Intelligent network operation and maintenance system based on big data

Bin Xiao\textsuperscript{1,*}, Weifeng Wang\textsuperscript{1}

\textsuperscript{1}University of Jinan, 250022, Shandong, Jinan, People R China.

*Corresponding author e-mail: 187345728@jlu.edu.cn

Abstract. With the deepening of network system construction, it is more and more important to monitor the operation status and performance of business system, network system, network equipment, host equipment, database and middleware; at the same time, the requirements for the depth, breadth and granularity of network information security monitoring are constantly improved, and the traditional network operation and maintenance is more and more inadequate. The "intelligent network operation and maintenance system based on big data" can realize the operation monitoring of network, host, business application, network equipment, desktop terminal, etc., and build an electronic platform based on big data to realize the standard operation and maintenance process, achieve the goal of real-time monitoring and closed-loop control.

Keywords: Big data, Network operation and maintenance, Network information security

1. Introduction

In the 21st century, computer technology has entered the era of intelligent big data, and more and more data can be analyzed and processed relying on intelligent computers. Since the beginning of the new century, big data has appeared frequently in various fields, and the construction of cloud computing platform to process the collected data is the top priority of big data processing. In terms of network equipment monitoring, the information detected every day is stored in the monitoring server. If the information is handled properly, it will be of great help to the monitoring, analysis and prediction of network equipment.

"Intelligent network operation and maintenance system based on big data" provides a solution for network equipment monitoring, analysis and prediction in the era of big data and high data visualization. In the process of implementation, the front-end through grafana two-dimensional data visualization processing and presentation; equipment monitoring uses ZABBIX, which is a distributed network monitoring and management solution based on web interface. Based on the server client architecture, it can be used to monitor all kinds of network services, servers and network equipment, and almost all network devices that need to be monitored, and the monitoring information set to the operation and maintenance server, real-time rendering to the front end. Through ZABBIX, the basic analysis of data can also be carried out. By setting the data trigger, the alarm can be triggered when
certain conditions are met. At the same time, the monitoring data can be processed twice by script, and all kinds of report documents can be generated regularly. The operation of general large and medium-sized network system can be realized by spending less manpower to complete a small amount of work Maintenance and monitoring. At the same time, ZABBIX's programmability, good open source API and custom script support make ZABBIX have unlimited development potential.

This article will build a mental health education system for college counselors based on web technology. The system consists of two subsystems, namely an online expert mental health knowledge system and an online psychological consultation system. College counselors use the online expert-type mental health knowledge system to discover possible mental health problems and to comprehensively and systematically learn mental health knowledge. In addition, through online psychological counseling system to conduct online counseling on all of your own mental health problems is suitable for your own method, so as to improve your psychological quality.

2. Design ideas and data analysis

2.1 Design ideas

The original intention of the design is to make the network management more convenient and intelligent, and reduce the workload of network operation and maintenance personnel. Because in many cases of campus network, the network traffic is too large and the instantaneous flow is too large. This kind of problem is extremely prominent, especially in the campus network, the network density is very large, the network type is extremely complex, such as video business flow, audio business flow, instant message flow, game flow, office traffic, campus network equipment management flow, etc., these traffic data volume is large and important. Therefore, when the network throughput is relatively large, especially when the network is congested, in order to take into account the more critical traffic, we need to analyze and classify the traffic, and set the priority for it.

This idea is similar to SDN, and the ideas are interlinked. We can classify network traffic based on port, network protocol analysis, deep packet detection and statistical learning. In short, packets generally have the same quintuple, namely source IP address, destination IP address, source port number, destination port number and protocol type. As long as one of the above five elements of a packet is different, they do not belong to the same stream. Therefore, we can classify the TCP flow or UDP flow in the network according to its application type in the application layer. In fact, take the example just mentioned. If we only look at the campus network, we can distinguish the traffic types according to the source IP, destination IP, source MAC, destination MAC and other information. In the classroom, office building and other different scenarios, different traffic ratios are analyzed. According to the different traffic types, different priority is set for different traffic and classified forwarding. For example, in the classroom, you can set the priority of the traffic type related to learning to the highest level, and the priority of the traffic flow of the game to the lowest. By analyzing the traffic, we can analyze the malicious attack traffic, and then limit the traffic and report to the administrator. The application scenario cannot be limited to the technical level, but also can analyze the proportion of various traffic in different places (such as office buildings, teaching buildings, dormitories, etc.) to investigate students' Internet access.

By setting up a network monitoring server, the whole network cluster is monitored, and the information is integrated and visualized. However, the utilization rate of data is not very high at this time. Through appropriate alarm settings, the monitoring situation is analyzed by big data, It maximizes the value of the data, gives the operation and maintenance oriented auxiliary information, and combines with the network security related functions to improve the security of the network cluster. There are real-time front-end data visualization feedback, and various kinds of operation and maintenance reports after data processing, which give the operation and maintenance personnel maximum assistance and improve the operation and maintenance efficiency and cluster security level.

2.2 Data analysis
Through the FCM algorithm and the algorithm optimization and supplement combined with the actual project, the data clustering and screening can be well carried out, and the corresponding analysis can be carried out according to the respective rules. In many fuzzy clustering algorithms, fuzzy c-means clustering algorithm is the most widely used and more successful. Fuzzy c-means clustering algorithm integrates the essence of fuzzy theory. Compared with K-means hard clustering, fuzzy c-means clustering algorithm provides more flexible clustering results. Because in most cases, the objects in the data cannot be divided into obviously separated clusters. It is hard to assign an object to a specific cluster, and errors may occur. Therefore, a weight is assigned to each object and each cluster to indicate the degree to which the object belongs to the cluster. By optimizing the objective function, the membership degree of each sample point to all class centers is obtained, so as to determine the category of sample points, so as to achieve the purpose of automatic classification of sample data. The problem to be solved by FCM is to minimize the weighted square error function.

\[
J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^m \| x_j - C_j \|^2 \quad (1 \leq m < \infty)
\]

Where \( m \geq 1 \), \( U_{ij} \) is the membership degree of \( X_i \) in cluster \( J \), \( X_i \) is the \( i \)th of \( d \)-dimensional measurement data, \( C_j \) is the \( d \)-dimensional center of the cluster, and \( |*| \) represents the similarity between the data and the centroid

\[
v_p = \frac{1}{\sum_{x \in C} (\| x - C \|_m^{\text{m-1}})} \quad C_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m}
\]

Fuzzy C-means (fuzzy c-means clustering) algorithm steps:

1. Initialize \( U = [u_{ij}] \) matrix \( U^{(0)} \)
2. step: calculate the centers vectors \( C^{(k)} = [C_j] \) with \( U^{(k)} \)

\[
C_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m}
\]
3. Update \( U^{(k)}, U^{(k+1)} \)

\[
v_p = \frac{1}{\sum_{x \in C} (\| x - C \|_m^{\text{m-1}})}
\]

4. If \( \| U^{(k+1)} - U^{(k)} \| < \varepsilon \) then STOP; otherwise return to step 2

FCM algorithm can help to complete fuzzy clustering, but there are some limitations in the scenario of this project, and the clustering results are not very good when the samples are not ideal. For example, if a certain value (outlier) is far away from each cluster center, the membership degree of each class is very small. However, due to the constraints of FCM algorithm, it will affect the final result of iteration (classification), or if the data differentiation is not high, it can't cluster well. Therefore, in order to optimize the above situation, supplement the algorithm The secondary verification mechanism is used to match the membership matrix obtained by the algorithm classification with the data. The classified data and the time stamp corresponding to the data are stored in the dictionary (key: timestamp, value: data). Each type of data has a dictionary, which is embedded into the array in the order from small to large. Then, through the secondary classification and integration, the classes with low differentiation are merged and the outliers are eliminated The impact of.

2.3 Equipment operation and maintenance report

Using Python docx library, it can automatically generate docx format documents such as traffic analysis, blacklist and access detection records, so as to save the work of manual writing report documents for operation and maintenance personnel.
Docx is a common document format used after Microsoft Office 2007. It replaces its current proprietary default file format with a new XML based compressed file format. The letter "X" is added after the traditional file name extension. Compared with doc format documents, docx is a better choice because it is based on the office open XML standard, all word processor software supports all advanced functions, and the file is smaller and easier to read and transfer.

The python docx we use is a python library used to create and update Microsoft Word. Docx files. Documents saved with it can be found in Microsoft office, Google Docs, and, OpenOffice.org 3 and so on.

In Python script, we optimize the API use of Python docx library to a certain extent, so as to facilitate the actual users to modify, add and delete the report content, and to a certain extent, integrate different functions to achieve multiple functions of a script.

3. Realization of equipment operation weekly report
After initial processing, the traffic data is stored in the dictionary in the form of key value pairs (Key: timestamp, value: traffic data), and then the traffic data is clustered by FCM algorithm, and the accuracy of clustering is improved by secondary verification (see the figure below). Then, the characteristic value is extracted to judge the flow up and down in each time period, and the analysis is made by matching the text database. The general idea of hardware analysis is similar.

The weekly report of equipment operation is oriented to users and provides the operation status of assets in the past period of time, while the safety report is for operation and maintenance personnel. It provides feedback on the equipment operation in the past week and provides operation and maintenance strategy suggestions. Therefore, when obtaining the original data, it distinguishes the requested data content according to different requirements. In the process of data analysis, the principle of clustering algorithm for data operation is the same.

4. Innovation and characteristics
1) The project uses ZABBIX as the basis of the monitoring server, and the front-end data visualization framework which is easy to use and can be highly customized, instead of the simple web interface provided by ZABBIX, can better achieve the experimental goal and present the data in real time. At the same time, make full use of the characteristics of ZABBIX platform to push the alarm centrally.

2) By pinning alarm instead of traditional alarm, all operation and maintenance personnel can get the alarm information at the first time and make response at the same time.

3) The project not only simply presents the monitoring data, but also analyzes and forecasts the big data at a higher level. It can regularly generate various reports, such as equipment operation weekly report, server security report, etc. the report strictly follows the production standard format, and gives analysis, prediction and suggestions on operation and maintenance.

4) Since the operation and maintenance of the client can hardly affect the operation and maintenance process of the project.

5) In the later stage, distributed deployment is attempted, which greatly improves the long-term scalability and reliability of the project, and reduces the pressure on a single server.

6) In the future, we will provide support for IPv6 devices and monitor IPv6 devices.

7) Cloud Computing Oriented infrastructure operation and maintenance will be the focus of the next stage of the market. After the project is put into operation, it will have good market benefits, which is consistent with the future market demands, and has great potential for commercialization.

References
[1] Yang Jiwu. Research on Computer Information Security Precautions in the Age of Big Data[J]. Communication World, 2019,26(03)
[2] Fu Jiaqi. Research on Intrusion Detection Technology Based on improved fuzzy c-means clustering algorithm [J]. International Journal of Mental Health & Addiction, 2017, 16(3):136-149.
[3] Li wen, Hao xiaopei, Fan Chunmei. The 12th China Intelligent Transportation annual meeting [J]. 2014, 14(4):873-879.

[4] Wang Fei. Building a Security Protection System to Consolidate User Data Security [J]. China New Communications, 2019, 21(04)

[5] Yuan Li, Xu Guiming, Wang Mengxiao. Engineering Design of Security Control for Display and Control Terminal[J]. Network Security Technology and Application, 2018(11)

[6] Zhang Jiansheng, Zhang Xiaohong, Peng Linhua. Computer Network Information Security and Protection Countermeasures under the Background of Big Data[J]. Information and Computer (Theoretical Edition), 2018(20)

[7] Chen Tao. Research on Data Security Protection Scheme of Information System in New Period[J]. Shanghai Construction Science & Technology, 2018(05)

[8] Chen Zhen. Research on the Construction of Big Data Information Security Risk Framework[J]. Information and Computer (Theoretical Edition), 2018(19)