Hindawi
Journal of Electrical and Computer Engineering
Volume 2023, Article ID 9759365, 1 page
https://doi.org/10.1155/2023/9759365

Retraction

Retracted: Application of Fast P2P Traffic Recognition Technology Based on Decision Tree in the Detection of Network Traffic Data

Journal of Electrical and Computer Engineering

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Zheng and J. Li, "Application of Fast P2P Traffic Recognition Technology Based on Decision Tree in the Detection of Network Traffic Data," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 8320049, 11 pages, 2022.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 8320049, 11 pages
https://doi.org/10.1155/2022/8320049

Research Article

Application of Fast P2P Traffic Recognition Technology Based on Decision Tree in the Detection of Network Traffic Data

Lin Zheng and Junjiao Li

Hengshui Open University, Hengshui 053000, China

Correspondence should be addressed to Junjiao Li; lijunjiao0402@163.com

Received 16 March 2022; Revised 27 April 2022; Accepted 3 May 2022; Published 3 June 2022

Academic Editor: Wei Liu

With the rapid development of large-scale enterprise informatization construction, the network scale has become huge and complex, and the data traffic carried by the network is increasing. Accurate network traffic identification is the basis of network management and is of great significance to enterprise informatization construction and operation and maintenance. In response to the network operation and maintenance requirements of large enterprises, this paper analyzes the network architecture and network traffic distribution of large enterprise groups from the perspective of enterprise network operators and introduces the current operation and maintenance process of enterprise network performance failures. Maintenance process optimization and reengineering are carried out to plan and find out the shortcomings of the current process links and put forward corresponding solutions. Based on the research of traffic identification in recent years, a fast P2P network traffic anomaly identification algorithm based on decision tree model is proposed, which improves the efficiency and accuracy of network traffic application identification and network traffic anomaly data identification.

1. Introduction

In recent years, the rapid development of network technology and the continuous enrichment of network application make the network become an indispensable part of people's life. Compared to networks from a decade ago, network users are getting larger, network structures are getting more complex, and network bandwidth is getting higher. The continuous development of the Internet has gradually changed the human way of life, and all aspects of human activities have begun to be completed on the Internet, such as the increasingly popular online shopping, the more and more extensive Internet communication and streaming media sharing. Another convenience for the development of the network is the network storage of information. More and more people begin to submit their personal data and daily behavior information to the Internet. The storage of personal information online brings convenience to people but also brings serious challenges to the network security. Complex network structure and diverse network applications make all kinds of security vulnerabilities appear, seriously affecting the security of individuals and enterprises; network security becomes increasingly important, closely related to the interests of individuals, governments, and enterprises, so necessary measures are to be taken to prevent network attacks and real-time monitoring.

The overall process of flow abnormal detection is divided into two parts:

(1) The presentation of traffic anomaly, through the transformation of domain, subspace, information entropy, and other methods to better show network traffic anomaly.

(2) For exception detection, the exception expressed in step (1) is detected by data mining, intelligent algorithm, machine learning, statistical analysis, and other methods.

Given the excellent performance of cross entropy in anomaly representation of network traffic anomaly and

decision tree classification method, this paper builds a decision tree model based on cross entropy to analyze and detect network traffic anomaly. In this way, the network operation and maintenance efficiency is improved, a large amount of equipment investment and operation and maintenance costs are saved, and the undefined network application types in the network can be found, and the research direction of network traffic identification can be enriched.

2. State of the Art

In the early twentieth century, Sen et al. [1] proposed a method to extract the features of the peer-to-peer (P2P) application. The key is to use information such as document and packet characteristics to identify the application signature of a specific P2P traffic, and to redevelop an online filter that can identify the signature for traffic identification. After testing, it can accurately identify P2P traffic at high bandwidth. This method belongs to the traffic identification method based on Deep Packet Inspection (DPI) deep package detection, with high identification accuracy, many identification protocols, and high effectiveness. However, the limitation is also large, because the need to check the payload of packets one by one, which involves user data privacy issues, and the method fails to identify encrypted traffic and unknown application types [2].

In order to overcome the problem that DPI technology can not identify encrypted traffic, Deep Flow Inspection (DFI) deep flow detection technology arises therefrom. DFI is different from DPI. It is an application recognition technology based on traffic behavior. The research object of DFI is no longer the data packet itself, but the data flow. DPI realizes traffic identification by analyzing and studying the flow state differences of different application streams. Recently, machine learning has been widely used in the field of network traffic recognition. Este et al. proposed a network traffic classification method based on Support Vector Machine (SVM) support vector machine, through which the SVM classifier performs correctly with only a few hundred samples. In China, Wang Zhenhua et al. [3] proposed, on the basis of deeply studying the Skype communication mechanism, the Skype traffic identification scheme based on comprehensive statistical features, and the traffic identification system model is realized and verified. Yuan Huabing [4] proposed a P2P network traffic recognition model based on DA-Elman machine learning. By taking packet feature attributes such as TCP traffic ratio as the input of DA-Elman model and network traffic type as the output of DA-Elman, the accuracy of P2P network traffic identification is effectively improved, providing new methods and ways for the identification of P2P network traffic. In recent years, the domestic industry manufacturers have also made great progress in the field of application recognition. In September 2019, Gartner measured and evaluated relevant network security vendors from both vision foresight and strategic execution dimensions in its latest Network Firewall Magic Quadrant Report [5, 6].

Based on the host attribute identification method, its detection content is the domain name, IP address, or port number, has a fast detection speed, high accuracy, and strong real-time, and can identify encryption applications and low performance requirements for equipment, and the disadvantage is high maintenance cost. Based on the identification method of deep package detection, the detection content includes application layer feature fields, ports, and package length, which has fast recognition speed, high accuracy, strong real-time performance, which can not identify encryption applications, high requirements for equipment performance, and high maintenance cost. In the identification method based on machine learning, the detection content includes packet size, quantity, and interaction time.

3. Modeling of P2P Fast Traffic Identification Based on Decision Tree

3.1. Decision Tree Algorithm Model Analysis. Decision tree classification algorithm is a machine learning algorithm, which is a process of automatically mining a set of regular patterns that are equally effective for data other than the training sample. In addition to decision tree classification algorithms, machine learning algorithms include naive Bayesian classification algorithms, support vector machine-based classification algorithms, neural network algorithms, k-means clustering algorithms, and fuzzy taxonomy. The decision tree constructed by a classification algorithm is a classification model in which each branch of the model represents a mapping of an attribute of an object to a certain value or type of value of an attribute. In the decision tree, each non-leaf node represents a judgment condition, and each judgment condition corresponds to an object attribute, while each branch path represents the attribute value that satisfies the judgment condition. Each leaf node L in the tree represents a set of values, and each value in the set satisfies every judgment condition in the path passing from the root node to the leaf node L. The decision tree is constructed from the root node, first selecting appropriate attributes to divide the sample set into several subsets, each forming a branch node, and then dividing each branch node until all samples in the node are consistent, or some end condition is met. Decision tree construction generally includes the following two steps: (1) the generation of the decision tree, that is, the process of generating the decision tree by using the training sample set; (2) the pruning of the decision tree, which needs to be verified, corrected, and modified after generating the decision tree. The algorithm proposed in this paper uses the method of multiple decision tree integration, each of which is a weak decision tree, does not appear overfitting, and therefore does not involve the pruning problem of the decision tree. The input sample set form of the decision tree construction algorithm is as follows:

$$I = \left\{\left(A_{00} \ldots, A_{0j} \ldots, A_{0n}, T_0\right) \ldots \left(A_{i0} \ldots, A_{ij} \ldots, A_{in}, T_i\right) \ldots\right\}. \tag{1}$$

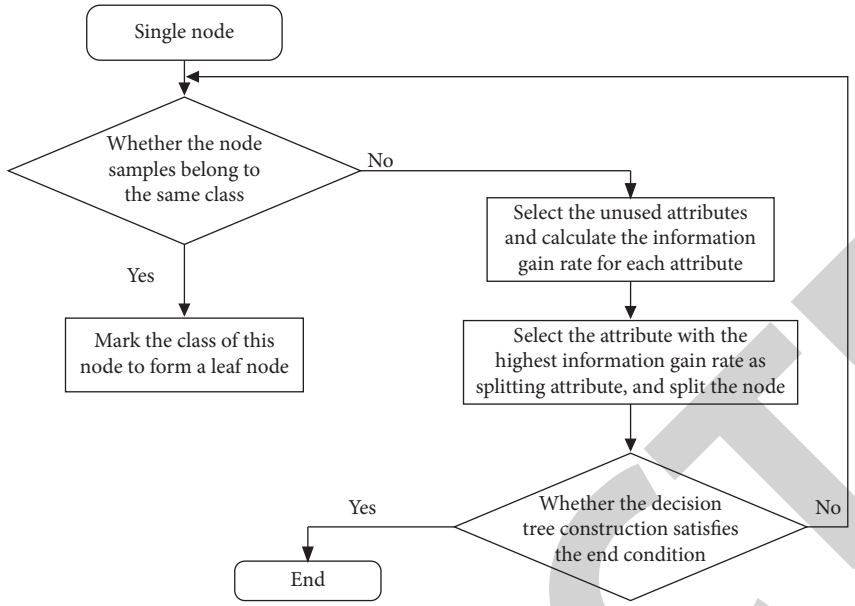

Figure 1: Decision tree construction process.

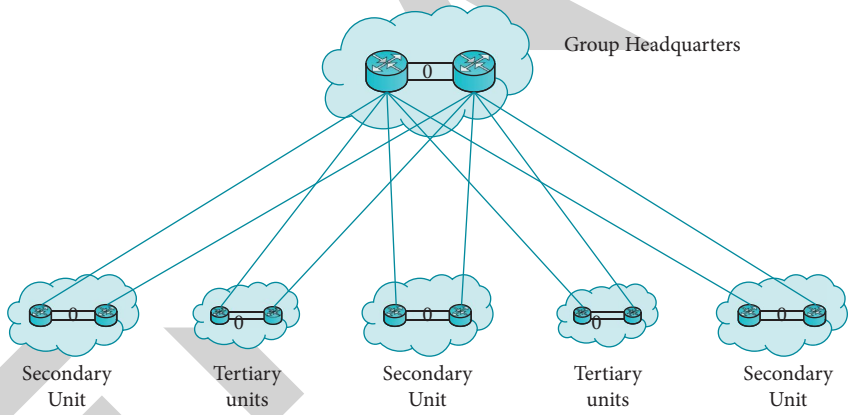

Figure 2: Star topology.

A_{ij} represents the value of the j th attribute of the i-bar sample in the set, T_i, the type marker of the sample in bar i. The result of decision tree construction is a binary or multiple fork tree, generally used for data sets whose properties are all Boolean logical judgments. The general flow of decision tree construction is shown in Figure 1:

Compared with other algorithms, the decision tree classification algorithm has high efficiency. The decision tree only needs to be constructed once and used repeatedly. The maximum number of calculations for each prediction does not exceed the depth of the decision tree. The time complexity of the decision tree algorithm is small and is the logarithm of the data points used to train the decision tree. And the decision tree algorithm has a small error rate, and the conclusion is more accurate. Therefore, this paper chooses the decision tree algorithm to apply it to the network traffic data anomaly identification and detection.

3.2. Network Traffic Identification Business Analysis and Operation and Maintenance Analysis. In order to demonstrate the necessity of improving the accuracy of the network traffic data anomaly identification and detection algorithm in network operation and maintenance, and to demonstrate how to optimize the network traffic data anomaly identification and detection algorithm from the perspective of overall network operation and maintenance, this paper conducts business analysis and analysis on network traffic identification, operation, and maintenance analysis.

Large enterprise networks generally adopt hierarchical design [7]. Structurally, usually using star structure or tree structure, the overall network architecture is shown in Figure 2 and 3. Star structure is a two-layer structure. The network core is generally deployed in the group headquarters, and each secondary and tertiary system unit is generally distributed in each province or municipality

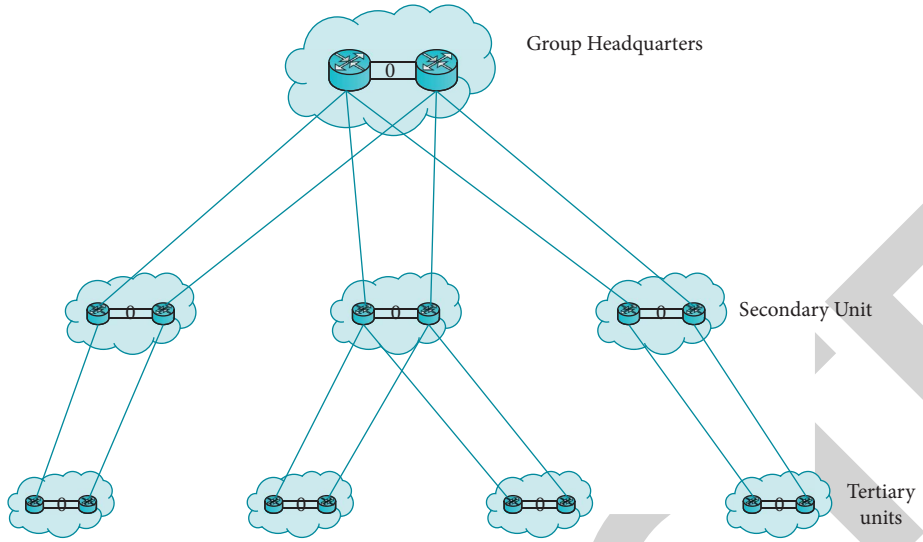

Figure 3: Tree-type topology.

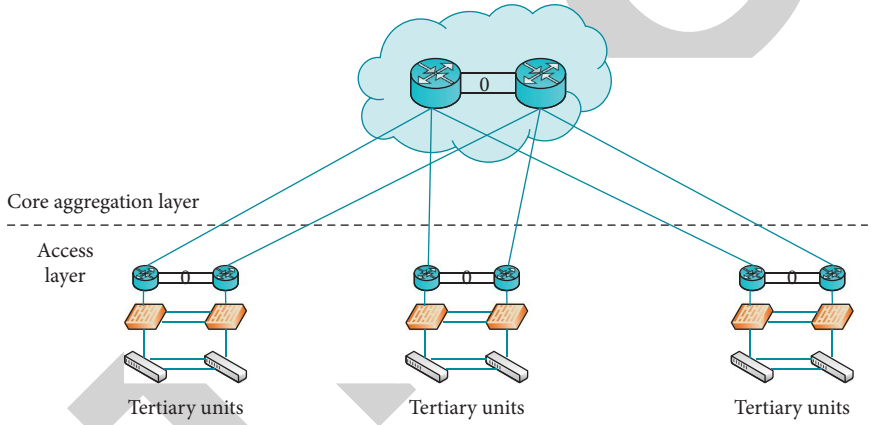

Figure 4: System unit network architecture.

directly under the Central Government. The network of each secondary and tertiary system unit is converged to the network core through the WAN network link, realizing network access and data interaction with the group headquarters. Tree structure is a kind of three layers or multilayer structure, the network core is also deployed in the group headquarters, in each secondary unit or regional company local set network convergence point, and each tertiary system unit is no longer directly connected to network core, but respectively access to the corresponding secondary unit or regional company, through the secondary unit or regional company network link to the network core, so as to realize network access and data interaction. The network within the group headquarters and each secondary and tertiary units is generally designed with the three-layer structure of the core layer, convergence layer, and access layer.

The WAN core convergence equipment is deployed at the group headquarters [7, 8]. Two modular high-end routers, with the technical function of separating routing control level and data forwarding level, are used to carry various types of business data between the group headquarters and system units and between system units and system units. As a WAN convergence node, the group headquarters rent the MSTP link of third-party operators to realize the link interconnection between the group headquarters and each unit of the system, and the link bandwidth varies from 4M to 100M.

Secondary and tertiary units deploy two mutually backup WAN routers locally [9]. Two different WAN links are connected to the group headquarters to achieve data interaction with the group headquarters and other units. The WAN link bandwidth of secondary or tertiary units connected to the group company is determined according to different service requirements, ranging from 4M to 100M. It can be seen that WAN link is the bottleneck of enterprise network bandwidth and the focus of network traffic analysis.

The network of each unit of the system includes the WAN access part connected to the group company and the LAN part of the system unit [7]. Topological maps are detailed in Figure 4. The WAN access part of the system unit side connected to the group company is mainly composed of the WAN access equipment, including two routers and two

firewalls: two routers, respectively, by renting the WAN link connection with the group headquarters, two links by route mutual link and link load balance, and two WAN routers through the network cable, configuration for equipment level standby, namely, a router failure when another router can complete to take over all the business. The two firewalls are connected using heartbeat lines, configured in AA mode, connected to two WAN routers, and then connected to the system unit LAN core switch, respectively.

The system unit LAN generally adopts a two-layer network architecture [8]. The core convergence layer is usually deployed with two core switches, and the two switches are virtual to one device through the network virtualization technology, as the network core convergence equipment of the system unit, used to gather the LAN access layer switches and realize the high-speed interaction of the system unit network data. The access layer is mainly divided into two categories; one is floor access, composed of a large number of floor access switches, mainly used for wired network access of ordinary office users; the other is server access, mainly used for accessing various servers of system units. The LAN firewall is generally hung on the core switch, which is used to divide the security area of the LAN and realize the access control between regions.

3.3. Analysis and Research on Network Operation and Maintenance Problems

3.3.1. Network Operation and Maintenance Problem Analysis. Analyzing from the perspective of network technology and network management [9], there are the following reasons:

(1) The overall investment in informatization is insufficient.

As an energy enterprise with electric power production as its main industry, information construction started early, in the 1960s and 1970s; it began to carry out a large range of applications in production, dispatching, and other fields. Since the beginning of this century, the management informatization has been continuously promoted and improved. Overall, the information attention in the field of production control is high, while the information construction of the management information region is not enough, and the overall investment is insufficient.

(2) Network construction attaches great importance to construction and ignores operation and maintenance management.

Network system is the foundation of the whole information construction, and each power group attaches great importance to the construction of network system and has built a high-speed and reliable backbone network and data center. Network operation and maintenance management system is a system used for network system monitoring and management, and the service object is the network operation and maintenance management personnel. Different from the construction of the network system and the application system, the core of the network management system is a complete set of network management software, often invisible, untouchable, and not providing value for users, so the attention is far from insufficient, and the capital investment is extremely low.

(3) Lack of operation and maintenance-oriented network performance management platform, low efficiency and high cost.

At present, many groups have built a set of network management system to realize the functions of topology discovery and display, link fault alarm, configuration backup, asset management, and so on. After years of continuous construction, a Power Group has built a network management system, which realizes topology discovery and display, link fault alarm, configuration backup, asset management, and other functions. However, the current network management system, focusing on technology in equipment and link level monitoring, has not yet included traffic monitoring into the monitoring category. With the continuous deepening of information construction, more and more application systems operate on the network, and the existing network monitoring system has been unable to meet the current operation and maintenance requirements.

3.3.2. Solution to Network Operation and Maintenance Problems. Through the above analysis, from the perspective of network operation and maintenance management and technology, to effectively improve the network operation and maintenance management level and improve user satisfaction, the following points should be done [13]:

First, the electric power enterprises should increase the overall investment in information technology, especially to improve the investment in the information system of the management information region and to solve the problem of the overall lack of investment.

The second is to change for a long time, in the network construction, "heavy construction, light operation and maintenance" inherent mode. They should be "construction" and "operation and maintenance," both grasped, and both hands should be hard. Increase the capital investment in the network operation and maintenance management.

The three are combined with the actual needs of a line network operations team, and users reflect the centralized problem, develop a set of operational network performance monitoring platform, through visualization, show the composition of the network flow and flow speed, simplify operation steps, reduce the difficulty of use, and improve network operations team work efficiency and fault response speed.

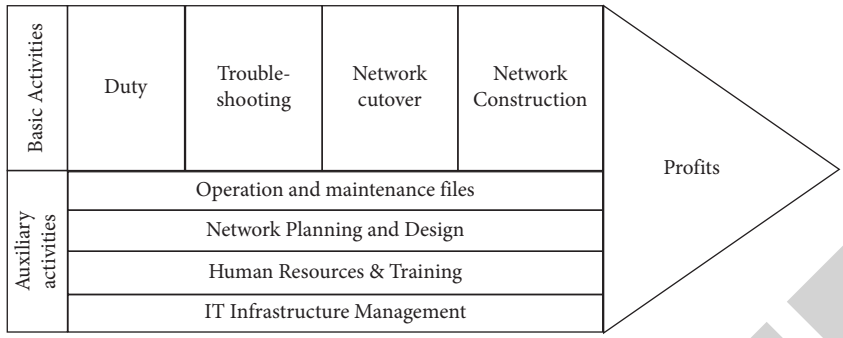

Figure 5: Network operation and maintenance value chain model.

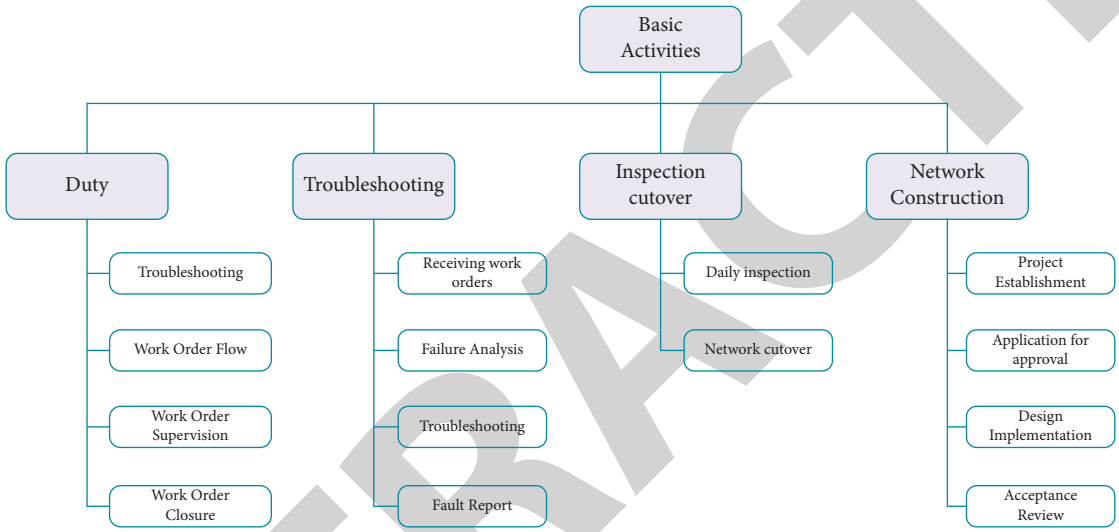

Figure 6: Enterprise Workflow tree.

3.4. Network Performance Fault Process Diagnosis and Analysis

3.4.1. Value Chain Analysis. All the different but interrelated business activities [13] constitute a dynamic process of creating value, namely, the value chain [14]. The value chain theory divides all related activities within an enterprise into two categories: according to whether the activity adds value, and whether the activity is the main activity of two methods for determination. The value chain model of the network operation and maintenance team is shown in Figure 5.

By sorting out and classifying the production activities of the network operation and maintenance team, a preliminary process is formed. In these works, direct operation and maintenance duty, troubleshooting, network coverage, and network construction are the basic activities that can create value. The above basic activities are further refined to form the enterprise workflow tree as shown in Figure 6.

3.4.2. Analysis of the Causes of the Existing Problems. Through the analysis of the fish bone map, we can see that there are four main factors that cause the slow investigation of network performance problems, namely, less investment, network operation and maintenance personnel, operation and maintenance management, and network operation and maintenance work, as shown in Figure 7. The lack of information investment thus affects the procurement of relevant network operation and maintenance tools and directly affects the network traffic identification ability of enterprise network operation and maintenance.

4. Experiments and Results

According to the research in the previous chapters, in the network traffic identification based on decision tree, this paper proposes the application of decision tree-based fast P2P traffic identification technology in network traffic data anomaly identification and detection. The basic idea is to make full use of the rapid network power identification capability of decision trees to avoid the defects of custom traffic identification. First, use the centralized management application system access control information to obtain the data information of the network traffic, and then calculate the cross entropy of the network traffic. Finally, use the split attribute to analyze the abnormal situation of network traffic; the overall framework process is shown in Figure 8.

4.1. Cross-Entropy of the Network Traffic Anomaly. Entropy is a concept introduced from the field of chemistry that gradually extends to the field of information theory and

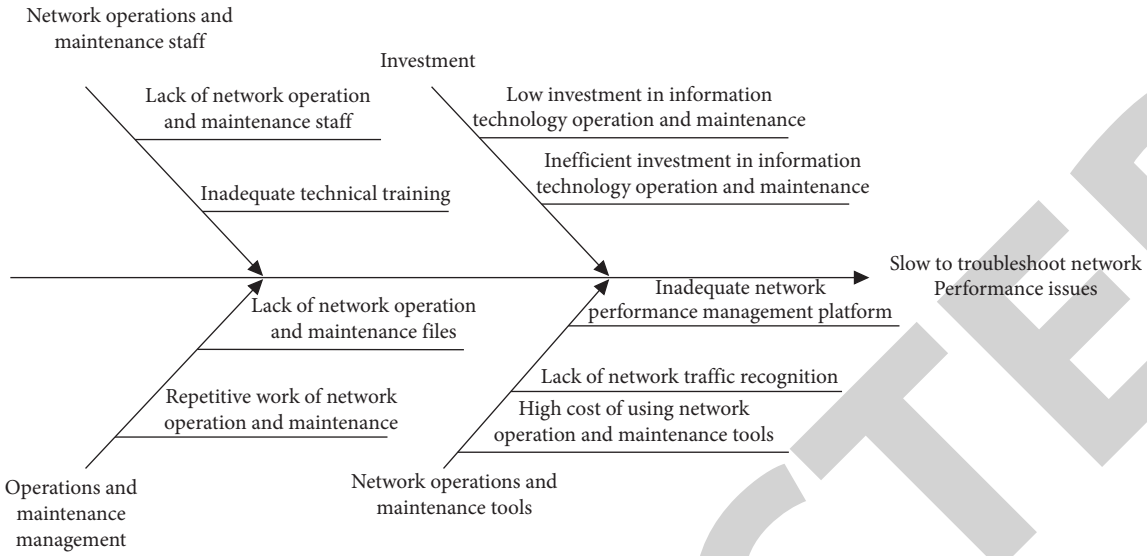

Figure 7: Flow chart of fish bone chart analysis problems.

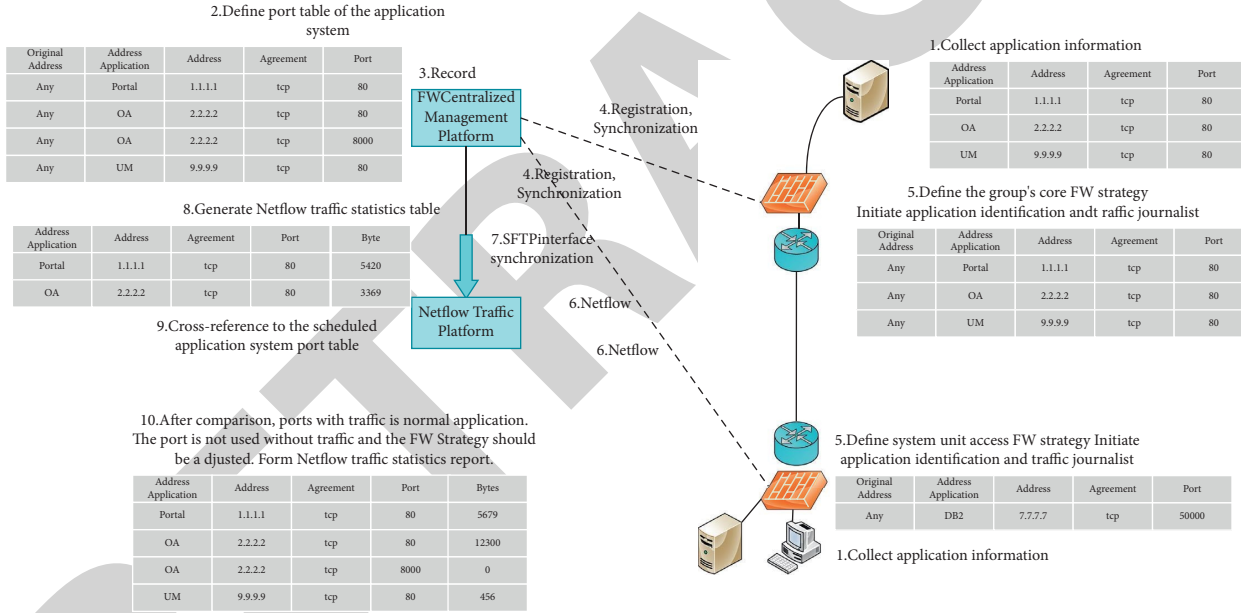

Figure 8: Framework process of fast P2P traffic identification based on decision tree.

becomes a highly dynamic method. Entropy value applies to indicate the disorder degree of a system, and the larger the entropy value, the more disordered the data in the system; the smaller the data is, the more "pure" it is. Cross-entropy is an important concept in Shannon's information theory, which is mainly used to measure the difference information between two probability distributions [14, 15].

Cross-entropy is used to measure the different degree in the statistical significance of the 2 distributions. This paper measures the statistical difference of traffic properties between two time periods by cross-entropy. The distribution functions of the source IP in the preceding and following time periods are P and Q respectively, so the cross-entropy H (P, Q) can be expressed by the following equation:

$$H(P,Q) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(i,j)\mathrm{lb}\frac{P(i,j)}{Q(i,j)} + \sum_{i=1}^{n}\sum_{j=1}^{m} Q(i,j)\mathrm{lb}\frac{Q(i,j)}{P(i,j)}. \quad (2)$$

When the network traffic is normal, the amount of information between inside and outside the network has a certain continuity within a certain period of time. In attacks such as network scanning or worm, adjacent traffic in the network segment will suddenly appear quite different. Network traffic contains complex background traffic, and it is difficult to find differences through only the number of mutations of the packets, as shown in Figure 9 and 10. From Figures 10 and 11, the cross-entropy has a strong description

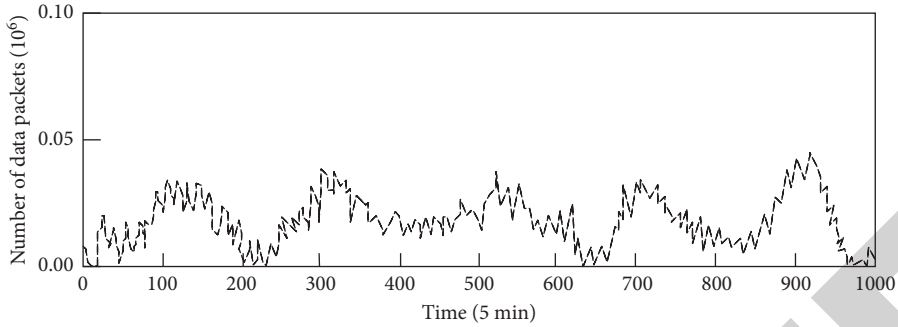

Figure 9: Cross-entropy of flow anomalies.

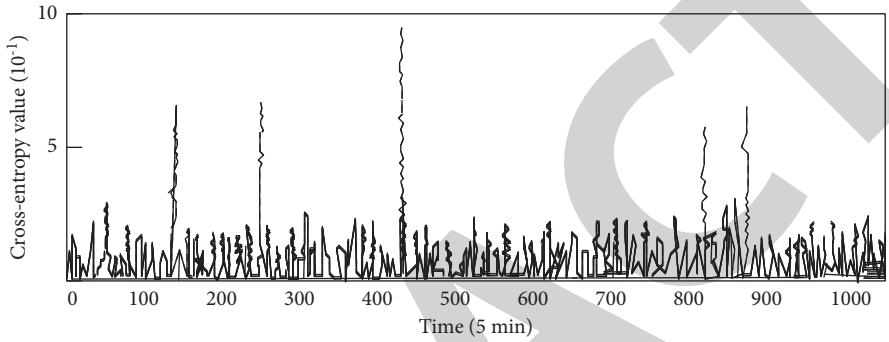

Figure 10: Source I P address cross entropy.

performance for anomalies, which are described in this paper.

4.2. Split Attribute Study. As mentioned above, different decision tree classification algorithms use different judgment conditions to choose the split attributes, and the two main judgment conditions are information gain and information gain rate.

4.2.1. Split Attribute Selection Is Performed Based on the Information Gain. Suppose that the set of training samples is S, and the attribute set is $P = \{p_1, \ldots, p_i, \ldots, p_n\}$. The type marker collection is $T = \{t_1, \ldots, t_i, \ldots, t_n\}$. The total sample size in S is $|S|$.use attribute p_i. The set of samples for the j class is S_{ij}, gather S_{ij}. The number of samples is $|S_{ij}|$. The proportion of samples belonging to class j in the sample data set is (2):

$$P(C_j) = \frac{|S_{ij}|}{|S|}. \tag{3}$$

Now, the information entropy value of the sample data set S is Entropy (S, p_i). For (3),

$$\text{Enproty}(S, p_i) = \sum_{j=1}^{n} -P(C_j)\log_2 P(C_j) = \sum_{j=1}^{n} -\frac{|S_{ij}|}{|S|}\log_2\frac{|S_{ij}|}{|S|}. \tag{4}$$

Assuming that, in the sample data set, there is the property p_i. The corresponding value domain is the one of v_i, and $S_i(v)$ represents the property p_i. Take a subset of samples with a value of v. Then, the sample set S pairs of the property p_i. The information gain is

$$\text{Gain}(S, p_i) = \text{Enproty}(S, p_i) - \sum_{v \in v_i} \frac{|S_i(v)|}{|S|}\text{Enproty}(S, p_i). \tag{5}$$

Calculate the set of $P = \{p_1, \ldots, p_i, \ldots, p_m\}$. The information gain value for each attribute Gain (S, p_i), $i = 1, \ldots, m$, and if the maximum information gain value is Gain (S, p_k), then select the property p_k as a split property of the current sample collection.

4.2.2. Split Property Selection Is Performed Based on the Information Gain Rate. After calculating the information gain of the sample set S, S is calculated at the property p_i. The split information on this section is

$$\text{SplitGain}(S, pi) = -\sum_{v \in v_i} \frac{|S_i(v)|}{|S|}\log_2\frac{|S_i(v)|}{|S|}. \tag{6}$$

Then, the sample data set S is relative to the property p_i. The information gain rate of GainRatio (S, p_j) for

$$\text{GainRatio}(S, pj) = \frac{\text{Gain}(S, p_j)}{\text{SplitInfo}(S, p_j)} = \frac{\text{Enproty}(S) - \sum_{v \in v_j}|S_j(v)|/|S|\text{Enproty}(S)}{-\sum_{v \in v_j}|S_j(v)|/|S|\log_2|S_j(v)|/|S|}. \tag{7}$$

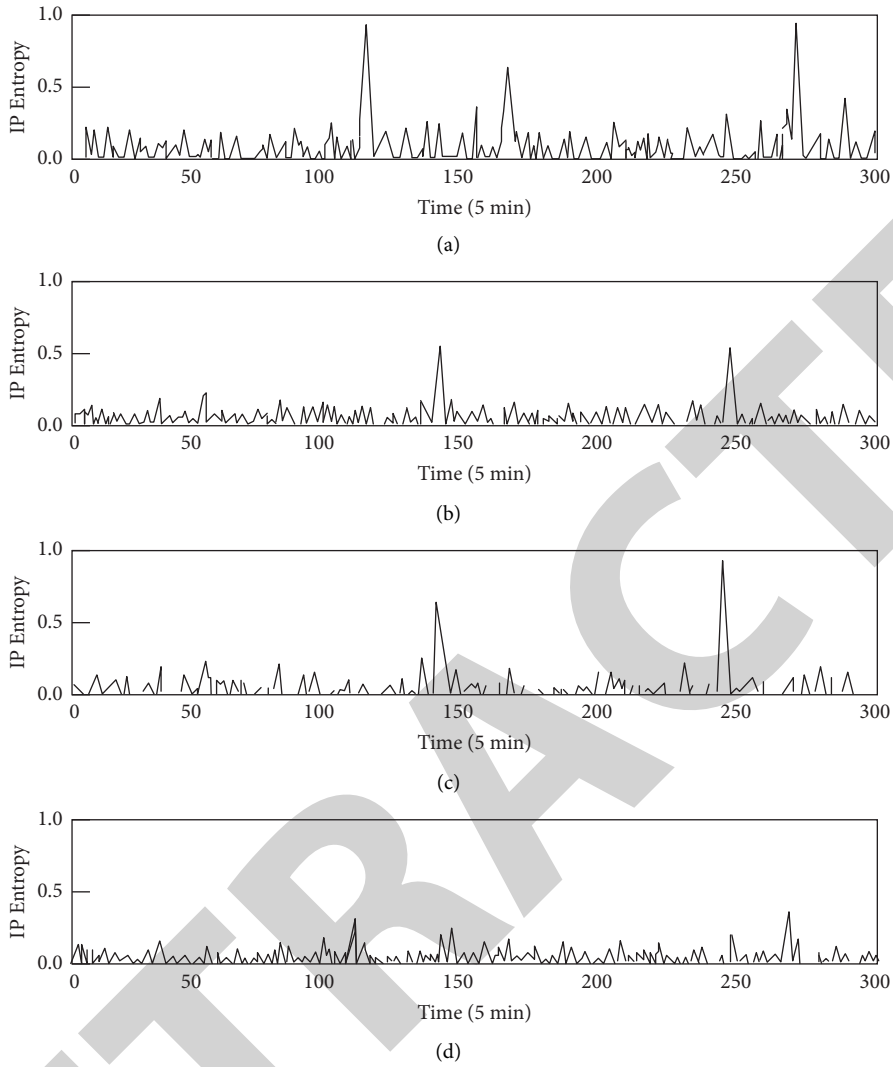

Figure 11: I P entropy change curves for different types of aggression. (a) Characteristics of traffic containing DDOS attack. (b) Traffic characteristics of worm propagation. (c) Traffic characteristics of network scanning. (d) Traffic characteristics of port scanning.

The information gain rate of all unselected attributes is calculated by the above formula to obtain the information gain rate set GainRatioSet (S) of all attributes to be selected:

$$\text{GainRatioSet}(S) = \{\text{GainRatio}(S, p_1), \ldots, \text{GainRatio}(S, p_j), \ldots, \text{GainRatio}(S, p_m)\}. \quad (8)$$

Select the attribute with the largest information gain rate in the set as the split attribute of the current node.

4.3. Analysis of the Experimental Results

4.3.1. Data Extraction and Preprocessing. Data extraction is about selecting the appropriate source database and extracting phase data from the database. Many attributes in the source data may be unrelated to the classification task that can slow down or mislead the learning steps. This paper takes a local area network inside the campus network as the experimental environment to obtain source data from the I: 1 switch. The data structure for anomaly detection is described above. The decision attributes include source IP address entropy, destination IP address entropy, source 1 : 1 entropy, destination port entropy, protocol, and IP address number.

4.3.2. Abnormal Detection Process and Results Analysis. The experiment is divided into two parts: detection and classification effect test. The test dataset is collected by injecting known attacks into the LAN. The detection effect test dataset consists of four subdatasets containing 288 data tuples, each containing 10 attacks of the same type. 5 min is a time unit, and 288 time intervals per day are randomly selected to inject similar attacks. Each time interval corresponds to a statistical data tuple. Four detection effect test data sets with different attack behaviors can be obtained by applying the data processing method in the previous section. The test data set acquisition method of classification effect

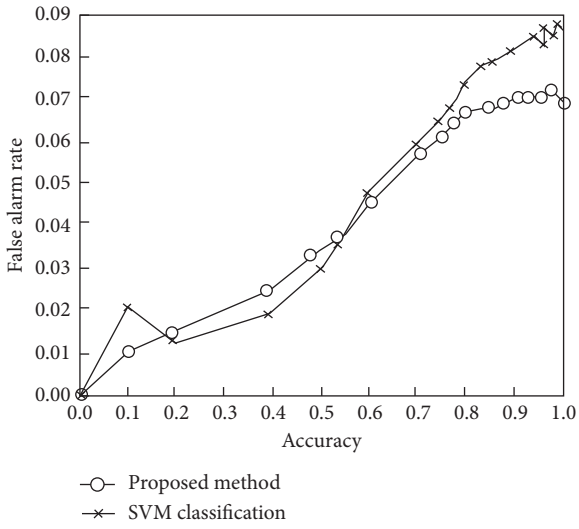

Figure 12: Abnormal detection of the ROC curves.

validation is the same as above, 288 time periods a day are selected, and 10 different attacks are injected into 40 time intervals. Only one type of attack is injected into each test period, without considering the traffic influence of different types of attacks. After the injection attack, the attribute will change significantly. Figure 11 describes the cross entropy change in 288 time periods before and after the source IP address as the injection attack.

Classification effect test: applying previously established model rules to the dataset containing 40 sets of aggressive behaviors yields the results, the number of classifications represents the number that the method divides the 288 data tuples into the corresponding categories, and the correct number indicates the number of correct classifications. It can be seen that the classification accuracy of this method reaches 92.5% and can classify abnormalities efficiently.

Test effect test: the models established in 3 sections are applied to the above 4 sub-data sets, respectively, and the detection results are shown in the following table. Specifically, the statistics quantification results for applying the decision tree method for anomaly detection and identification are listed.

The number of false positives is included in the detection number, the ratio of the correct number of detection to the total number of anomalies, and the false positive rate is the ratio of the number of false positives to the total number of data (the total number here is 288).We know that this method has strong detection ability for DDOS attacks and worm attacks in the network and achieves high detection rate at low false positive rate. The detection false alarm rate of attack behaviors such as network scanning and port scanning is slightly higher, because the same characteristic network traffic changes may be caused by normal network operations due to other types of centralized network operations (see Figure 12).

It compares the decision tree method with the SVM taxonomy. It can be seen that the false alarm rate can be stable in a low determination range when the detection rate reaches a certain rate, but the same false alarm rate has a high detection rate, and the final anomaly detection false alarm rate is stable in a lower range.

5. Conclusion

From the perspective of enterprise network operators, this paper introduces the current operation and maintenance process of enterprise network performance failures, plans the optimization and reengineering of the operation and maintenance process through the value chain analysis method, finds out the shortcomings of the current process links, and proposes solutions. Then, in view of the shortcomings of traditional anomaly detection methods based on traffic overall characteristics that some well-planned hidden attacks cannot be detected, an anomaly detection method based on traffic feature attribute structure is proposed. Taking the cross entropy of the distribution of traffic attribute values per unit time as the research object, it is proposed to apply the decision tree method to the detection and classification of traffic anomalies, and to use Weka as the experimental tool and a local area network of the campus network as the data source. Experiments show that this method has a high detection rate, especially for DDoS and terminal *i*:1 scanning. It provides a new method for the rapid detection and classification of traffic anomalies and improves the accuracy of detection and classification. Because every step of the decision tree is accurate classification without fuzzy concept, a slight deviation in the selection of model nodes will directly affect the abnormal classification results. In the future, the concept of fuzziness will be introduced in the selection process of classification nodes, and subsequent nodes will be introduced in the current node segmentation process, to further improve the classification detection performance.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by Hengshui Open University.

References

[1] H. Zhang, B. Fang, and M. Hu, "Review of the Internet measurements and analysis," *Journal of Software*, vol. 14, no. 01, pp. 110–116, 2003.

[2] J. Bi, *The Internet Behavioral Measurement and Analysis Study*, Graduate School of Chinese Academy of Sciences (Institute of Computational Technology), Beijing, China, 2002.

[3] Z. Guo, *Implementation and Application of Man Man Network Deep Package Detection System*, Dalian University of Technology, Dalian, China, 2016.
[4] H. Tan, S. Yang, and H. Kan, "Research on P2P traffic in campus network," *Journal of Changsha University*, vol. 29, no. 02, pp. 70–72, 2015.
[5] D. Kang, *Research on Feature Code Extraction and Traffic Control for Internet P2P Application*, Yunnan University, Kunming, China, 2013.
[6] Li Xin, *Research on the Internet Traffic Control Countermeasures of Operators*, Beijing University of Posts and Telecommunications, Beijing, China, 2011.
[7] L. Zheng and X. Yang, "Analysis and Implementation of Instant Messaging Software Protocol Based on DPI," *Information network security*, no. 01, pp. 51–58, 2016.
[8] Daili.Research, *Implementation of the DFI Flow Classification Technique*, Beijing University of Posts and Telecommunications, Beijing, China, 2011.
[9] Y. Su, *Research on Network Encryption Traffic Classification Based on Data Mining*, Harbin University of Science and Technology, Harbin, China, 2019.
[10] J. Liao, H. Tan, and Y. Liu, "Deep service perception and telecom network P2P service," *Journal of the University of Electronic Science and Technology*, no. 06, pp. 1338–1341, 2007.
[11] J. Hu, "Preliminary research on the application of network traffic management and control technology in campus network," *The Network and the Information*, vol. 23, no. 05, pp. 18-19, 2009.
[12] Li Ning, *Research Based on NetFlow Technology in Enterprise Network Application*, Nanjing University of Science and Technology, Nanjing, China, 2009.
[13] K. C. Claffy, H. W. Braun, and G. C. Polyzos, "A parameterizable methodology for Internet traffic flow profiling," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1481–1494, 1995.
[14] W. Ding, *Network Anomaly Flow Detection and Filtering Based on Decision Tree Classification*, University of Electronics Technology, Sichuan, China, 2013.
[15] Q. Zhang, "Financial data anomaly detection method based on decision tree and random forest algorithm," *Journal of Mathematics*, vol. 2022, Article ID 9135117, 10 pages, 2022.