# Research on Network Security Authentication Method Based on Data Mining Technology

Xiao-gang Ma[1(✉)] and Huan-yu Wang[2]

[1] Shandong Vocational College of Science and Technology, Weifang 261053, China
maxiaogang698@aliyun.com
[2] College of Earth Environment and Science, Southwest Jiao Tong University, Chengdu 611756, China

**Abstract.** In order to solve the problems of low authentication accuracy, long authentication time and poor authentication security in traditional network security authentication methods, this paper uses data mining technology to design a new network security authentication method. First, analyze the types of attacks on the network by illegal nodes on the network and the principles of authentication, and then mine the data to be authenticated through the binary network. In order to reduce the mining error, the acquired data is punished and integrated. In this process, in order to ensure the effective iteration of the data, the neural network algorithm in the machine learning algorithm is introduced for in-depth mining. The experimental results show that the authentication accuracy of this method can reach up to 98%, and the authentication time is always less than 2 s. The above results show that: after adopting this method, the network security performance can be improved.

**Keywords:** Network security · Data mining · Machine learning algorithm · Safety certification

## 1 Introduction

With the rapid development of information technology, the Internet has become an inseparable part of people's daily life, and has been widely used in various fields such as business, communications, entertainment, education, and personal daily affairs [1]. Therefore, a large amount of valuable information needs to be disseminated through the network. However, this also provides a way for various illegal intrusions against computer information systems [2].

At present, the applications of wireless public networks in people's lives are becoming more and more popular and diversified. The large amount of data and complex structure of the wireless public network means that it has characteristics such as complexity and dynamics. In this way, internal and external security risks have brought great pressure and challenges to network security work [3, 4]. And with the rapid development and extensive coverage of the Internet in recent years, the forms of cyber attacks have

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
S. Wang et al. (Eds.): IoTCare 2021, LNICST 415, pp. 19–29, 2022.
https://doi.org/10.1007/978-3-030-94182-6_2

become more diverse, and security threats have become deeper and deeper. The issue of network security needs urgent attention.

Reference [5] proposed a security authentication method for wireless sensor networks. In this method, the biometric key of the user is obtained by fuzzy extraction technology, and then the mutual authentication and session key negotiation between the legitimate user and the sensor node are completed with the help of the hash algorithm, which reduces the calculation cost. Finally, heuristic security analysis, BAN logic and ROM model are used to verify the security and effectiveness of the proposed scheme. Reference [6] proposed a computer network security routing method based on genetic algorithm. This method quantifies link security based on authentication access control and encryption mechanism, and then constructs a multi-objective secure routing model based on QoS parameters. According to the allocation of the common buffer pool and the minimum reserved bandwidth, the multi-objective secure routing model is selected to optimize the target. The objective function minimization problem of computer network secure routing is transformed into a maximization problem. Finally, the operator is selected for crossover and mutation, and the individual with the best fitness value is determined by genetic algorithm to realize the optimization of computer network secure routing. Reference [7] put forward the research method of encryption authentication and network security for Ethernet. The security of Ethernet is the research object of this method. On the basis of establishing the Ethernet model, AES-128 encryption algorithm and HMAC-SHA1 security authentication method and technology are introduced into Ethernet, so as to effectively prevent external intrusion, data jump and other risks, and further improve the security performance of Ethernet and bus network.

Although the above methods have achieved certain results, there are still problems such as low certification accuracy and long certification time. For this reason, this paper proposes a network security authentication method based on data mining technology. The design idea of the new method is as follows:

(1) Based on the analysis of the types of network attacks by illegal nodes and the principles of authentication, the data to be authenticated are mined through the binary network.
(2) In order to reduce the error of mining, the obtained data are punished and integrated. In this process, in order to ensure the effective iteration of the data, the neural network algorithm of machine learning algorithm is introduced to carry out deep mining.

## 2 Method Research

### 2.1 Data Mining Type Analysis

The process of data mining can be roughly divided into five stages: problem definition, data collection, data preprocessing, data mining implementation, and the interpretation and evaluation of mining results [8]. The principle of data mining is shown in Fig. 1.

Step 1: Problem definition. At this stage, a clear pre-achieved goal and user needs need to be determined.

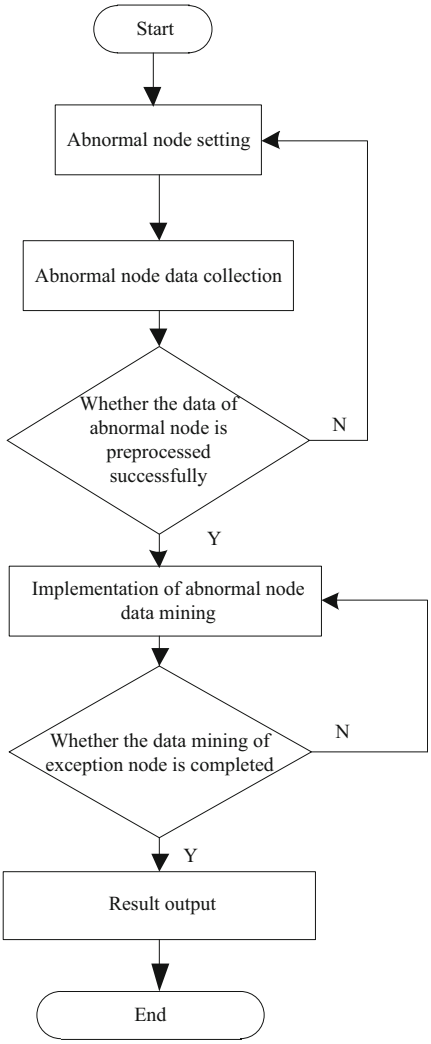


**Fig. 1.** Schematic diagram of data mining

Step 2: Data collection. After determining the target data objects involved in the data mining task, extract the data sets related to the mining task from the relevant data sources.
Step 3: Data preprocessing. Data preprocessing includes operations such as noise point elimination, data formatting, and data reduction.
Step 4: Data mining implementation. According to the task of data mining and existing methods (classification, clustering, outlier mining, etc.), data mining algorithms are selected, and the results are obtained by mining the obtained data.
Step 5: Interpretation and evaluation of results. At this stage, it is necessary to evaluate and analyze the results obtained by data mining in order to discover meaningful knowledge patterns.

Before adopting data mining technology, it is also necessary to have a detailed understanding of the data that needs to be mined. The types of risk data included in the network security certification are many and complex. Failure to effectively classify them will result in failure of the certification security or affect its safe operation. Therefore, this article first effectively divides the data types that need security authentication on the network, and implements effective factor design methods for them. The types of data that need to be mined for network security certification include the following:

(1) The continuous or offline value of the real number field is the most common target data form in the data mining method:
(2) Discrete integer field values are also quite common. In addition to expressing the meaning of their own attributes, discrete integer field values are also used to escape character data and enumerated data to support mining work.

### 2.2 Data Mining Algorithm Analysis

In network security certification, data mining technology is an important technology to improve its security. There are many mining algorithms in this technology. In order to achieve network security certification, this article analyzes the data mining algorithm in detail.

The binary network mining algorithm is an algorithm that can reduce the algorithm space. It determines the data relationship in the network security authentication through the network topology relationship [9], which can be expressed as:

$$A_i = \frac{1}{n}\sum_{i=1}^{n}\frac{a_i b\delta}{n(b_i)} \tag{1}$$

$$f(x) = \sum_{j=1}^{m} A_i f'(x) \tag{2}$$

Due to some errors in the data relationship obtained above, this paper punishes the trust relationship between data in network authentication through the high-order relationship, the process is as follows:

$$f(x) = \left(Ab^2 + Bb^3\right) f(a_i) \tag{3}$$

$$f'(x) = \sum_{i=1}^{n}\left(Ab^2 + Bb^3\right) f(a_i) \tag{4}$$

On the basis of binary network algorithm mining, it is believed that there is a greater degree of association and similarity between network security authentication data. Therefore, it is necessary to process the similarity of the network security authentication data, namely:

$$A_i = \frac{1}{k}\sum_{i=1}^{m}\frac{a_i\delta_i}{k(a_i\delta_i)} \tag{5}$$

In the process of network security authentication, due to the constant changes in the network environment, the number of data in the network that needs to be authenticated and the degree of data security need to be updated constantly, which increases the complexity of the authentication process, resulting in poor results and efficiency of authentication. For this reason, machine learning algorithm is introduced in the authentication process of network authentication data.

Machine learning algorithms are more familiar with the intelligent algorithm of neural network. By constructing the network level, the network security authentication data is reasonably trained, and the weight information is transmitted through the key cells in the network, namely:

$$C(x) = \sum_{x=1}^{n} p_n\left(v_n^t X\right) \tag{6}$$

After the above-mentioned security authentication data weight is transmitted, the information in the hidden layer is effectively converted, and the processing method of the intermediate layer can be obtained:

$$u_i = \frac{1}{1 + \exp\left(\sum_{i=1}^{n} w_i + \theta_j\right)} \tag{7}$$

After the hidden information is revealed, the non-secure data that needs to be authenticated can be obtained after valid authentication, namely:

$$p(a, b) = \frac{1}{1 + l_i uy} \tag{8}$$

Logical regression and support vector machines both train the parameters of the model through feature samples. When the parameters approached the optimal training solution, the training was stopped, and the results were outputted according to the model of unknown sample vectors. Typical logistic regression training methods include Newton's method, conjugate gradient method, quasi Newton's method and so on. At present, the Hessian matrix is used to solve the logistic regression model through the quasi-Newton method, which overcomes the disadvantage that the parameter gradient descent direction of the traditional Newton method cannot converge along the direction of the optimal solution. At the same time, the traditional Newton method is transformed into the least square iterative optimization process, which greatly improves the convergence speed of network security authentication [10].

In the research of data mining technology, the mining of non - secure data is the key. Therefore, the fully authenticated data is first mined through a binary network. In order to reduce the error of mining, we punish the acquired data, and automatically introduce the neural network algorithm in machine learning algorithm. Through deep mining, we complete the deep mining of data in the process of network security authentication.

### 2.3 Data Analysis and Network Security Certification

Network security certification is actually a process of accumulating normal data node data and non-secure data and ordering the certification process. The schematic diagram of the network security authentication node is shown in Fig. 2.

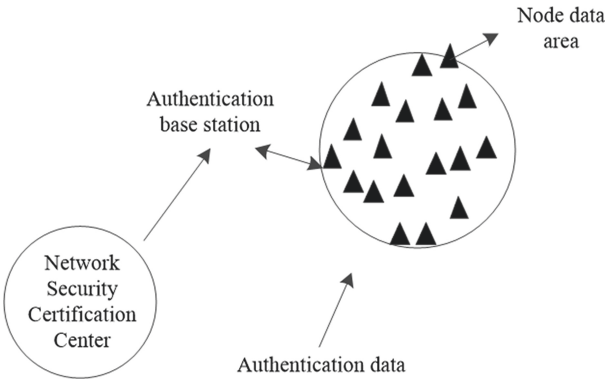


**Fig. 2.** Schematic diagram of network security authentication node

(1) Identity password stealing attack: The illegal node obtains the identity password of the legal node to access the system resources authorized to the legal node.
(2) Traffic analysis attack: The illegal node analyzes the information exchanged between the communicating parties, trying to judge or recover the original information.
(3) Retransmission attack: The illegal node intercepts the information sent by the sender, and then retransmits it to the receiver.
(4) Modification or forgery attack: The illegal node intercepts the information sent by the sender, replaces or modifies the information and then transmits it to the receiver, or pretends to be a legitimate node to send information [11].
(5) Denial of service attacks: prevent legitimate nodes from legally managing and using system resources.
(6) External attacks and internal attacks: External attacks refer to attacks initiated by external nodes when WSN communicates with external nodes. This external node is a newly added node.

In response to the above-mentioned attacks on illegal nodes, network security services should provide the following authentication functions:

(1) Reliability: The source of the information is credible, that is, the recipient of the information can confirm that the information obtained is not sent by a counterfeit.
(2) Integrity: The integrity of the information is required to be guaranteed during the transmission process, that is, the information recipient can confirm that the acquired information has not been modified, delayed or replaced during the transmission process [12].
(3) Non-repudiation: It is required that the sender of the information cannot deny the information it sends, and similarly, the receiver of the information cannot deny the received information.
(4) Access control: Illegal nodes cannot access the system resources of the network, and legal users can only access the resources authorized and designated by the system.

## 3 Experiment Analysis

In order to verify the effectiveness of the network security authentication method under the above-designed data mining technology, the method in this paper is compared with the method in Reference [5] and the method in Reference [6]. Through multiple iterations, the verification and analysis of the effectiveness of the method in this paper are realized.

### 3.1 Lab Environment

In the design of the experimental environment, since network authentication is an abstract process, the commonly used MATLAB platform is used for simulation analysis. The experimental operating system is WINDOWS 10.

### 3.2 Experimental Parameter Design

In order to meet the requirements of the experiment, the parameters of the experiment are designed, and the specific parameters are shown in Table 1.

**Table 1.** Experimental parameter design

| Parameter | Numerical value |
|---|---|
| Network security node/a | 1000 |
| Number of secure nodes | 600 |
| Number of non-secure nodes/piece | 400 |
| Authentication interval/s | 0.5 s |
| Certification error/% | <2 |
| Number of certification tests/time | 100 |

According to the design of the above-mentioned experimental environment and parameters, the accuracy of the authentication node data, the authentication time-consuming and the post-authentication network security performance of the method in this paper, the method in Reference [5] and the method in Reference [6] are compared.

### 3.3 Analysis of Results

**Analysis of the Accuracy of Node Authentication in Different Methods**

The network security certification is mainly to authenticate the data of non-secure nodes with threats. Therefore, in the experiment, the method in this paper, the method in Reference [5] and the method in Reference [6] are compared to authenticate the sample security node data. In the experiment, the results are obtained multiple times to ensure the accuracy of the obtained results. The results obtained are shown in Fig. 3. Shown.

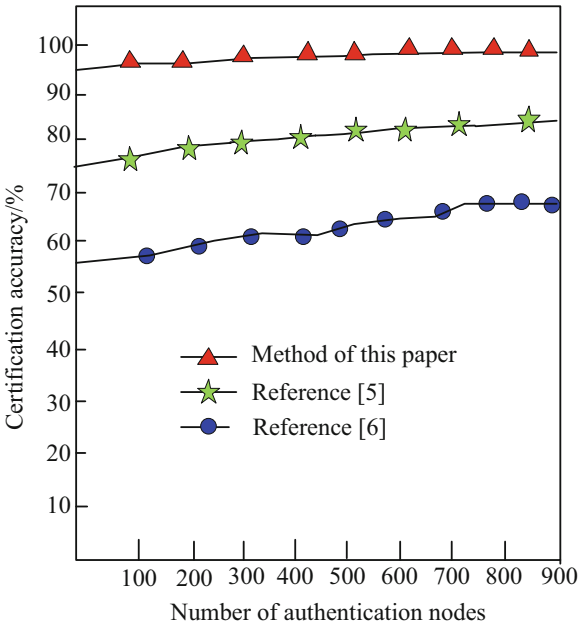


**Fig. 3.** Comparison results of node authentication accuracy of different methods

Analyzing the data in Fig. 3, it can be seen that with the continuous change of the number of authentication network nodes. The method in this paper, the method in Reference [5] and the method in Reference [6] have certain differences in the accuracy of the sample network node authentication. Among them, when the number of authenticated data nodes is 300, the accuracy of the method in this paper for authentication of sample network nodes is about 95%, and the accuracy of the method in Reference [5] for authentication of sample network nodes is about 87%, Reference [6] The accuracy of the method for authenticating sample network nodes is about 56%; when the number of authenticated data nodes is 600, the accuracy of the method in this paper for authenticating sample network nodes is about 96%. The method in Reference [5] is effective for authenticating sample network nodes. The accuracy is about 80%, and the accuracy of the Reference [6] method for the authentication of the sample network node is about 62%; when the number of authentication data nodes is 900, the accuracy of the method for the authentication of the sample network node is about 97%. The Reference [5] method has an accuracy of about 89% for the sample network node authentication, and the method in this paper has an accuracy of about 65% for the sample network node authentication. It can be seen that the authentication accuracy of the three methods shows an upward trend, but In comparison, the accuracy of the method in this article is higher than the method in Reference [5] and the method in Reference [6].

This is because the method in this paper determines the node attributes of authentication before authentication, and uses data mining technology to dig it deeply, which improves the authentication accuracy of the proposed method and has certain advantages.

**Time-Consuming Analysis of Node Authentication in Different Methods**

In network security certification, the time-consuming certification can show the performance of the network. If the certification time is shorter, the network performance is better. For this reason, comparing the time consumption of the method in this paper, the method in Reference [5] and the method in Reference [6] to authenticate the sample secure node data, the results obtained are shown in Fig. 4.

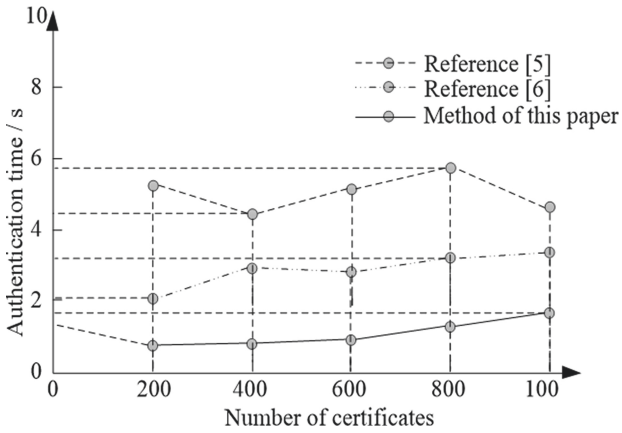


**Fig. 4.** Time-consuming analysis of node authentication by different methods

Analyzing the data in Fig. 4, it can be seen that under the same experimental environment, there are certain differences in the time-consuming authentication of the sample network nodes using the method of this paper, the method of Reference [5] and the method of Reference [6]. It can be seen from the curve in Fig. 4 that the verification time of the method in this paper is always less than 2 s, but the safety authentication time of the other two methods is lower than that of the method in this paper. This is because the method in this paper uses artificial intelligence algorithms to mine hidden data during data mining, which reduces the complexity of mining and improves the effectiveness of the method in this paper.

**Security Analysis of Different Method Nodes after Authentication**

The purpose of network security certification is to improve the security performance of the network. Therefore, the experiment analyzes the network security performance of the sample network nodes after the method of this paper, the method of reference [5] and the method of reference [6]. Among them, the higher the degree of security, the more secure the network performance. The results obtained are shown in Table 2.

**Table 2.** Comparison of security after authentication of different methods of nodes (%)

| Number of iterations/time | Method of this paper | Method of Reference [5] | Method of Reference [6] |
|---|---|---|---|
| 10 | 95 | 89 | 87 |
| 20 | 95 | 87 | 88 |
| 30 | 94 | 87 | 89 |
| 40 | 94 | 85 | 87 |
| 50 | 95 | 85 | 87 |
| 60 | 95 | 89 | 87 |
| 70 | 92 | 87 | 85 |
| 80 | 95 | 89 | 85 |
| 90 | 94 | 86 | 87 |
| 100 | 92 | 85 | 84 |

Analyzing the experimental data in Table 2, it can be seen that with the increase of the number of experimental certifications, the method of this paper, the method of Reference [5] and the method of Reference [6] have changed the security performance of the sample network node after authentication. Although the three methods have certain advantages in the security performance of the network security certification, the security performance of the method in this paper is better than that of the Reference [5] method and the Reference [6] method. This is due to the method in this paper. Continuously authenticate the mined data, strengthen the authentication details of non-secure data, and then enhance the effectiveness of the method in this paper.

## 4 Conclusion

Authentication of network security is the key to protect users' rights and interests. Therefore, this article uses data mining technology to effectively authenticate network security. Firstly, by analyzing the types of node data in the network and its characteristic attributes, in-depth mining of the unsafe nodes that exist through data mining technology, the research of network security authentication is realized. Compared with traditional methods, this method has the following advantages:

(1) The accuracy of using this method for network security node authentication is about 98%, which can effectively authenticate network security nodes;
(2) The time-consuming for network security node authentication using this method is always less than 2 s, which can quickly and effectively authenticate network security nodes;
(3) The network security performance is improved after the method of this paper is used to authenticate the network security node, and the network security node can be effectively authenticated.

Although at the present stage, the method in this paper has achieved certain success and has certain help for network security authentication, there is still room for improvement. Therefore, in the future research center will carry out more in-depth research, and constantly enhance the strength of network security certification.

## References


1. Hao, T.: The identity authentication of Wi-Fi system based on network security. Annals of telecommunications - annales des télécommunications **41**(10), 1–8 (2020)
2. Muqtadir, S.: Network security based on chaotic maps based authentication in vanets. J. Appli. Sci. Computa. **31**(5), 44–57 (2019)
3. Gao, S., et al.: A lightweight fingerprint-based device authentication architecture for wireless industrial automation networks. In: 2019 1st International Conference on Industrial Artificial Intelligence (IAI). IEEE (2019)
4. Guiga, L., Roscoe, A.: Neural network security: hiding CNN parameters with guided grad-CAM. In: 6th International Conference on Information Systems Security and Privacy (2020)
5. Alexander, R.: Using the latin square design model in the prioritzation of network security threats: a quantitative study. J. Inf. Secur. **11**(2), 92–102 (2020)
6. Haider, N., Baig, M.Z., Imran, M.: Artificial intelligence and machine learning in 5G network security: opportunities, advantages, and future research trends. Research Gate **23**(9), 303–319 (2020)
7. Zhao, M., et al.: CPTmilODIM method for bipolar fuzzy multi˚ ttribute group decision making and its application to network security service provider selection. Int. J. Intell. Syst. **17**(4), 18–26 (2021)
8. Ahmad, M.B., Agarwal, P.: Viability of adaptive network security exercising tradeoff between performance and security. Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science) **13**(5), 893–900 (2020)
9. Shuai, L., Liu, G.C., Zhou, H.Y.: A robust parallel object tracking method for illumination variations. Mob. Netw. Appl. **4**(1), 5–17 (2019)
10. Zhao, D., Song, H., Li, H.: Fuzzy integrated rough set theory situation feature extraction of network security. J. Intell. Fuzzy Sys. (1), 1–12 (2021)
11. Liu, S., Liu, G.C., Zhou, H.Y.: A robust parallel object tracking method for illumination variations. Mob. Netw. Appl. **24**(1), 5–17 (2019)
12. Deng, C.H., et al.: Research on attribute security verification method of space network based on complex network theory. Electr. Desi. Eng. **28**(3), 79–83 (2020)