Research on Optimization of Communication Network Transmission Based on Big Data

Chao Liang(✉), Pengfei Ma, and Ge Yang

Systems Engineering Research Institute, Beijing, China
greenday_xh@hotmail.com

Abstract. Research the information dissemination on the communication network in combination with big data technology. A large number of comprehensive and timely data can ensure the accuracy of decision-making or control. A large number of data is conducive to drawing correct conclusions from a statistical point of view. Comprehensive data can avoid one-sided understanding of things, and timely data can reduce delays. By analyzing the information transmission on the communication network, build and optimize the communication network structure, and improve the information transmission on the communication network.

Keywords: communication network · big data · retransmission · delay

1 Introduction

The rapid development of modern information technology, especially big data technology, allows us to collect all kinds of recorded data and monitoring data in near real time, thus greatly expanding the scope of data and research available in research [1]. The processing process of big data is to collect data first. A big word of big data reflects the huge amount of data and the complexity of data types. In view of the characteristics of big data, data collection is the first step of big data research and the most basic part. The second step is to process the collected data, eliminate the unnecessary data, and leave the required data for unified storage. The third step is to analyze the collected and stored data. The analysis of data is the core work of big data research. In this process, the value and laws of data can be found. The last step is to explain and explain the data analysis results [2].

The rapid acquisition and in-depth analysis of a large amount of data can be completed through big data technology, which can realize the value of data through mastering huge data information and specialized processing of these data. Due to the problems of perceptual operation such as feeling and chasing hot spots, the application of big data technology can turn perception into rationality, speak with data, and transform decision-making to data-driven. First, quickly extract the relevant massive data, and then deeply analyze the data through big data technology, analyze the correlation between multiple factors and give the weight of each factor. Sampling survey is the basic research method

N. Akhtar et al. (Eds.): PMIS 2023, AHIS 8, pp. 1118–1124, 2023.
https://doi.org/10.2991/978-94-6463-200-2_118

of social science. In the information age, we can use big data technology to achieve comprehensive collection of large amounts of data, and can achieve rapid and complex processing of large amounts of data [3].

In the era of big data, the methods and means of scientific research will undergo major changes. Through the acquisition, processing, analysis and optimization of massive and diverse data from different sources, the decision-making behavior is based on data. Conduct accurate, multidimensional and multifaceted consideration of multiple factors. Although big data originated from information and communication technology, its impact on society, economy and life is not limited to the technical level [4]. In essence, big data provides us with a new way to look at the world, that is, decision-making behavior will increasingly be based on data, rather than relying more on experience and intuition as in the past.

The processing and analysis of big data is becoming the focus of the new generation of information technology fusion application. The application of new generation of information technologies such as social networks, mobile Internet, Internet of Things, digital home, e-commerce and so on continues to generate big data [5]. Through the management, processing, analysis and optimization of data from different sources, feedback the results to the application will create huge economic and social value. Big data has the energy to promote social progress [6].

Big data is a new engine for the sustained and rapid growth of the information industry. New technologies, new products, new services and new formats for the big data market are emerging. The application of big data will become the key factor to improve the decision-making level and core competitiveness. Decision-making in all walks of life is changing to data-driven [7]. In the business field, the analysis of big data can enable retailers to grasp the market dynamics in real time and respond quickly, provide decision support for businesses to develop more accurate and effective marketing strategies, and help enterprises provide more timely and personalized services to consumers. In the medical field, it can improve the diagnostic accuracy and drug effectiveness. In the field of public utilities, big data has also begun to play an important role in promoting economic development and maintaining social stability.

2 Communication Network Information Dissemination Combined with Big Data Technology

2.1 Resend Mechanism

Error control is the use of coding methods in digital communication to control the errors generated in transmission to improve the accuracy of digital message transmission. The error is mainly caused by random noise generated by the electrical characteristics of the line itself, attenuation or distortion of signal amplitude, frequency and phase, reflection and echo effect of electrical signal on the transmission medium, crosstalk of adjacent lines, external electromagnetic interference and equipment failure.

Error control in digital communication, the coding method is used to control the errors generated in transmission to improve the correctness and effectiveness of transmission. Error control includes error detection, forward error correction and automatic request

retransmission. According to the different nature of error, error control is divided into error control for random error and error control for burst error. Random error means that the channel error is evenly distributed in different time intervals; Burst error means that the channel error is concentrated in a very short period of time. Sometimes several error control methods are mixed, and it is required to have certain error control capability for random error code and burst error code.

We need a method to ensure the integrity and accuracy of the received data. The actual communication line is always imperfect. Data may become disordered or lost during transmission. In order to catch these errors, the modem at the sending end performs a mathematical operation on the data to be sent, and sends the operation result together with the data. The modem receiving the data performs the same operation on the data it receives, and compares the two results. If the data is damaged during transmission, the two results will be inconsistent, and the modem receiving the data will request the sender to resend the data. Errors in the communication process can be roughly divided into two categories: one is random errors caused by thermal noise; the other is the burst error caused by conflict noise. Sudden errors affect the local, while random errors affect the global.

There are three possibilities for automatic request retransmission: the first possibility is affirmative confirmation. If the receiving end does not find any error after checking the received check frame, it will send a positive confirmation signal to the sending end, which is represented by ACK. The second possibility is to confirm. After receiving a frame, the receiving end will send back a negative confirmation signal, expressed by NAK, if it finds an error after verification. The third possibility is a timeout retransmission.

When data transmission fails, retransmission is often used, so a reasonable retransmission mechanism is needed. Timeout retransmission is another important mechanism for the transmission control protocol to ensure data reliability. Its principle is to start a timer after sending a certain data. If the ACK confirmation message of the sent datagram is not received within a certain time, the data will be retransmitted until the transmission is successful.

The most important mechanism in the reliability of transmission control protocol is to handle data timeout and retransmission. The transmission control protocol requires that every message segment sent at the sending end should start a timer and wait for confirmation information; The receiver returns the confirmation message after successfully receiving the new data. If the data is not confirmed before the timeout of the timer, the transmission control protocol considers that the data in the message segment has been lost or damaged, and needs to reorganize and retransmit the data in the message segment. Although the concept of timeout retransmission is very simple, the mechanism of handling timeout retransmission in the transmission control protocol is quite complex compared with other reliability protocols.

The retransmission mechanism is an important transmission method. However, the retransmission mechanism may lead to a long time delay, as well as a large transmission energy consumption, and even enter a dead cycle and eventually lead to node energy depletion. The retransmission shall be conducted at random intervals to improve the success rate. Set a reasonable number of retransmissions. The setting of the number of retransmissions needs to consider the delay or energy issues. For example, set the number

of retransmissions to three. If the retransmission fails several times, another path will be found for data transmission. If no other path can be found for data transmission, increase the transmission power of the node one by one to find other transmission paths. If the node cannot find another path for data transmission at the maximum transmission power, the transmission will be terminated and will not be sent.

The timeout retransmission mechanism is one of the most important and complex technologies in the transmission control protocol, and it is also the guarantee of the data transmission reliability of the transmission control protocol. Each time the sender sends a message segment, the transmission control protocol will reserve a copy, set a timer and wait for the confirmation information. If the timer times out and the data in the sent message segment is not confirmed, the message segment will be retransmitted until the transmission is successful. It can be seen that the calculation of retransmission timeout is the key part of timeout retransmission. The transmission control protocol requires that the current network condition can be roughly estimated.

If there is a timeout, there must be retransmission, but even retransmission has a strategy, rather than simply sending data. The number of retransmissions of transmission control protocol messages also varies according to the system settings. In some systems, a message will only be retransmitted three times. If the message has not been confirmed after three retransmissions, then no retransmission will be attempted and the transmission control protocol connection will be reset directly. However, in some highly demanding business application systems, discarded messages will be retransmitted continuously, to ensure the normal interaction of business data as much as possible.

Due to the unreliability of the IP protocol and the complexity of the network system, a small amount of message loss and retransmission of the transmission control protocol are normal. However, if there are a large number of retransmissions of the transmission control protocol in the process of business interaction, it will seriously affect the efficiency of business system interaction, resulting in slow or even unresponsive business system. In general, a large number of retransmissions of transmission control protocols indicate that the status of network communication is very bad. It is necessary to analyze the causes of packet loss and retransmission from the perspective of network layer.

The reason for retransmission of transmission control protocol is to ensure the reliability of transmission control protocol. It is precisely because of the retransmission mechanism of transmission control protocol that those business applications based on transmission control protocol no longer worry about a series of application problems caused by packet loss and packet damage during network interaction.

2.2 Communication Link Evaluation Based on Big Data

When a path fails to send data, it will involve retransmission after failure. In addition to retransmission, data can also be transmitted through other paths. In general, with the increasing number of nodes in the communication network, there are often multiple links that can be reached from one node to another. How to make full use of multiple links for parallel transmission or as link backup is of great significance. For example, users' mobile phones generally have two network access modes: 5G and WIFI. When users use the network, if they connect to WIFI, they will use WIFI first, but if the intermediate WIFI link is broken, they will switch to 5G and then use it.

For multi-path during transmission, the path needs to be managed so that the link condition can be known and the relevant information can be updated in time when the link is removed or added. From the implementation level, multi-path transmission can be divided into network layer implementation, transport layer implementation and application layer implementation. The implementation of the application layer is the most expensive, because existing applications need to be changed. However, in the face of traffic control and other problems, it is difficult to implement in the network layer. Only in the transport layer can we use the reliability mechanism of the transport control protocol to transform.

Since there are two possibilities for deploying a new communication base station: meeting the coverage requirements and not meeting the coverage requirements, the number of deployed communication base stations X follows a binomial distribution, and q represents the probability of meeting the coverage requirements.

$$\Pr\{X = i\} = \binom{m}{i} q^i (1-q)^{m-i} \tag{1}$$

$$\binom{m}{i} = \frac{m!}{i!(m-i)!} \tag{2}$$

Figure 1 is the statistical analysis of the number of neighbour nodes. The simulation experiment is carried out with MATLAB, and N is the number of nodes. The big data technology is used to count the links with communication retransmission, and the probability of each link with retransmission is obtained as the basis for evaluating different communication links. All decisions should be based on reality and data, with quantitative instead of qualitative and rational instead of perceptual. How to eliminate

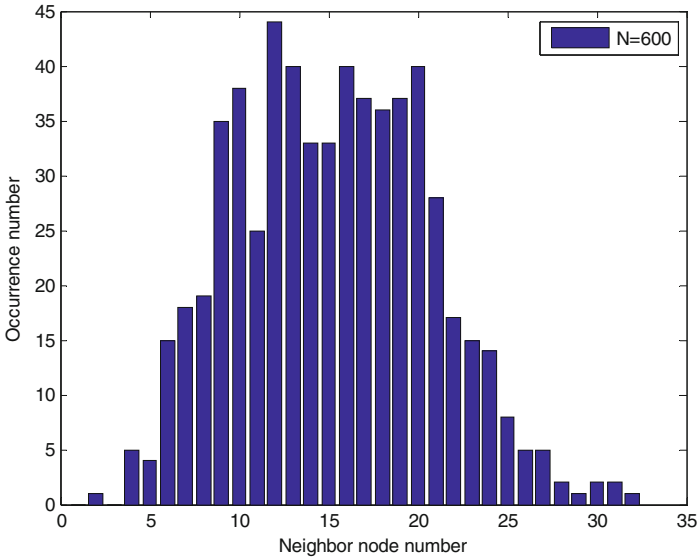

Fig. 1. Statistical analysis of the number of neighbour nodes

the false and retain the true, and how to get the true data from the rough is a key. In this way, big data technology can present the real situation and laws. In the era of big data, researchers can monitor and track the massive behaviour data generated by research objects in real time, conduct mining analysis, reveal the regularity, and draw research conclusions and countermeasures. The transmission path switch is used to switch to the bypass transmission path to send the data to be transmitted if the next node on the first data transmission path fails or the path with the next node fails. Establish the optimal path set path selection method, and realize the automatic path switching in multi-path simultaneous transmission based on the optimal path set method. The more neighbour nodes, the more transmission paths are available. The increase of transmission paths can reduce the number of communication retransmissions, thus reducing the time delay.

3 Conclusion

Through big data technology, the link quality of the communication network and the probability of retransmission of each link can be obtained, and the data transmission path can be planned and switched. When the base station receives the path switching request sent by the transmission node, it will switch to the alternative path according to the location information and status information of the node in the communication network topology. The number of alternative paths depends on the number of neighbour nodes of nodes in the communication network. The multi-path transmission of communication network information combined with big data technology can effectively reduce the number of retransmissions and delay.

References

1. Sauer Molly, Vasudevan Prarthana, Meghani Ankita et al. Situational assessment of adult vaccine preventable disease and the potential for immunization advocacy and policy in low- and middle-income countries[J].Vaccine, 2021, 39(11) :23–36.
2. Kalpna Guleria,Anil Kumar Verma. Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks[J].Wireless Networks,2019,18(3):52–59.
3. Chiara Petrioli, Michele Nati, Paolo Casari, Michele Zorzi, Stefano Basagni. ALBA-R: load-balancing geographic routing around connectivity holes in Wireless Sensor Networks[J]. Parallel and Distributed Systems, 2014,25(3):529-539.
4. Kexin Zhao, Jun Ling, Jian Li. Enhanced Mobile Multiple-Input Multiple-Output Underwater Acoustic Communications[J]. International Journal of Distributed Sensor Networks, 2013, 13(6):91-97.
5. Jing Han, Sundeep Prabhakar Chepuri, Qunfei Zhang et al. Iterative Per-Vector Equalization for Orthogonal Signal-Division Multiplexing Over Time-Varying Underwater Acoustic Channels[J]. IEEE Journal of Oceanic Engineering, 2019, 44(1):32-42.
6. Liangrui Tang, Zhilin Lu, Bing Fan. Energy Efficient and Reliable Routing Algorithm for Wireless Sensors Networks [J]. Applied Sciences,2020,14(5):61-67.
7. Gang Qiao, Muhammad Bilal, Songzuo Liu et al. Symmetry Oriented Covert Acoustic Communication by Mimicking Humpback Whale Song[J]. Symmetry, 2019, 11(6):71-77.