A novel selection method of network intrusion optimal route detection based on naive Bayesian

Yu Nuo
Experimental Training Center,
Xi’an University of Finance and Economic,
Xi’an 710061, China
Email: yunuoyun@163.com

Abstract: In order to improve the network security performance and resist the increasingly complex and diversified network intrusion, and reduce the false alarm rate of network intrusion and improve the detection efficiency, this paper proposes the selection method of the network intrusion optimal route detection based on naive Bayesian. We selected the feature subset of network route data by the principal component analysis and accordingly processed the network route detection sample set, getting the input characteristics of network route detection. The research selected the new low dimensional feature of network route data through linear or nonlinear transformation, and used the naive Bayesian network structure to classify the new network route data set. Simulation results show that the proposed method can improve the detection rate of network intrusion optimal route and reduce the false alarm rate, getting a more perfect result of network intrusion detection.

Keywords: network intrusion detection; principal component analysis; PCA; normalisation; optimal route of network intrusion.


Biographical notes: Yu Nuo is engaged with the Experimental Training Center at Xi’an Institute of Finance and Economics. He received his Master’s degree while he was also an engineer, His main field of study is e-commerce and network security.

1 Introduction

With the continuous development of the computer and the increasing of network users, the network is moving to people's lives. At the same time, the network security risk factor is constantly increasing (Bian et al., 2015; Ambard et al., 2016; Kros et al., 2016). As a major means of security prevention, the firewall has been unable to meet the needs of network security. Network intrusion detection technology is a kind of security measures which can find out the complete, confidential and available hazardous information resources. The network attack form gradually shows the complexity and diversity (Korcynski et al., 2016; Fossaceca et al., 2015; Vasudevan et al., 2016; Blum et al., 2015). As a method of protecting network security performance (Aziz et al., 2016; Saravanan and Senthilkumar, 2015), the network intrusion detection method can detect
the malicious intrusion and alarm, and notifies the network to take appropriate measures for responding. In order to protect the internal attacks and external attacks and misuse of network, and intercept the corresponding intrusion before the network is endangered, we need to research on selection method of network intrusion detection optimal route. At present, the selecting method of network intrusion optimal route detection based on feature selection uses different discretisation and feature selection algorithms to generate several subsets of network routes with different characteristics, and classifies the alone network route feature subset, on the basis of improving classification algorithm, and learning and modelling for the feature of network route after extracting, we select the optimal route of network intrusion detection. In the process of network intrusion detection, there are some problems such as low accuracy and high false positive rate of detection attacks (Xie et al., 2015; Kang and Kim, 2016; Bulbul et al., 2015; Wang and Zhang, 2015. Therefore, this has triggered the research and discussion of many domestic and overseas scholars. The research results of the optimal route selection in network intrusion detection have been obtained at the same time.

The reference Sang (2016), Ma (2016) and Li (2013) proposes the selection method of network intrusion optimal route detection based on feature selection and sample selection. Firstly, we extracted the feature of network route data, and then made it normalised, and then using the network kernel principal component analysis (PCA) to select feature of network route data and network route samples. Finally we used the extreme learning machine to set up a network route detection classifier, selecting the optimal route of network intrusion detection. In the process of the actual network intrusion detection, this method caused the network intrusion detection model to have the problems of low accuracy and long training time due to the interference of network data redundancy and noise characteristics. The reference Chen and Guo (2016) proposed the selection method of the network intrusion detection optimal route based on support vector machine. Firstly we used the kernel PCA method to extract the network route features, discretising the network features to get the new network route features, using support vector machine classifier to get some better network route. Finally we selected the best route through the network intrusion detection technology.

In the process of network intrusion detection, the detection efficiency and accuracy of this method was not high due to the fluctuation of the network route characteristic parameters. The reference Liu and Wang (2016) proposed a method of selecting the optimal route of network intrusion detection based on the sensor. Firstly the sensor collected the network route data and stored it in correlative network database. We carried out the feature selection and data conversion for network route detection data, which was integrated into the relational database of network route, using the network route classification module to classify network route input data, and selected the optimal route. This method had the good classification effect, but had the low detection accuracy and high false alarm rate.

According to the problems of the above methods, we proposes the selection method of the network intrusion optimal route detection based on naive Bayesian. Our paper firstly extracted the original data characteristics of network route by the PCA of network route detection, after normalised processing, the research selected the new low dimensional feature of network route data through linear or nonlinear transformation, and used the naive Bayesian network structure to classify the new network route data set. According to the classification results we selected the optimal route network. Experimental simulation shows that the proposed method can improve the detection rate
of network intrusion optimal route and reduce the false alarm rate, getting a more perfect result of network intrusion detection.

2 Research network intrusion optimal route detection method

2.1 Feature selection for network route detection data

In the process of the network route detection data feature selection, firstly the method select the network intrusion detection data subset through the PCA theory. The network route detection sample sets are normalised to get the input feature of network route detection. The method performs linear/nonlinear transformations on input features selecting the low dimensional feature of new network route detection data. The network route detection feature selection framework is shown in Figure 1.

![Figure 1 Frame diagram of network route detection](image)

Supposing that the network intrusion route detection input space sample set is \( X = (x_1, x_2, \ldots, x_N) \), \( \Phi \) says the mapping \( X^d \rightarrow F \) from the input space of network intrusion detection route to high dimensional feature space, \( N \) says the low dimensional feature space of network intrusion detection route, then the covariance matrix of network intrusion route detection is:
Based on the spatial mapping relationship of the formula (1), we construct the kernel function of network intrusion route detection features selection:

\[ K = k(x, y) = \langle \phi(x), \phi(y) \rangle = \phi(x)^T \phi(y) \] (2)

Among them, \( k(x, y) \) represents the kernel function of the network intrusion route detection space. \( \phi(x) \) represents the kernel function of the high dimensional space of the network intrusion route detection. \( \phi(y) \) represents the kernel function of the low dimensional space of the network intrusion route detection. We solve the characteristic equation by calculating the kernel function of network intrusion route detection feature:

\[ Cv = \lambda \] (3)

In formulas, \( l \) represents the mean vector of the network intrusion route detection sample. \( v \) and \( \lambda \) represents the corresponding network intrusion route detection feature vector and eigenvalue of \( C \).

Supposing that \( \alpha \) represents the input weights of the network intrusion route detection samples, and \( i \) represents the output weights of the network intrusion route detection samples. The calculation equation of the feature vector of the network intrusion route detection is as follows:

\[ v = \sum_{i=1}^{N} \alpha_i \phi(x_i) \] (4)

The above formula (1) and (2) are brought into the formula (3) and (4):

\[ N\lambda \alpha = K\alpha \] (5)

The feature vector and eigenvalue of the network intrusion route detection are obtained by solving formula (5). Then the projection of the random network intrusion route detection sample on the feature space is:

\[ (v, \phi(x)) = \sum_{i=1}^{N} \alpha_i \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^{N} \alpha_i k(x_i, x) \] (6)

PCA theory can be used to select the feature subset of network intrusion route detection effectively, and the network intrusion detection sample set is processed accordingly to get the input feature of network intrusion route detection. According to the input feature of the network intrusion route detection, the noise-signal ratio \( (d) \) is defined:

\[ d = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2} \] (7)

In formulas, \( \mu_1 \) and \( \mu_2 \) respectively represent the feature mean of the network intrusion route detection. \( \sigma_1 \) and \( \sigma_2 \) respectively represent the standard deviation of the network intrusion route detection feature. The calculation of Bhattacharyya distance between different network intrusion detection features needs to consider the mean value in the route detection samples, also needs to consider the sample variance distribution, that is:
A novel selection method of network intrusion optimal route detection

\[
B = \frac{1}{4} \left( \frac{\mu_1 - \mu_2}{\sigma_1^2 + \sigma_2^2} \right) + \frac{1}{2} \ln \left( \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1\sigma_2} \right)
\]

(8)

Through the calculation of the formula (8), we can select the network intrusion route detection samples, the specific process is as follows.

The PCA of network intrusion detection route sample linear feature selection in Definition 1 gives the network intrusion route detection sample input data matrix \(X_{N \times n}\) (usually \(N < n\)), network intrusion detection route input data matrix usually consists of some centralised sample data \(\{x_i\}_{i=1}^n\).

Suppose \(x_i \in \mathbb{R}^n\), and \(\sum_{i=1}^n x_i = 0\) the network intrusion route detection input data vector \(x_i\) is transformed into a new network intrusion route detection input data vector \(s_i\):

\[
s_i = U^T x_i
\]

(9)

In formulas, \(U\) is the \(N \times N\) orthogonal matrix of the network intrusion route detection model, \(u_i\) in its \(i^{th}\) column is the covariance matrix of the network intrusion detection data sample:

\[
C = \frac{1}{n} \sum_{i=1}^n x_i x_i^T
\]

(10)

The formula (10) is the \(i^{th}\) network intrusion route detection intrinsic quality. \(T\) represents the feature vector of the network intrusion route detection sample. The calculation formula of network intrusion route detection PCA intrinsic quality:

\[
\lambda_i u_i = C u_i, \quad i = 1, 2, \ldots, N
\]

(11)

In formulas, \(\lambda_i\) represents the eigenvalue of the network intrusion route detection by the PCA of \(C\). \(u_i\) represents the corresponding network intrusion detection eigenvector. When using the previous \(p\) network routes to detect the PCA intrinsic vector \(U\) (corresponding to the descending order of the network intrusion route detection eigenvalue \(\lambda_i\)), and the score matrix is obtained:

\[
S = \{s_i\}_{i=1}^n
\]

(12)

\[
X = \{x_i\}_{i=1}^n
\]

(13)

The orthogonal transformation of the formula (12) and (13) is as follows:

\[
S = U^T X
\]

(14)

In the formula, a new component \(S\) is called principal component, analysis, when using previous several eigenvector of PCA, the number of principal components in \(S\) will be reduced. Then the PCA method achieves the effect of reducing the dimensions for the input data network intrusion route detection.

The principal component of PCA is used for the network intrusion route detection, which has the following characteristics:
the row vector of network intrusion route detection vector $S(i)$, $i = 1, 2, \cdots, p$ is
uncorrelated
$S(i)$, $i = 1, 2, \cdots, p$ has the largest variance in order
the foremost principal component based on PCA is used to represent the original
input.

In formula (14), $U^T$ projects $X$ into space $S$, we assume that using the linear regression
equation of PCA for the network intrusion route detection:

$$Y = SB + \varepsilon$$  \hspace{1cm} (15)

Among them, $Y$ represents the corresponding variable matrix of network intrusion route
detection; $B$ represents the coefficient matrix of network intrusion route detection; $\varepsilon$
represents the noise matrix of network intrusion route detection matrix. The least squares
solution of the linear regression equation using PCA for network intrusion detection is:

$$B = \left(S^TS\right)^{-1}S^TY$$ \hspace{1cm} (16)

Supposing that $t$ represents the characteristic value of test sample of the network intrusion
route detection. The network intrusion route detection test sample $X_t$ is projected into the
space $S$ to get $S_t$ before the network intrusion detection is applied to the principal
component regression:

$$X_t = \left\{x_{i}^{n+m_{t}}\right\}_{i=n+1}$$ \hspace{1cm} (17)

$$S_t = U^TX_t$$ \hspace{1cm} (18)

$S_t$ obtained from the calculation of the formula (18) multiplied by the network intrusion
route detection coefficient matrix $B$, getting the regression expression the network
intrusion route detection:

$$Y_t = S_tB$$ \hspace{1cm} (19)

**Definition 2**: nonlinear feature selection for network intrusion route detection. The
network intrusion route detection input data vectors $x_i$ are mapped to a high dimensional
feature space, and we carry out the linear PCA in the feature space. The PCA problem in
the high dimensional feature space $F$ of network intrusion route detection (dimension is
set as $M$) can be described as a diagonalisation. A covariance matrix estimated by $n$
network intrusion route detection samples:

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n} \Phi(x_i)\Phi(x_i)^T$$ \hspace{1cm} (20)

In formulas, $\Phi(x_i)$ represents the centralisation nonlinear mapping of the input variables
$\left\{x_{i}\right\}_{i=1}^{N} \in R^N$ of the network intrusion route detection. Diagonalisation represents that the
network intrusion route detection input data is transformed into the space coordinate
constructed by the PCA eigenvector $v$. We find the principal components analysis
eigenvalue $\lambda \geq 0$ of the network intrusion route detection and non-zero eigenvector $v \in F$
to meet the following eigen conditions:
A novel selection method of network intrusion optimal route detection

\( \lambda v = \hat{C} v \) \hspace{1cm} (21)

Since all solutions of \( \lambda \neq 0 \) are located in the spanning space \( \Phi(x_1), \ldots, \Phi(x_n) \), we can deduce the following equivalent eigenvalue problem of network intrusion route detection:

\[ n \lambda x = K \alpha \] \hspace{1cm} (22)

In the formula, \( \alpha \) represents the network intrusion route detection coefficient column vector \( T_1, T_2, \ldots, T_n \) satisfying the following formula:

\[ v = \sum_{i=1}^{n} T_i \Phi(x_i) \] \hspace{1cm} (23)

\( K \) is a symmetric \( n \times n \) Gram matrix in network intrusion route detection:

\[ K_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle = K(x_i, x_j) \] \hspace{1cm} (24)

Network intrusion route detection non zero eigenvalue \( \lambda_k = n \hat{\lambda}_k \) of normalisation matrix \( K \) corresponding to the eigenvector \( v^k \) of network intrusion route detection, even \( \hat{\lambda}_k (T^k, T^k) = 1 \), by calculating the projection from \( \Phi(x) \) to the network intrusion route detection eigenvector \( v^k \) we get the \( k^{th} \) network intrusion route detection nonlinear principal component of \( x \) as follows:

\[ P_k(x) = \langle v^k, \Phi(x) \rangle = \frac{1}{\sqrt{n \hat{\lambda}_k}} \sum_{i=1}^{n} T_i^k K(x_i, x) = \hat{\lambda}_k^{1/2} \sum_{i=1}^{n} T_i^k K(x_i, x) \] \hspace{1cm} (25)

We use \( V \) to represent the matrix constructed by taking the network intrusion route detection eigenvector \( \{v^p\}_{i=1}^{p} \) of \( \hat{C} \) as column element. \( \hat{V} \) represents the matrix constructed by taking the network intrusion route detection eigenvector \( \{\alpha^p\}_{i=1}^{p} \) as the column element. \( \Lambda \) represents the diagonal matrix \( \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p) \) formed corresponding to network intrusion route detection eigenvalue. \( \hat{\Lambda} \) represents the network intrusion route detection high dimensional space vector transformation. \( \hat{\lambda} \) represents the network intrusion detection test sample element. By using these representation methods, we can project the training data points \( \{x_i\}_{i=1}^{n} \) of the network intrusion route detection converting the formula (24) into the following matrix:

\[ P = \Phi V = \Phi \Phi^T \hat{V} \hat{\Lambda}^{-1/2} = K \hat{V} \hat{\Lambda}^{-1/2} = \hat{V} \hat{\Lambda}^{-1/2} \] \hspace{1cm} (26)

In formulas, we use \( V = \Phi^T \hat{V} \hat{\Lambda}^{-1/2} \), similarly, for network intrusion route detection test data points \( \{x_i\}_{i=n+1}^{n+p} \), there are:

\[ P_t = \Phi_t \Phi^T \hat{V} \hat{\Lambda}^{-1/2} = K \hat{V} \hat{\Lambda}^{-1/2} \] \hspace{1cm} (27)
In formulas, $\Phi_t$ the matrix $n_t \times M$ composed of the mapping $\{\Phi(x_i)\}_{i=1}^{n_t}$ of the network intrusion route detection test sample points, $K_t$ represents the network intrusion route detection $n_t \times n$. Its element is:

$$(K_t)_{ij} = \left(\Phi(x_i), \Phi(x_j)\right) = K(x, x)$$  

(28)

In the actual computation, the centralisation in the network intrusion route detection feature space make the following transformation for the detection matrix $K$ and $K_t$:

$$K \leftarrow \left( I_n - \frac{1}{n} 1_n 1_n^T \right) K \left( I_n - \frac{1}{n} 1_n 1_n^T \right)^T$$  

(29)

$$K_t \leftarrow \left( K_t - \frac{1}{n} 1_n 1_n^T K \right) \left( I_n - \frac{1}{n} 1_n 1_n^T \right)^T$$  

(30)

In formulas, $1_n$ and $1_{n_t}$ represent the full 1 vector of the network intrusion route detection matrix with $n$ and $n_t$ length respectively. $I$ represents the $n$-dimensional network intrusion route detection identity matrix.

After the network intrusion route detection training data $X = \{x_i\}_{i=1}^n$ is projected to the high dimensional feature space and selecting several principal components in front, we obtain the nonlinear principal components $P$ of network intrusion route detection. We can use the following formula to express the nonlinear regression equation of the network intrusion route detection:

$$Y = PB + \varepsilon$$  

(31)

Among them, $\varepsilon$ represents the network intrusion route detection noise. $B$ represents the network intrusion route detection regression coefficient matrix. The least squares solution of the nonlinear regression equation of network intrusion route detection of $B$ can be calculated by using the formula (32):

$$B = (P^T P)^{-1} P^T Y$$  

(32)

The following network intrusion route detection model:

$$Y_i = \text{sgn}(P_i B)$$  

(33)

Among them, $P_i$ represents the nonlinear principal component of the network intrusion route detection corresponding to the network intrusion route detection test sample $X_i = \{x_i\}_{i=1}^{n_t}$.

### 2.2 Naive Bayesian classification of network route detection

According to the network intrusion route detection model in Section 2.1, using the calculation method of posterior probability of probability of detection to determine the probability of network intrusion route detection sample category, we associate the prior probability of network intrusion route detection with the correlation of posterior probability, then using the prior information and sample data information of network intrusion route detection to determine the posterior probability of the network route
A novel selection method of network intrusion optimal route detection

According to the Bayesian formula of network intrusion route detection data, we suppose that \( A_1, A_2, \ldots, A_n \) indicates a group of two incompatible events in the network intrusion route detection data, but \( B \) can and can only occur simultaneously with one network intrusion route detection event. So the following formula is true:

\[
P(A|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \sum_{i=1}^{n} P(B|A_i)P(A_i)
\]

Naive Bayesian classification in the network intrusion route detection data is based on a simple setting. That is the given sample set for the network intrusion route detection. The probabilities of each category in network intrusion route detection data are known, and the influence of network intrusion route detection data attribute for categories are independent of each other. If the target value of the network intrusion route detection data instance is given, we can see that joint probability is equal to the probability product of each network intrusion route detection data individual attributes.

Supposing that the network intrusion route detection sample set has \( n \) attributes, \( A_1, A_2, \ldots, A_n \), network intrusion route detection attribute value forms the feature vector of the sample. Possible categories have \( m \), in details, \( \{C_1, C_2, \ldots, C_m\} \). We set the feature vector of the sample to be classified in network intrusion route detection as \( \{x_1, x_2, \ldots, x_n\} \), and calculate the probability \( P(C_i|X) \) of that \( X \) in the network intrusion route detection belongs to each category. The calculating formula of the probability \( P(C_i|X) \) of network intrusion route detection:

\[
P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}
\]

In the formula, because all \( C_i \) and \( P(X) \) are the same, it is only necessary to compare the molecular parts \( P(X|C_i) \) and \( P(C_i) \) in the formula. \( P(C_i) \) can be obtained from the network intrusion route detection training set. This value is equal to the proportion of that the network intrusion route detection training set classification of the sample is \( C_i \). On the premise of the independent network intrusion route detection, the following formula is true:

\[
P(X|C_i) = \prod_{j=1}^{n} P(x_j|C_i)
\]

In formulas, \( P(x_j|C_i) \) is estimated by the training set of network intrusion route detection. Assuming that \( A_j \) represents network intrusion route detection classification attributes, \( P(x_j|C_i) \) is equal to the attribute \( A_j \) in category as \( C_i \) of network intrusion route detection training samples, and is equal to the proportion of \( x_j \). We need to calculate \( P(X|C_i)P(C_i) \) of each class \( C_i \) in the network intrusion route detection for the classification of \( X \). If the network intrusion route detection sample is classified into \( C_i \), we set \( m \) as the total number of network intrusion route detection data class, making \( P(X|C_i)P(C_i) \) get the
class $C_i$ with maximum value, that is the class that the $X$ sample to be tested belongs to in network intrusion route detection, which needs to meet the following conditions:

$$
\begin{align*}
    P(X|C_i) &= \prod_{j=1}^{n} P(x_j|C_i) P(C_i) \\
    P(X|C_j) P(C_j) &> P(X|C_i) P(C_i)
\end{align*}
\quad (37)
$$

1 \leq j \leq m, j \neq i

In the network structure of naive Bayesian network, there is only one node. Other nodes represent various attributes of classification of network intrusion route detection. Each attribute node has only one class node, and each attribute node in network intrusion route detection is independent of each other. Naive Bayesian classification structure chart of network intrusion route detection data is shown in Figures 2 and 3.

**Figure 2** Classification structure of naive Bayesian

![Network route node types](image)
A novel selection method of network intrusion optimal route detection

Figures 2 and 3 show that directed edge shows the probability dependency relation between nodes in naive Bayesian classification structure of the network intrusion route detection. Naive Bayesian classification in the network intrusion route detection data is to calculate the posterior probability of each node, and uses the posterior probability to update the prior probability, and takes maximum in the network intrusion route detection probability as the classification basis, completing network intrusion route detection data classification.

2.3 Selecting method of network intrusion optimal route detection based on naive Bayesian classification

We classify the data set using the maximum of network intrusion route detection probability, setting the network intrusion route detection training data set as $M_0$. On the basis of classification, the optimal route selection of network intrusion is completed. Specific processes are as follows:

- The first step: pre-process the network intrusion route detection training set $M_0$ so as to get the training data set $M_1$ of network intrusion route detection.
- The second step: use the PCA method of network intrusion route detection to select the feature of $M_1$, and get a new data set $M_2$ of network intrusion route detection containing fewer attributes.
- The third step: the network intrusion detection data set $M_2$ are discretised to get a new network intrusion detection data set $M_3$.
- The fourth step: get the structure of Bayesian network with network intrusion route detection training for $M_3$.
- The fifth step: Based on naive Bayesian network route detection classification method, get some good network route.
- The sixth step: select the optimal route of network intrusion detection based on the naive Bayesian classification. The specific process is as follows.
Set two pointers for the network intrusion route detection training data set, respectively, they are the head pointer \( P_1 \) of the network intrusion route detection (pointing to the head of network intrusion route detection training data set) and the tail pointer \( P_2 \) of network intrusion route detection (pointing to the tail of network intrusion route detection training data set); at the same time, set two pointers for the test data set of network intrusion route detection, the head pointer and tail pointer of network intrusion route detection are expressed as \( Q_1 \) and \( Q_2 \), in which \( P = P_1, Q = Q_1 \).

1. The data that the network intrusion route detection pointer \( P \) points to is stored in the data set \( \mathcal{M} \), \( P = P_1 + 1 \).
2. Repeat the operation in step (1) until \( P > P_2 \).
3. The parameter \( C \) of each node in the Bayesian network structure are calculated by the network intrusion route detection data set.
4. Use the network intrusion route detection Bayesian network classification structure and the parameter \( C \) from the calculation to detect the network route data that the pointer \( Q \) points to in the training data set, and the data is appended to the tail of the training data set \( \mathcal{M} \) of network intrusion route detection, \( P_2 = P_2 + 1, Q = Q_1 + 1 \).
5. Repeat the operation in step (3) until it satisfies \( Q = Q_1 + N \) or \( Q > Q_2 \).
6. If \( Q > Q_2 \), go on next step, otherwise \( P_2 = P_1 + N, P = P_1 \), rerun the step (1).
7. Finish the test, select the optimal route of network intrusion route detection.

3 Simulation results and analysis

In the inteli72620QM, 32GBRAM, 2TB hard disk, Windows7 operating system environment workstation of laboratory, we use Matlab network intrusion route detection model for simulation experiments. The data comes from the KDDCUP99 database. KDDCUP99 database includes a variety of different types of network intrusion attack data, selecting the four common classes:

1. Troy Trojan attack
2. false message attack
3. denial-of-service attack
4. remote user unauthorised access attack.

Due to the huge amount of data in the KDDCUP99 database, it is impossible to extract and use all the data. We randomly selected some network intrusion data in KDDCUP99 database for the experimental simulation. The number of selected network intrusion samples is as follows:

1. 200 Troy Trojan attack samples
2. 300 false message attack samples
3. 500 denial-of-service attack samples
4  100 remote user unauthorised access attack samples
5  2,000 normal samples of network systems.

The performance evaluation of network intrusion route detection method is based on the false alarm rate, the detection rate and the detection time of network intrusion detection:

\[
\text{false alarm rate} = \frac{\text{number of normal sample data of false alarm for network intrusion}}{\text{total number of normal sample data}} \times 100\% 
\]

\[
\text{Detection rate} = \frac{\text{Detected total sample data of network intrusion}}{\text{Total number of network intrusion data samples}} \times 100\%
\]

In order to reduce the influence of the characteristics of sample data of few network intrusion route on Bayes classification performance, the network intrusion route sample data is normalised:

\[
x'_i = \frac{x_i - \bar{x}}{x_{std}}, \quad i = 1, 2, \cdots, n
\]

In formulas, \(x_{std}\) represents the characteristic standard deviation of the network intrusion route sample. \(\bar{x}\) represents and the feature variance of the network intrusion route sample.

### Table 1  Comparison results of detection rate (%) and false alarm rate (%) of network intrusion optimal route in several different methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detection rate</td>
<td>False alarm rate</td>
</tr>
<tr>
<td>Trojan horse attack</td>
<td>63.2</td>
<td>20.5</td>
</tr>
<tr>
<td>False message attack</td>
<td>56.8</td>
<td>19.6</td>
</tr>
<tr>
<td>Denial of service attack</td>
<td>60.3</td>
<td>25.6</td>
</tr>
<tr>
<td>Remote user unauthorised access attack</td>
<td>65.8</td>
<td>18.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network intrusion class</th>
<th>Method of reference Xie et al. (2015)</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detection rate</td>
<td>False alarm rate</td>
</tr>
<tr>
<td>Trojan horse attack</td>
<td>66.4</td>
<td>27.6</td>
</tr>
<tr>
<td>False message attack</td>
<td>59.6</td>
<td>26.1</td>
</tr>
<tr>
<td>Denial of service attack</td>
<td>62.1</td>
<td>18.3</td>
</tr>
<tr>
<td>Remote user unauthorised access attack</td>
<td>59.0</td>
<td>18.8</td>
</tr>
</tbody>
</table>
In order to validate the comprehensive effectiveness of network intrusion optimal route detection selection method based on naive Bayesian proposed in this paper and the superiority of the false alarm rate and the detection rate of the optimal route detection in the network intrusion, we respectively compare with the detection rate and the false alarm rate using the method in reference Aziz et al. (2016), Saravanan and Senthilkumar (2015) and Xie et al. (2015) and proposed method. We use the same network intrusion route training samples to train the proposed method and several methods of comparison, and then use the same network intrusion route test data for simulation experiment in the same environment. The results are shown in Table 1.

Figure 3  The optimal route detection rate of different methods under different types of attacks, (a) Trojan horse attacking sample size/number (b) false information attack sample number/number (c) service denial attack sample number/number (d) the remote user unauthorised access attack sample number/number

From the comparison result of experimental simulation test in Table 1, we can see that although the network optimal route detection rate of reference Aziz et al. (2016) is high relative to different types of network intrusion classification, but it has low false alarm rate. The network optimal route detection rate of reference Xie et al. (2015) is higher relative to different types of network intrusion types, but its false alarm rate is low.
However the proposed algorithm improves the detection rate of network optimal route, and reduces the false alarm rate. In Figures 3 and 4, we compare and analyse the network optimal route detection rate and network optimal route false detection rate of different methods under different types of attacks. We can see that no matter under what type of network attack, the detection rate of this method is higher, and the false alarm rate of network optimal route is the lowest. Simulation results show that the optimal route of network intrusion detection selected in this paper has better detection performance.

Figure 4 The optimal route false detection rate of different methods under different types of attacks, (a) Trojan horse attacking sample number/number (b) false information attack sample number/number (c) service denial attack sample number/number (d) the remote user unauthorised access attack sample number/number

According to the data in Table 1, the detection time of different network intrusion detection methods is compared and analysed. The results are shown in Table 2.

From the data comparison in Table 2, we can see that compared with the method in reference Aziz et al. (2016), Saravanan and Senthilkumar (2015) and Xie et al. (2015), the training time and testing time of this method in the network optimal route detection. This method improves the detection efficiency of the network optimal route, and it is the optimal route in network intrusion detection.
Table 2  Comparison of network optimal route detection time (s) with several different methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training time</td>
<td>Detection time</td>
</tr>
<tr>
<td>Trojan horse attack</td>
<td>23.2</td>
<td>10.5</td>
</tr>
<tr>
<td>False message attack</td>
<td>16.8</td>
<td>11.6</td>
</tr>
<tr>
<td>Denial of service attack</td>
<td>8.3</td>
<td>5.6</td>
</tr>
<tr>
<td>Remote user unauthorised access attack</td>
<td>8.8</td>
<td>8.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network intrusion class</th>
<th>Method in reference Xie et al. (2015)</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training time</td>
<td>Detection time</td>
</tr>
<tr>
<td>Trojan horse attack</td>
<td>7.4</td>
<td>7.6</td>
</tr>
<tr>
<td>False message attack</td>
<td>9.6</td>
<td>6.1</td>
</tr>
<tr>
<td>Denial of service attack</td>
<td>12.1</td>
<td>6.3</td>
</tr>
<tr>
<td>Remote user unauthorised access attack</td>
<td>9.0</td>
<td>8.8</td>
</tr>
<tr>
<td>Network intrusion class</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Conclusions

In the process of the optimal route selection for network intrusion detection by using current methods, there are some problems such as low detection rate and high false alarm rate. This paper proposes a method the optimal route selection for network intrusion detection based on Naive Bayes. Simulation results show that the proposed method improves the detection rate of network optimal route, reduce false alarm rate in the process of network intrusion detection optimal route selection. The detection efficiency in the detection process is also improved. In the face of different types of network attacks, this method can take appropriate measures in time, and it has a good practical value, and it lays a foundation for practical application in the future.

References


A novel selection method of network intrusion optimal route detection


