Research on Iolation Dtection Agorithm of Ntwork Itrusion Data

Xuan Huang[1,2]
[1]School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, 330013, China
[2]Faculty of Management, Nanchang University, Nanchang, 330031, China

Abstract: In terms of the network intrusion data isolation detection, the conventional isolation detection algorithm has disadvantages. The isolation detection accuracy is low and the isolation success rate is not high, so this paper proposes the research on isolation detection algorithm of network intrusion data. Based on the structure design of isolation detection algorithm, the anomaly detection mechanism of statistical analysis, neural network anomaly detection mechanism and data mining anomaly detection mechanism are constructed to realize the operation design of network intrusion data isolation detection algorithm. Relying on the process design of isolation detection algorithm and the parameter optimization of intrusion isolation detection algorithm, the isolation detection algorithm of network intrusion data is realized. The experimental data shows that the proposed isolation detection algorithm can improve the isolation detection accuracy by 45.17% and the isolation success rate by 26.6%, compared with the conventional isolation detection algorithm. It is suitable for the isolation detection of network intrusion data.

Keywords: Network intrusion; Data isolation; Detection algorithm; Isolation detection; Data mining; Anomaly detection

1. Introduction

Relying on computer virus theory, the conventional isolation detection algorithm studies the propagation mode. However, for the network intrusion data isolation detection, due to the limitation of algorithm operation results, the isolation detection accuracy is low and the isolation success rate is not high. It is not suitable for the isolation detection of network intrusion data[1].Therefore, the research on the isolation detection algorithm of network intrusion data is proposed. Based on the structure design of isolation detection algorithm, the structure of isolation detection algorithm is established, the anomaly detection mechanism of statistical analysis and the anomaly detection mechanism of neural network are constructed, so as to complete the operation design of network intrusion data isolation detection algorithm. Relying on the process design of isolation detection algorithm and the parameter optimization of intrusion isolation detection algorithm, the isolation detection algorithm of network intrusion data is studied. In order to ensure the effectiveness of the designed isolation detection algorithm and simulate the network intrusion test environment, two different isolation detection algorithms are used to simulate the isolation detection accuracy and isolation success rate.

2. Operation Design of Network Intrusion Data Isolation Detection Algorithm

2.1. Constructing structure system of isolation detection algorithm

Isolation detection algorithm is a method to detect and analyze abnormal behavior, and then to judge and discover the unknown attack mode. Generally, computers work in normal mode, but intrusion behavior is different from normal behavior. All the normal behaviors are summarized, and the normal behavior model is constructed. The existence of intrusion behavior is judged by discovering abnormal behavior. The structure of the isolation detection algorithm is shown in Figure 1[2].

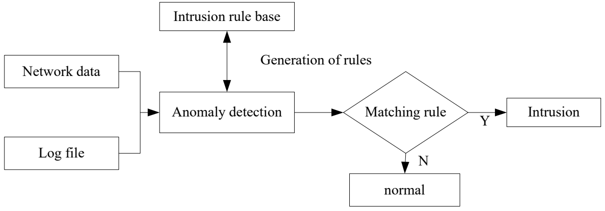

Figure 1. Structure system of the isolation detection algorithm

The structure of isolation detection algorithm takes normal behavior as the detection object, and establishes the structure of isolation detection algorithm under perfect normal behavior. Normal behavior characteristics can be expressed by various characteristic parameters required by the normal operation of the system, such as CPU utilization rate, memory utilization rate, file checksum, and network flow, etc [3].Anomaly detection collects various characteristic parameters of the system in normal operation. It establishes a normal behavior model according to these parameters. Then draws up a normal behavior model as well as an evaluation function to detect and calculate. In the actual detection process, an abnormal index limit is usually set. If an abnormal behavior is found to exceed the threshold value, the behavior is marked as abnormal which will be handled to determine whether there is an intrusion. The main techniques for anomaly detection include statistical analysis, data mining, neural network, computer immune system and so on.

2.2. Constructing anomaly detection mechanism for statistical analysis

Anomaly detection of statistical analysis is the most widely used technology. Statistical analysis first sets reasonable statistical variables for observed events, and describes the behavior of users or systems by using statistical variables, such as the frequency, interval, CPU usage rate, and memory occupancy rate, etc. All statistical events are compiled into a systematic statistical model and the statistical variables are audited. If the audit results deviate significantly from the statistical model, the anomaly will occur. The IDES of SRI and NIDES adopt statistical analysis technology to realize anomaly detection. The anomaly detection mechanism of statistical analysis is shown in Figure 2.

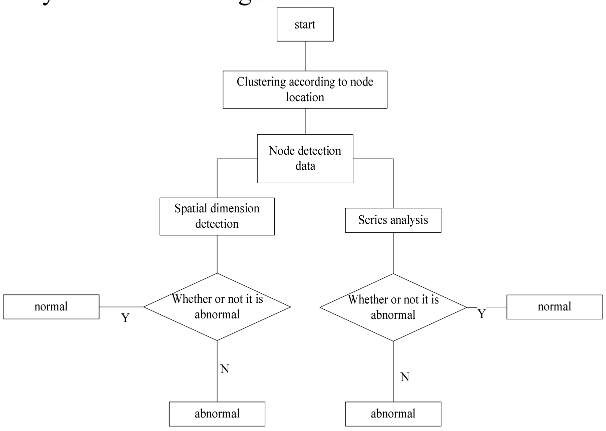
start
Clustering according to node location
Node detection data
Spatial dimension detection
Series analysis
normal
Y
Whether or not it is abnormal
Whether or not it is abnormal
Y
normal
N
N
abnormal
abnormal

Figure 2. Anomaly detection mechanism of statistical analysis

2.3. Constructing the anomaly detection mechanism of neural network

Neural network is composed of a large number of processing units. It is a multilayered network structure made up of connection units with weights. Artificial neural network (ANN) is used for anomaly detection, and adaptive learning is used to mark anomaly behavior in modeling. K.L.Fox first proposed using neural networks to construct behavior patterns of observers. Neural networks trained from historical behavior data can predict the next event in an event sequence. The prediction accuracy can be used as an anomaly index to measure an event sequence [5]. JakeRyan trains a back-propagation neural network called Neural Network Intrusion Detector (NNID) according to the command operation habits of each UNIX user. The NNID is used to predict the next command of the user during detection. If the prediction results do not match or can not give a clear prediction, an exception is triggered. The schematic diagram of the anomaly detection mechanism for constructing the neural network is shown in Figure 3 [6].

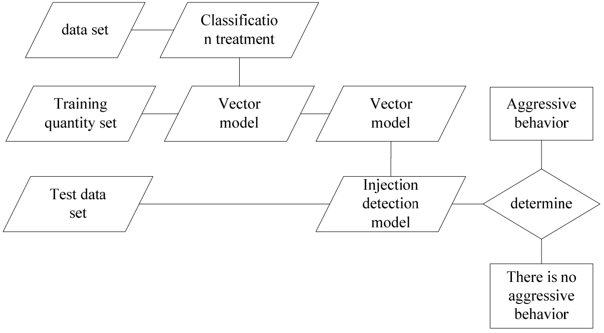

Figure 3. Construction of the anomaly detection mechanism of neural network

2.4. Constructing anomaly detection mechanism for data mining

Data mining anomaly detection is a fast anomaly detection method based on Ensemble. It has linear time complexity and high precision. It is a state-of-the-art algorithm that meets the requirements of big data processing. The iForest of data mining is suitable for anomaly detection of continuous data. The anomaly is defined as an isolated outlier, which can be understood as a point with sparse distribution and far away from the high density group. Statistically, a sparsely distributed region in the data space indicates that the probability of data occurring in this region is very low, so the data in these regions can be considered abnormal.

iForest belongs to the Non-parametric and unsupervised methods, which do not use defined mathematical models or need labeled training. iForest uses a very efficient strategy to find out which points are easy to be isolated. Supposed that a random hyperplane is used to cut the data space, cut it once and two subspaces will appear. Then continue to cut each subspace with a random hyperplane and repeat it until there is only one data point in each subspace. Intuitively, it can be seen that clusters with high densities can be cut many times to stop cutting, but those with low densities can easily stop in a subspace very early, as shown in the following test sample [8].

It can be seen that D is most likely to be an exception, because it was isolated at the earliest time, thus the anomaly detection for data mining is realized.

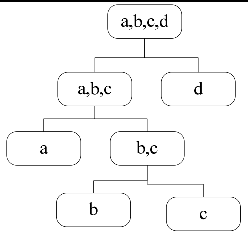

Figure 4. Test sample

3. Realization of Isolation Detection Algorithm of Network Intrusion Data

3.1. Determination of isolation detection algorithm process

The basic flow of this algorithm is shown in Figure 5, and the following steps will be explained.

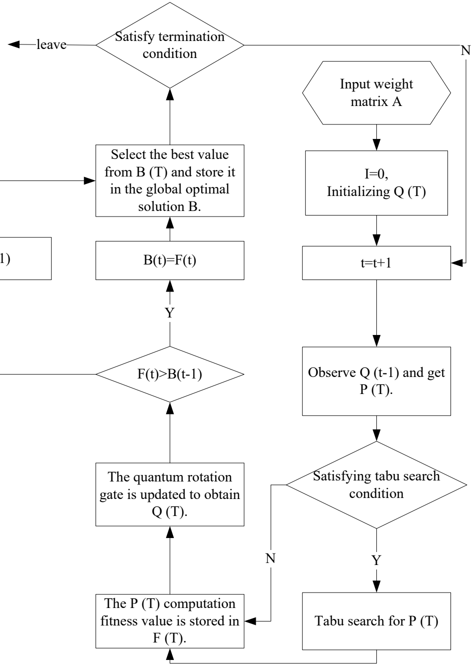
leave
Satisfy termination condition
N
Input weight matrix A
Select the best value from B (T) and store it in the global optimal solution B.
I=0, Initializing Q (T)
B(t)=B(t-1)
B(t)=F(t)
t=t+1
Y
N
F(t)>B(t-1)
Observe Q (t-1) and get P (T).
The quantum rotation gate is updated to obtain Q (T).
Satisfying tabu search condition
N
Y
The P (T) computation fitness value is stored in F (T).
Tabu search for P (T)

Figure 5. Basic flow of isolation detection algorithm

1. Input weight matrix A to get the scale of the problem.
2. Initial algebra $t=0$.
3. Initialize intrusion data population $Q(t)=\{q_1^t,...,q_i^t,...q_n^t\}$.
4. $t=t+1$.
5. Observe the intrusion data population $Q(t-1)$ and get the determined binary intrusion data population $P(t)=\{X_1^t,X_2^t,...,X_m^t\}$.
6. If the tabu search condition is satisfied, go to step 7; otherwise, go to step 8.
7. For every intrusion data of $P(t)$, the Tabu search algorithm is called.
8. Calculate the fitness value $F(t)=\{f_1^t,...,f_i^t,...,f_n^t\}$of each intrusion data in population $Q(t)$.
9. Update the intrusion data population $Q(t)$ with quantum rotation gate.
10. Compare each fitness value f_i^t in $F(t)$ with b_i^{t-1} in the best solution $B(t-1)=\{b_1^{t-1},...,b_i^{t-1},...,b_n^{t-1}\}$ of the previous generation, and the bigger one is stored in the optimal solution of the current generation $B(t)$.
11. Compare all the values in $B(t)=\{b_1^t,...,b_i^t,...,b_n^t\}$ and store the best solution in the global optimal solution b.
12. If the termination condition is not satisfied, then jump to step 4.
13. Save the global optimal solution b to the Notepad, and end the isolation detection.

3.2. Algorithm process description

The network access is usually realized in step 3. All the intrusion data, such as q_j^t, α_i^t and β_i^t is initialized to be $\frac{1}{\sqrt{2}}$, that is, the probability of the intrusion data q_j^0 being all superposition states is equal, and its state is as follows: $\left|\psi_{q_j^0}\right\rangle=\sum_{k=1}^{2^m}\frac{1}{\sqrt{2^m}}\left|X_k\right\rangle$

$X_k=(x_1,...,x_i,...,x_j,...,x_m)$ in the formula represents the superposition state of K, and $x_i(i=1,2,...,m)\in[0,1]$. Recognize the intrusion data. A set of definite binary solutions are generated from $P(t)=\{X_1^t,X_2^t,...,X_m^t\}$ by observing $Q(t-1)$ in step5. $X_j^t(j=1,2,...,n)$ is a binary string that is corresponding to X_k. It is generated by observing all the bits of the intrusion data. This observation means that a decimal r between 0 and 1 is randomly generated, and then the number is compared with $[\alpha]^2$ of q_i^t. If r< $[\alpha]^2$, $x_i=0$; otherwise, $x_i=1$.

Step 6 and step 7 are the first keys in this algorithm. By judging whether the TS-condition of tabu search is satisfied, if it is satisfied, each intrusion data in the population must be searched locally, so as to search for the local optimal solution within a local range of the current intrusion data [9].

Generally, $N=\{1,2,......,n\}$ is used to represent the component ordinal number of an X vector, and the input Q matrix is treated as a lower triangular matrix. Δi is used to represent the mobile value that is flipped from the variable x_i. $q_{(i,j)}$ is used to represent q_{ij} when i>j or $q_{(i,j)}$ is used to represent q_{ji} when j>i. In this way, each increment can be estimated in advance with the following formula in linear time: $\Delta i = (1-2x_i)(q_{ii} + \sum_{j\in N, j\neq i, x_j=1} q(i,j))$

In $q(i,j)$, when $i > j$, $q(i,j) = q_{ij}$. When $i < j$, $q(i,j) = q_{ji}$. In addition, once a component is inverted, it means that the next prediction begins on the basis of the inverted vector. In particular, the following computational procedure may be performed to re-estimate the estimated added value of the flipped component itself and other components.

$\Delta_i = -\Delta_i$ for each $j \in N - \{i\}$ $\Delta_j = \Delta_j + \sigma_{ij} q(i,j)$

Among them, $\sigma_{ij} = 1$. If $x_j = x_i$, $\sigma_{ij} = -1$.

The algorithm contains a tabu list as a storage structure based on the "nearest access" principle, so as to ensure that solutions accessed within a certain cycle interval are not re-accessed. During execution, each time x_i is flipped, TabuTenure (i) is randomly assigned a value (this value is used to identify whether it is a tabu) to avoid being flipped again when circling x_i [10]. Set up here as follows:

$TabuTenure(i) = tt + rand(10)$

Among them, TT is a given constant, and Rand (10) takes a random number from 1~10.

When the function value of the best solution can no longer be improved in the given move of α, the tabu search stops, which is called the optimal cut-off level. In step 8, according to the formula:

$$(x_1,\ldots,x_i,\ldots,x_j,\ldots,x_n)\begin{pmatrix} q_{11} & \cdots & q_{1j} & \cdots & q_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ q_{i1} & \cdots & q_{ij} & \cdots & q_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ q_{n1} & \cdots & q_{nj} & \cdots & q_{nn} \end{pmatrix}(x_1,\ldots,x_i,\ldots,x_j,\ldots,x_n)^T$$

the fitness value of each intrusion data is calculated, and the solution formula can be simplified here.

$$f(X) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} q_{ij} x_j$$

Such is easy to program.

The terminating conditions commonly used in the algorithm are as follows:

1) Limit the maximum evolution algebra.

2) The probability Prob (b) of each individual converges to the optimal solution b is greater than a specific value γ_0 . $Prob(b) = \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{i=1}^{m} p_{j,i}\right)$. Among them, $p_{j,i} = \begin{cases} |\alpha|^2 \ ,\ , & if(b_i = 0) \\ |\beta|^2 & if(b_i = 1) \end{cases}$. n is the number of genes in an individual, and m is the number of quantum bits in each individual.

As can be seen from the above formula, the calculation of Prob (b) is related to the probability amplitude of all quantum bits, so the value of γ_0 is difficult to determine. When $\gamma_0 \geq 0.1$, all the results are basically the same. As γ_0 increases, the algorithm converges better, but the execution time will increase significantly. In order to obtain the optimal solution and achieve isolation detection, usually the execution time is increased to achieve research on isolation detection algorithm of network intrusion data.

4. Experimental Results and Analysis

In order to ensure the validity of the proposed algorithm for isolation and detection of network intrusion data, simulation experiments are carried out. In the process of the experiment, different network intrusions are taken as the test objects. The simulation test of isolation detection precision and isolation success rate is carried out. The type of network intrusion, the invasion structure and the type of intrusion are simulated. In order to ensure the validity of the test, the conventional isolation detection algorithm is used as the comparison object, and the two simulation results are compared. The test data is presented in the same data chart.

4.1. Comparison of isolation detection accuracy

During the experiment, two different isolation detection algorithms are used to work in the simulation environment, and the change of isolation detection accuracy of network intrusion is analyzed. The results of the test are shown in Table 1.

Table 1. Comparison table of isolation detection accuracy

Case type number	Routine isolation detection algorithm/%	Isolation detection algorithm proposed/%
1#	42	95
2#	56	92
3#	54	94
4#	45	93
5#	36	90
6#	48	91

The arithmetic average processing of the isolation detection accuracy of the proposed isolation detection algorithm and the conventional isolation detection algorithm show that the isolation detection accuracy of the proposed isolation detection algorithm is 45.17% higher than that of the conventional isolation detection algorithm.

4.2. Comparison of isolation success rate

During the experiment, two different isolation detection algorithms are also used to work in the simulation environment, and the change of isolation success rate of network intrusion is analyzed. The comparison curve of the test results are shown in Figure 6.

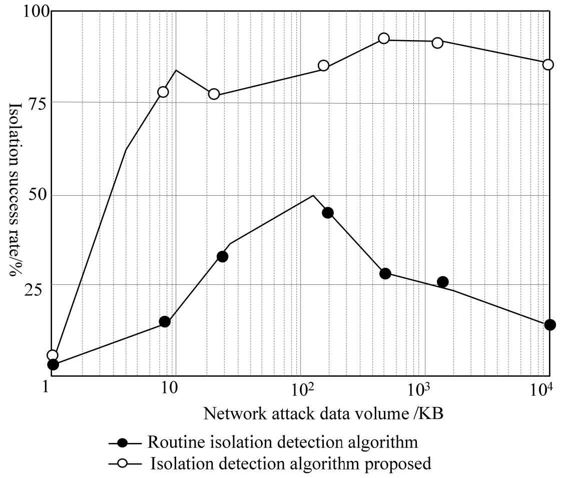

Figure 6. Comparison Chart of isolation success rate

The arithmetic average processing of the isolation success rate of the proposed isolation detection algorithm and the conventional isolation detection algorithm show that the proposed isolation detection algorithm is 26.6% higher than the conventional isolation detection algorithm, which is suitable for the isolation detection of network intrusion data.

5. Conclusion

In this paper, the research on isolation detection algorithm of network intrusion data is proposed. Based on the operation design of network intrusion data isolation detection algorithm, isolation detection algorithm flow design and parameter optimization, the research of this paper is realized. The experimental data shows that the method designed in this paper is very effective. It is hoped that the research in this paper can provide a theoretical basis for the isolation detection algorithm of network intrusion data.

6. Acknowledgments

The authors would like to thank all the referees for their constructive and insightful comments that allowed us to improve this paper. The contributions of this research are the following: The contributions of this research are the following: Primary Research & Development Plan of Jiangxi Province (No. 20181ACE50033); the Jiangxi Province Social Science Planning Project (No.18YJ28); the National Natural Science Foundation of China (No. 61762062, 61601215, 61862042).

References

[1] LuoTingting, ZhouHangfei. Research on effective isolation simulation of private network script virus [J]. computer simulation, 2018(1):275-278.

[2] Teixeira A, Shames I, Sandberg H, et al. Distributed fault detection and isolation resilient to network model uncertainties[J]. IEEE Transactions on Cybernetics, 2017, 44(11):2024-2037.

[3] Rina K, Nath S, Marchang N, et al. Can Clustering be Used to Detect Intrusion During Spectrum Sensing in Cognitive Radio Networks?[J]. IEEE Systems Journal, 2018, 12(99):1-10.

[4] Zhang B, Hu G, Zhou Z, et al. Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base[J]. Etri Journal, 2017, 39(4):592-604.

[5] Ramakrishnan S, Devaraju S. Attack's Feature Selection-Based Network Intrusion Detection System Using Fuzzy Control Language[J]. International Journal of Fuzzy Systems, 2016, 19(2):1-13.

[6] Chiba Z, Abghour N, Moussaid K, et al. A Cooperative and Hybrid Network Intrusion Detection Framework in Cloud Computing Based on Snort and Optimized Back Propagation Neural Network[J]. Procedia Computer Science, 2016, 83:1200-1206.

[7] Sultana N, Chilamkurti N, Peng W, et al. Survey on SDN based network intrusion detection system using machine learning approaches[J]. Peer-to-Peer Networking and Applications, 2018(1-2):1-9.

[8] Roshan S, Miche Y, Akusok A, et al. Adaptive and online network intrusion detection system using clustering and Extreme Learning Machines[J]. Journal of the Franklin Institute, 2017.

[9] Sou S I, Lin C S. Random Packet Inspection Scheme for Network Intrusion Prevention in LTE Core Networks[J]. IEEE Transactions on Vehicular Technology, 2017, PP(99):1-1.

[10] Yu Y, Ye Z, Zheng X, et al. An efficient cascaded method for network intrusion detection based on extreme learning machines[J]. Journal of Supercomputing, 2016(2):1-16.