Network Intrusion Monitoring Technology Based on Decision Tree Method

Network Intrusion Monitoring Technology Based on Decision Tree Method

Xinyu Liu [1*]

[1]Beijing University of Posts and Telecommunications, Beijing, Beijing, 100876, China

[*]Corresponding author's e-mail: liuxinyu8252@163.com

Abstract. In order to solve the problem of poor real-time network security intrusion detection, this paper proposes a network intrusion monitoring technology based on decision tree method. Firstly, the network intrusion monitoring module is set up in the existing network topology model, then the network information collection, information processing and information analysis modules are combined. Finally, the decision tree method is used to identify the abnormal data, effectively realize different types of abnormal data monitoring, and finally realize the network intrusion monitoring. Experimental results show that the proposed intrusion monitoring method can better monitor network intrusion behavior, and has higher monitoring accuracy than other methods.

1. Introduction

With the development of Internet technology, network plays an important role in social production and life, and network information security is becoming more and more important [1-3]. Due to many loopholes in the existing network system, the network is still facing the threat of hacker attack. Therefore, it is necessary to study new active intrusion detection methods to maintain network security [4-5]. Network intrusion usually needs to be scanned before large-scale intrusion, so as to generate abnormal data. Finding abnormal data in advance is of great significance for network defense. The static defense method belongs to passive defense, which will miss the best defense time. Therefore, it is necessary to use dynamic network defense method to monitor abnormal data in real time [6-7].

The existing intrusion monitoring methods are mainly divided into quantitative monitoring and qualitative monitoring. The quantitative monitoring mainly uses the pre-set decision threshold to monitor the abnormal data, mainly including the Bayesian based anomaly monitoring method. The intrusion detection method based on qualitative monitoring refers to the method of Intrusion Detection Based on pattern matching, in which some behaviors are defined as illegal intrusion. To some extent, the existing methods can monitor the abnormal intrusion, but the recognition probability still can not meet the requirements.

In this paper, the network intrusion monitoring system is added between the Internet and the server. Based on the three modules of information collection, information processing and information analysis, the decision tree method of data dimensionality reduction is introduced to identify the abnormal data to realize the network intrusion monitoring. Experimental results show that the proposed method has good applicability for different types of intrusion recognition.

2. System model design

In order to realize the real-time monitoring of abnormal network data and ensure the health of the whole network, the intrusion detection module is added to the original network, and its network architecture is shown in Figure 1 in this paper. The Internet connects with network terminals and servers through routers and switches. Meanwhile, the external devices of the whole network include intruders and security devices. In order to discover the intrusion device in time, this paper uses the intrusion monitoring module to monitor the network abnormal data before the network terminal, so as to ensure the safety of the whole network.

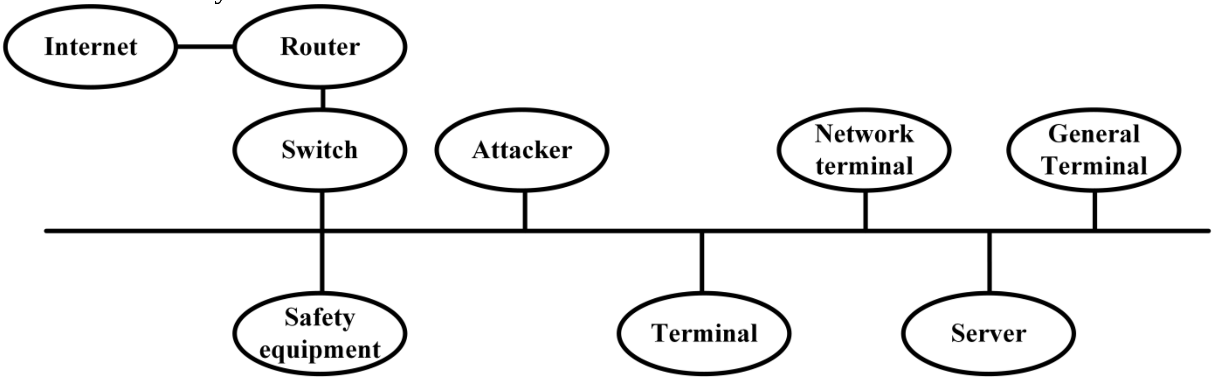

Figure 1. Intrusion monitoring network architecture

The intrusion monitoring module is located between the network access point and the terminal equipment, which can effectively monitor the abnormal data and alarm the system, so as to take corresponding measures to achieve early defense. For the intruder, because of the existence of intrusion monitoring module, the server and device terminal of the system are invisible to the intruder. The intruder can only access the router address and switch information, but can not get other effective information further, and can not carry out further attacks. Therefore, network security can be effectively maintained.

According to the function modules, the whole network can be divided into three layers, namely user interaction layer, application logic layer and basic layer. The user UI interaction layer mainly uses B / S structure to realize the front-end display and interaction of the system. The application logic layer is the core of the system, including data acquisition, data processing and data analysis sub modules. The module of data processing and data analysis performs the function of intrusion monitoring, alarms the system after finding the abnormal data, and controls the access of the abnormal intruder. The basic layer is the bottom layer of the whole system, which is mainly responsible for storing the collected network information and interacting with the basic equipment.

3. Intrusion monitoring module architecture

As shown in Figure 2, the structure of the intrusion monitoring module designed in this paper can be divided into three parts: data collection, data processing and data analysis.

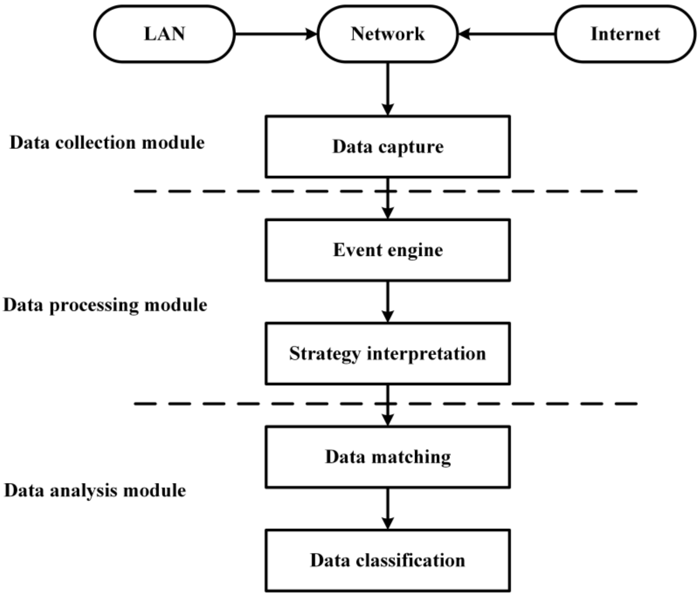

Figure 2. Composition of network intrusion monitoring module

The data collection module mainly obtains the data packets of LAN or Internet through data packet grabbing to realize the grabbing of massive data. In the face of massive data, it can be expanded with big data mining technology.

After data collection, using the event engine and policy interpretation to preprocess the data that conforms to certain rules, and feed back the processing results to the data analysis module.

The data analysis module is mainly responsible for identifying the data after preprocessing, classifying the data by using the corresponding decision algorithm through data matching, so as to find and process the abnormal data; in addition, the data analysis module can use the association detection method to predict whether the unknown data is abnormal, so as to improve the efficiency of system data processing.

4. The key technologies of intrusion monitoring

In this paper, decision tree algorithm based on dimension reduction processing is used to analyze network information, so as to realize intrusion monitoring. Its training model is shown in Figure 3

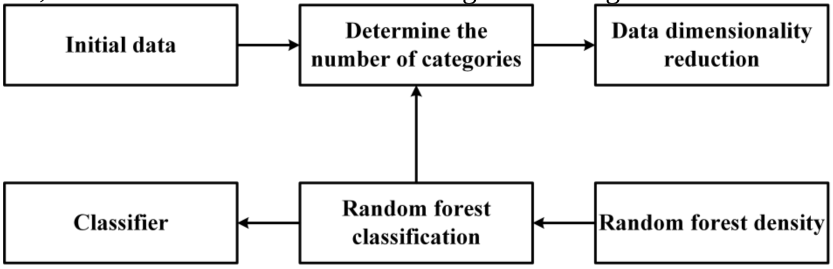
Initial data
Determine the number of categories
Data dimensionality reduction
Classifier
Random forest classification
Random forest density

Figure 3. Intrusion monitoring algorithms training flow chart

The core of building decision tree is to divide data attributes according to certain criteria. Ideally, the data in the attribute set should belong to the same category. But in the actual implementation process, it is difficult to obtain the ideal partition. Therefore, recursion is needed in decision tree algorithm.

In this paper, the gain of data information is taken as the partition criterion, and the set of network information samples collected by the system is assumed to be *D*. Where *d* is an attribute of set *D*, assuming that there are n data attributes, set *D* is a set of *N* data samples, which can be expressed as $D=\{d_1,d_2,\cdots,d_N\}$. The information gain can be expressed as

$$G(D,d)=E(D)-\sum_{n=1}^{N}\frac{\left|D^{n}\right|}{D}E(D^{n}) \tag{1}$$

Where, *G*(*D*, *d*) represents the information gain of data d in sample set *D*, and *E*(*d*) represents the information entropy of sample set *D*. If the proportion of the i-th sample is p_i, then the information entropy can be defined as

$$E(D)=-\sum_{i=1}^{N}p_{i}\log_{2}p_{i} \tag{2}$$

After calculating the information gain of all samples in the network information sample set D, the data is divided according to the information gain size, and the attribute with the largest information gain is regarded as the attribute of the data. In attribute set partition, not the larger the set association degree is, the more reasonable the data partition is; in extreme cases, when only one element exists in an attribute set, the set association degree is the largest. In order to avoid this kind of situation, this paper defines the information gain rate and Gini coefficient to divide the optimal attributes, where the information gain rate can be defined as

$$GR(D,d)=\frac{G(D,d)}{I(d)} \tag{3}$$

Where, *I*(*d*) represents a parameter value related to D, and the evaluation expression is

$$I(d)=-\sum_{n=1}^{N}\frac{\left|D^{n}\right|}{\left|D\right|}\log_{2}\frac{\left|D^{n}\right|}{\left|D\right|} \tag{4}$$

The information gain rate is sensitive to the number of sample attributes. The attribute with the largest information gain rate can be selected as the partition attribute.

Gini coefficient represents the probability of inconsistency between two data attributes in the data set. The Gini coefficient of attribute *d* in set *D* can be defined as the sum of the Gini coefficients of all attribute subsets. When the data association degree in attribute set is larger, the Gini coefficient of data is smaller. Therefore, the attribute with the smallest Gini coefficient is used as the data attribute. The Gini coefficient can be defined as:

$$Gini(D,d)=\sum_{n=1}^{N}\frac{\left|D^{n}\right|}{\left|D\right|}Gini(D^{n}) \tag{5}$$

Gini(*d*) can be defined as

$$Gini(D)=\sum_{i=1}^{N}\sum_{j=i}p_{i}p_{j}=1-\sum_{i=1}^{N}p_{i}^{2} \tag{6}$$

Then the final attribute decision function can be defined as

$$G(D，d)=a\cdot GR(D,d)-(1-a)\cdot Gini(D) \tag{7}$$

Where, *a* is the adjustment coefficient, and the value range is 0~1.

The steps of intrusion detection method based on dimension reduction decision tree proposed in this paper are as follows:

Input: original data set *D*; number of attributes *M*.

Step1: Eigenvalue decomposition of data;

Step2: The first k features with the largest eigenvalue are retained, and PCA method is used to reduce the dimension of data;

Step3: According to equation (7), the decision function of data set is calculated;

Step4: According to the size of attribute decision function, the decision tree method is used to classify the data;

Step5: If the data attribute classification remains unchanged, turn to 6; otherwise, turn to Step3

Step6: The end of the algorithm;

Output: attribute subset D_n.

5. Experimental verification and result analysis

For different types of intrusion detection probability, in order to verify the performance of the intrusion detection method based on dimension reduction decision tree, this paper compares the proposed method with several common methods in the same data set.

The monitoring probability of different types of network intrusion is shown in Table 1. As can be seen from table 1, the intrusion monitoring system designed in this paper has the highest monitoring probability for probe bypass attack. When the percentage of data records is only 12.2%, 95.4% of the monitoring probability can be achieved. The adaptability of the system to NORMAL attack is the worst, but it can also achieve 90% of the monitoring probability.

Table 1. Different types of intrusion monitoring probabilities

Attack type	Record rate /%	Monitoring rate /%
NORMAL	78	90
DOS attack	15.6	91.4
USER-to-Root	24	90.2
PROBE BYPASS	12.2	95.4
Remote-to-Login	7.1	93

In the same data set, the probability ratio of intrusion detection between the proposed reduced dimension decision tree method and the three traditional methods is shown in Table 2. Compared with the three traditional methods, the intrusion detection probability of the method designed in this paper is significantly improved. When there are two attack types, the monitoring probability can reach 90.7%. When there are five attack types, the monitoring probability can reach 86.5%, which is more than 8% higher than the existing algorithm.

Table 2. Comparison of monitoring probabilities of different methods

Monitoring algorithm	Type 2 attack monitoring probability /%	Type 5 attack monitoring probability /%
SVM	82.4	78.4
Bayes	78.5	74.9
Pattern matching	80.1	77.1
Dimension reduction decision tree	90.7	86.5

6. Conclusion

In this paper, a network intrusion monitoring technology based on decision tree method is proposed. By adding intrusion monitoring module in the existing network and using dimension reduction decision tree method to identify data, the probability of intrusion monitoring is greatly improved and real-time abnormal data monitoring is effectively realized. Through the analysis of experimental results, we can see that the proposed algorithm has higher intrusion detection probability than the existing algorithm, and has better adaptability for different types of intrusion.

References

[1] Bi, M., Wang, A., Xu, J. (2018) Anomaly behavior detection of database user based on discrete-time Markov chain. Journal of Shenyang University of Technology., 1: 70–76.

[2] Zhang, B. (2017) Research on discrete event system for LAN attack. Electronic Science and Technology., 30: 169–172.

[3] Ru, B., He, X. (2017) Improved algorithm for intrusion suspected boundary problem in cloud computing environment. Journal of Shenyang University of Technology., 5: 70–76.

[4] Liu, H., Zhang, H., Bi, J., et al. (2018) Overview of distributed and cooperative network intrusion detection technologies. Computer Engineering and Applications., 8: 1–6, 20.

[5] Lin, W., Chen, M., Zhan, Y., et al. (2017) Researchon intrusion detection algorithms based on PCA and random forest classification. Information Network Security., 11: 54–58.

[6] Sun, W. (2014) Research of distributed intrusion detection based on network management system. Electronic Design Engineering., 22: 165–167.

[7] Wang, L. (2017) Network intrusion detection method based on data mining. Information Security Research., 3: 810–816.