	0.0087	0.9808
NB Tree	Ref. [6]	0.9990	0.0020	0.9985
Naive Bayes	Ref. [6]	0.9360	0.1340	0.9010
AD Tree	Ref. [6]	0.9890	0.0190	0.9850
FCBF	Ref. [7]	-	-	0.8704
SVC	Ref. [11]	0.9340	0.1400	0.8970
SVM	Ref. [16]	0.8200	0.1500	0.8350
IDM-MI	Our method	0.9893	0.0012	0.9940

6 Conclusions

An intrusion detection model based on mutual information for industrial Internet was presented. Adopting features selection method based on mutual information, the issues of high dimension of data, attributes redundancy and high computing cost were effectively resolved. The dynamic feedback mechanism was added into the intrusion detection model based on perceptual hash. When the new normal or abnormal records appearances, the new hash digest was added into the hash digest base. And the corresponding intrusion detection rule was updated. The adaptability of the proposed method was enhanced. The NSL-KDD data set was utilized to validate the efficiency and accuracy of detection. As the experimental results show that the *TP* is 0.9893, the *FP* is 0.0012 and the *Acc* is 0.9940, which proof the good performance of detection. In the future, the algorithm optimization work of our method is vital.

Acknowledgments

This work is supported by the National Natural Science Foundation of China ($No.61363078$), the Natural Science Foundation of Gansu Province of China ($No.1310RJYA004$), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) ($No.201700005$). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] M. H. Aghdam and P. Kabiri, "Feature selection for intrusion detection systemusing ant colony optimization," *International Journal of Network Security*, vol. 18, no. 3, pp. 420–432, 2016.

[2] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Transactions on computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[3] R. A. R. Ashfaq, X. Z. Wang, and J. Z. Huang, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484–497, 2017.

[4] B. Babu, T. Ijyas, P. Muneer, and J. Varghese, "Security issues in scada based industrial control systems," in *2nd International Conference on Anti-Cyber Crimes (ICACC'17)*, pp. 47–51, Abha, Saudi Arabia, Mar. 2017.

[5] S. M. H. Bamakan, H. Wang, Y. Tian, and Y. Shi, "An effective intrusion detection framework based on mclpsvm optimized by time-varying chaos particle swarm optimization," *Neurocomputing*, vol. 199, pp. 90–102, 2016.

[6] D. H. Deshmukh, T. Ghorpade, and P. Padiya, "Intrusion detection system by improved preprocessing methods and naive bayes classifier using NSL-KDD 99 dataset," in *International Conference on Electronics and Communication Systems (ICECS'14)*, pp. 1–7, Coimbatore, India, Feb. 2014.

[7] D. H. Deshmukh, T. Ghorpade, and P. Padiya, "Improving classification using preprocessing and machine learning algorithms on NSL-KDD dataset," in *International Conference on Communication, Information and Computing Technology (ICCICT'15)*, pp. 1–6, Mumbai, India, Jan. 2015.

[8] R. Dong, WD. u, and Q. Zhang, "The integrated artificial immune intrusion detection model based on decision-theoretic rough set," *International Journal of Network Security*, vol. 19, no. 6, pp. 880–888, 2017.

[9] H. Duan, Q. Zhang, and M. Zhang, "Fcbf algorithm based on normalization mutual information for features selection (in Chinese)," *Journal of Huazhong University of Science & Technology (Natural Science Edition)*, vol. 45, no. 1, pp. 52–56, 2017.

[10] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ann," in *International Conference on Signal Processing and Communication Engineering Systems (SPACES'15)*, pp. 92–96, Guntur, India, Jan. 2015.

[11] E. De la Hoz, A. Ortiz, and J. Ortega, "Network anomaly classification by support vector classifiers ensemble and non-linear projection techniques," in *8th International Conference on Hybrid Artificial*

Intelligent Systems (HAIS'13), pp. 103–111, Salamanca, SPAIN, Sept. 2013.

[12] Y. Farhoui, "Design and implementation of an intrusion prevention system," *International Journal of Network Security*, vol. 19, no. 5, pp. 675–683, 2017.

[13] A. Mosa, H. M. El-Bakry, S. M. Abd El-Razek, S. Q. Hasan, "A proposed E-government framework based on cloud service architecture," *International Journal of Electronics and Information Engineering*, vol. 5, no. 2, pp. 93–104, 2016.

[14] O. Osanaiye, H. Cai, K. K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for ddos detection in cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, pp. 1–10, 2016.

[15] H. H. Pajouh, G. H. Dastghaibyfard, and S. Hashemi, "Two-tier network anomaly detection model: a machine learning approach," *Journal of Intelligent Information Systems*, vol. 48, no. 1, pp. 61–74, 2017.

[16] M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing svms," in *8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA'14)*, pp. 1–6, Dhaka, Bangladesh, Dec. 2014.

[17] E. Popoola, A. O. Adewumi, "Efficient feature selection technique for network intrusion detection system using discrete differential evolution and decision," *International Journal of Network Security*, vol. 19, no. 5, pp. 660–669, 2017.

[18] S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced intrusion detection system," in *21st International Conference on Emerging Technologies and Factory Automation (ETFA'16)*, pp. 1–8, Berlin, Germany, Sept. 2016.

[19] Q. S. Qassim, A. M. Zin, and M. J. A. Aziz, "Anomalies classification approach for network-based intrusion detection system," *International Journal of Network Security*, vol. 18, no. 6, pp. 1159–1172, 2016.

[20] S. Rodda and U. S. R. Erothi, "Class imbalance problem in the network intrusion detection systems," in *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT'16)*, pp. 2685–2688, Chennai, India, Mar. 2016.

[21] M. B. Shahbaz, WX. ang, A. Behnad, and J. Samarabandu, "On efficiency enhancement of the correlation-based feature selection for intrusion detection systems," in *7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON'16)*, pp. 1–7, Vancouver, BC, Canada, Oct. 2016.

[22] B. Subba, S. Biswas, and S. Karmakar, "A neural network based system for intrusion detection and attack classification," in *Twenty Second National Conference on Communication (NCC)*, pp. 1–6, Guwahati, India, Mar. 2016.

[23] Z. Tan, A. Jamdagniand, X. He, P. Nanda, and R. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis," *IEEE transactions on parallel and distributed systems*, vol. 25, no. 2, pp. 447–456, 2014.

[24] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A detailed analysis of the KDD cup 99 data set," in *Symposium on Computational Intelligence for Security and Defense Applications (CISDA'09)*, pp. 1–6, Ottawa, ON, Canada, July 2009.

[25] A. Yang, L. Sun, X. Wang, and Z. Shi, "Intrusion detection techniques for industrial control system," *Journal of Computer Research and Development*, vol. 53, no. 9, pp. 2039–2054, 2016.

[26] B. B. Zarpelao, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in i nternet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017.

Biography

Dong Rui-hong Vice researcher, worked at school of computer and communication in Lanzhou university of technology. His research interests include network and information security, information hiding and steganalysis analysis, computer network.

Wu Dongfang In 2015, Wu Dongfang obtained his bachelor of engineering degree from Northwest University for Nationalities. Currently, he is studying for his masters degree at Lanzhou University of Technology. His research focuses on the industrial control network security.

Zhang Qiu-yu Researcher/PhD supervisor, graduated from Gansu University of Technology in 1986, and then worked at school of computer and communication in Lanzhou University of Technology. He is vice dean of Gansu manufacturing information engineering research center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, multimedia communication technology.

Duan Hong-xiang lecture/PhD student of School of Computer and Communication in Lanzhou University of Technology, she received M.Sc. degree in Lanzhou University of Technology in 2011. Now she is working on feature selection of high-dimensional data in multimodal human-computer interaction and image understanding and pattern recognition.