Research on the Classification Method of Network Abnormal Data

Bozhong Liu(✉)

School of Electronic and Information Engineering, Guang'an Vocational Technical College, Guang'an, China
liubozhong77@163.com

Abstract. As people use the network more and more and release more and more personal information to the Internet, it also caused the leakage of personal information. According to the above background, the optimization research on the classification detection method of network anomaly data was proposed. Correlation analysis was carried out for the conventional algorithm, and the related model was constructed. A new algorithm was proposed to detect the network anomaly data to improve the processing ability of the network anomaly data. The experimental data showed that the proposed network anomaly data classification detection optimization algorithm improved the processing range by 31% when processing abnormal data, and the efficiency of processing data was increased by 36%. It proved the effectiveness of the new method and provided a theoretical basis for the processing of future abnormal data.

Keywords: Network anomaly · Data classification · Detection method · Improved design

1 Introduction

Under the development of modern science and technology, people's information security is also reduced. Many information is badly bought and sold online, and people's privacy cannot be effectively guaranteed. This is not conducive to the improvement of network security, causing great damage to personal information, affecting the security of personal information. In this context, the network technology department has also improved the network data anomaly detection, and fully enhanced the algorithm to improve the security of information. By making relevant predictions about the data that may cause attacks in advance, the security of information is improved, and relevant prevention of virus attacks is carried out in advance. In the construction of firewall, its performance should be fully improved. On the basis of traditional prevention, protection of personal privacy information should be strengthened to improve personal security performance. In the process of technical improvement, the research on the classification detection method of network abnormal data should be strengthened. There are different methods of classification detection for different network intrusion methods to maintain the data, so as to protect the data well [1]. At present, China's research on the classification detection method of network abnormal data is still in its infancy, and the technology is not very mature. It is not well

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2019
Published by Springer Nature Switzerland AG 2019. All Rights Reserved
G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 254–262, 2019.
https://doi.org/10.1007/978-3-030-36402-1_27

protected against malicious online information attacks, resulting in the loss of personal information, which has a great negative impact on individuals. Therefore, the protection of personal information should be constantly strengthened, the relevant technology of network abnormal data classification detection method should be improved continuously, the operation of its algorithm should be improved, and its safety performance should be increased. It provides effective technical support for the protection of personal information and future security performance, and guarantees the privacy of individuals. As shown in Table 1:

Table 1. Attribute representation of abnormal traffic data

Abnormal	The source IP address	Destination IP address
DDOS	Scattered	Concentrated
Attack	Concentrated	Concentrated
The worm	Concentrated	Scattered
IP scanning	Scattered	Concentrated
Port scanning	Scattered	Scattered

2 Classification and Detection of Network Abnormal Data

2.1 The Method of Artificial Intelligence Is Used for Relevant Detection

Using artificial intelligence to classify network abnormal data is the trend of network abnormal data detection in the future. The relevant procedures are edited manually and the machine is used for the related detection operations. The relevant detection can be improved through the related automatic calculation, saving manpower, improving the efficiency of detection and reducing the working pressure of people [2]. At present, with the improvement of science and technology, the attack technology of network information is further improved. A series of simple methods used in the past to monitor data have been unable to meet the needs of network abnormal data detection. The traditional method is relatively simple. Now, the detection method for complex data should be updated so that it can carry out stronger detection for the update method of complex network abnormal data. Improve related processing, use artificial intelligence to perform related detection, and edit complex programs in advance so that they can cope with complex problems and solve them accordingly. The artificial intelligence can continuously update and learn according to the relevant complexity, can automatically deal with related problems, improve the processing technology for information, and improve the efficiency of processing problems. It makes the processing of information more difficult, and the scope of processing related problems is expanded accordingly, and various complex data can be solved in a related manner. Artificial intelligence can link related abnormal data, so that the processing of network abnormal data can be solved in a series of ways. It has changed the traditional solution to only a single abnormal data, and improved the processing efficiency of abnormal data [3]. As shown in Fig. 1:

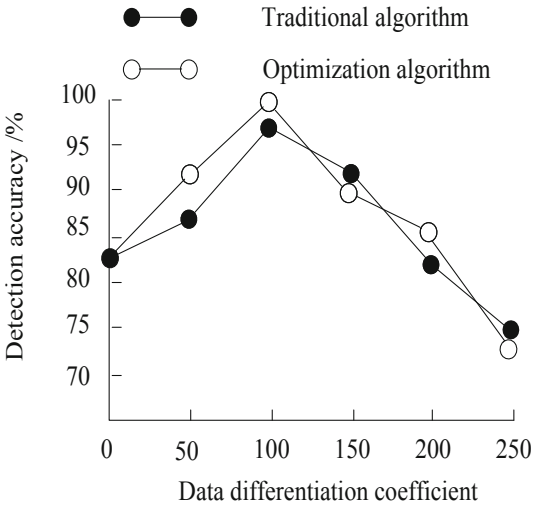

Fig. 1. Algorithm detection results

2.2 Data Mining for Correlation Detection

When using the data mining method for related detection, it is necessary to conduct a large amount of mining of the originality of the abnormal data to find out the root cause of the problem. Correlate the root cause of the abnormal data and solve the problem of abnormal data processing. Using this method to detect and analyze related data, the source of the relevant abnormal data can be solved and analyzed, and the source of the virus is processed from the root cause. On this basis, the information processing process can be fully simplified, and it is no longer necessary to search for abnormal data one by one, which wastes a lot of time. When performing related detection through data mining [4], not only can the problem be processed, but also relevant differential detection can be performed by analyzing the data. Analyze the degree of abnormality of the data, compare it with the previous data, compare the source and difference of different abnormal data, in order to make more reasonable editing of the development and update of the algorithm, and simplify the processing of the abnormal data afterwards. In the response to the processing of various complex information, the way of data mining is particularly important, which can highlight the handling of related complex problems and solve the problems from the root. As shown in Fig. 2:

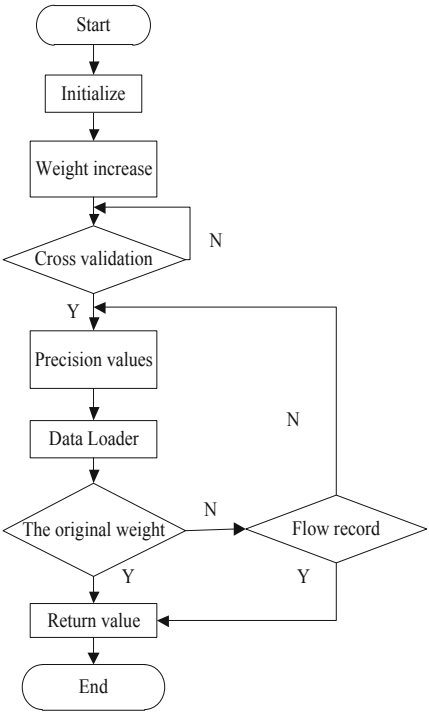

Fig. 2. Algorithm general flow

2.3 Using Information Entropy Method to Detect Abnormal Data

The method of information entropy is to detect the details of the anomaly information, and use the relevant algorithm to further detect and analyze the abnormal data in detail. The detection of network problems is not limited to the processing of the surface, but the details are analyzed. On this basis, the editing of the relevant algorithms is performed according to the details. On this basis, the relevant calculations are carried out, and the details of the problem are analyzed in detail, so that the details are handled very delicately, not only the surface is processed, but the abnormal data is repeated and cannot be cured [5]. In the process of using the information entropy method to detect abnormal data, it is necessary to continuously update the abnormal data, and continue to perform detailed operations on the algorithm to improve the processing of details and improve personal information security. Under the premise of safe handling of information, it is guaranteed that individuals can effectively protect the security of individuals and strengthen the processing of information. The relevant abnormal data is detected in advance, and the detailed problems are processed firstly on the basis of the detection, and the security of the personal information is improved on the basis of the processing [6]. As shown in Fig. 3:

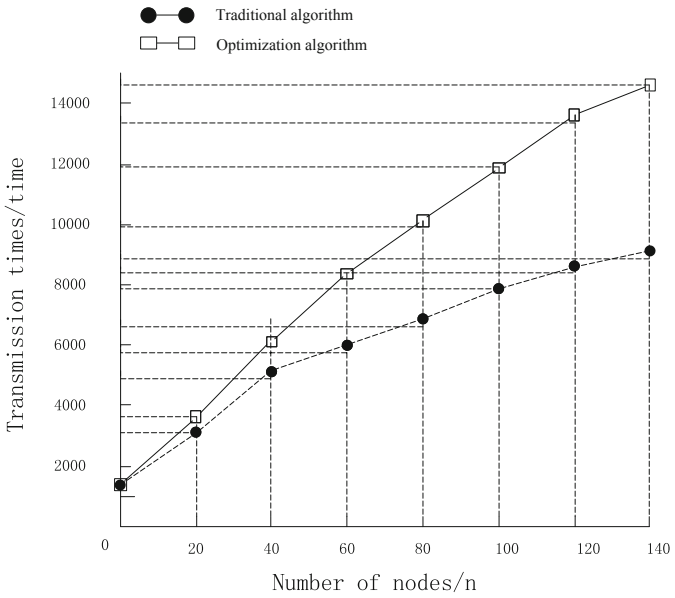

Fig. 3. Comparison of experimental results of different algorithms

3 Case Analysis

In order to ensure the effectiveness of the network anomaly data classification detection method proposed in this paper, it is necessary to analyze its effectiveness. In the test process, different classification detection methods are used as the test object, and the optimization ability of the detection method is analyzed in detail, and the test objects of different specifications are used for simulation analysis. In order to ensure the accuracy of the experimental process, the parameters of the experiment should be set [7]. In this paper, the test data is used as the test object, and two different design methods are used to conduct the classification test. The simulation results are analyzed. Since the analysis results obtained in different methods are different from the analysis methods, it is necessary to ensure the consistency of the test environment parameters during the test. The test data setting results in this paper are shown in Table 2:

Table 2. Different methods for detecting network result data tables

Abnormal characteristic number	K mean accuracy	Accuracy of neural network detection
12	81	77
21	83	80
31	79	78
41	74	77
51	78	81
15	74	78
37	76	79
42	77	91

3.1 Analysis of Results

Algorithm:

$$\sum_{j=1}^{n} z_j b_j = 1, b_j > 0 \tag{1}$$

In the form, $\sum_{j=1}^{n} z_j b_j$ represents the label information of network abnormal data subblock b_j; z_j represents Acquired j an effective network abnormal data block; n represents the total number of data blocks. Perform related update operations according to the above algorithm to optimize the related algorithms:

$$(z)y = sign\left(\sum_{j=1}^{n} b_j z_j + c\right) \tag{2}$$

In the form, $(z)y$ represents the optimized network abnormal data feature fusion error domain; c represents the center of mass of the sample. Using this algorithm, effective monitoring of abnormal data can be realized, and analysis and processing of abnormal data can be optimized [8]:

$$\max \sum_{j=1}^{n} \theta_k = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} z_j b_j (z) y \tag{3}$$

In the form, $\max \sum_{k=1}^{m} \theta_k$ represents the network anomaly data detection model.

As shown in Table 3:

Table 3. Sample data statistics

Number of experiment	Sample quantity	Identifiable abnormal data characteristics
1	324	67
2	342	45
3	532	55
4	123	37
5	432	27
6	124	34
7	864	43

According to the method described above, the updated algorithm is used to analyze and process the abnormal data, reduce the existence of abnormal data, improve the security of information, and strengthen the protection of information [9]. J represents the correlation coefficient of the processed information, k represents the analysis of the coefficient, and m represents the security of the data, b represents the range of data processing, z represents the range of calculations used by the running data, and the detection of abnormal data [10]. As shown in Fig. 4:

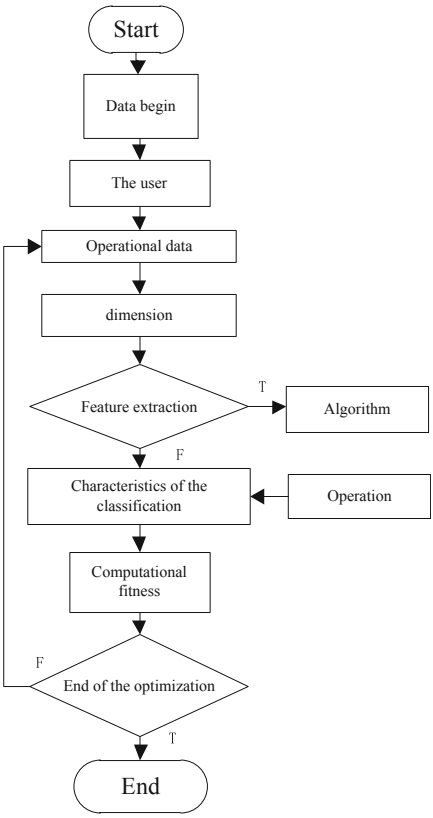

Fig. 4. Abnormal data detection flow chart

In the process of performing related operations, as shown in Fig. 5:

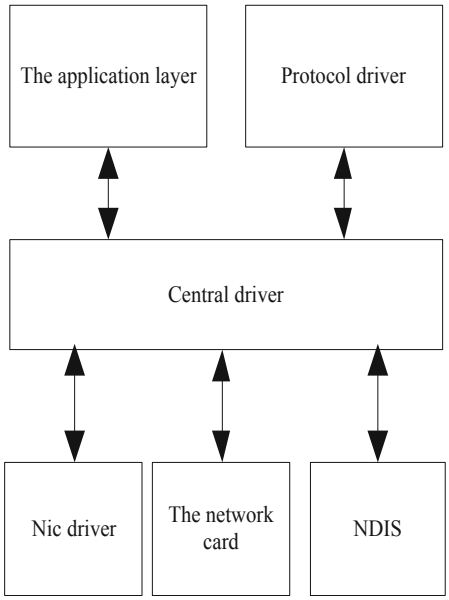

Fig. 5. Driver structure

4 Conclusions

In this paper, the network anomaly data is analyzed and processed in the context of informationization. Based on the analysis of the traditional algorithms, the related algorithms are optimized. Based on the classification and detection of network anomaly data, the optimization algorithm is processed. A large number of algorithms are used to process related data, which ensures the security of user information, improves the protection of user information, and enhances the information security for users.

References

1. Nemanja, B., Duan, K.Z.: Improved real-time data anomaly detection using context classification. Expert Syst. Appl. **23**(14), 23–36 (2017)
2. Sun, Y., Wong, A.C., Kamel, M.S.: Classification of imbalanced data: a review. Int. J. Pattern Recognit. Artif. Intell. **12**(3), 501–520 (2017)
3. Wang, B., Zhao, Y., Hu, F., et al.: Anomaly detection with subgraph search and vertex classification preprocessing in Chung-Lu random networks. IEEE Trans. Signal Process. **66**(20), 5255–5268 (2018)
4. Hussain, J., Lalmuanawma, S., Chhakchhuak, L.: A two-stage hybrid classification technique for network intrusion detection system. Int. J. Comput. Intell. Syst. **9**(5), 863–875 (2016)
5. Xin, D.U., Huang, X., Hongga, L.I., et al.: Research on classification of plant community using projection pursuit learning network algorithm on high resolution remote sensing images. J. Geo-Inf. Sci. **18**(1), 124–132 (2016)

6. Al-Obeidat, F., El-Alfy, E.S.M.: Hybrid multicriteria fuzzy classification of network traffic patterns, anomalies, and protocols. Pers. Ubiquitous Comput. **2**, 1–15 (2017)
7. Nguyen, H.A., Choi, D.: Application of data mining to network intrusion detection: classifier selection model. Lecture Notes in Computer Science **36**(12), 230–241 (2017)
8. Pang, M., Hao, X.: Traffic flow prediction of chaos time series by using subtractive clustering for fuzzy neural network modeling. In: Second International Symposium on Intelligent Information Technology Application, vol. 2(13), pp. 12–26 (2017)
9. Jerome, R.B., Hätönen, K.: Anomaly detection and classification using a metric for determining the significance of failures. Neural Comput. Appl. **28**(6), 1–11 (2016)
10. Ying, W.: Wireless network traffic abnormal data information detection simulation. Comput. Simul. **34**(9), 408–411 (2017)