Power Data Sharing and Secure Interaction System Technology Based on k-Nearest Neighbor

Yun Dong [1, a], Zhaoli Chen [1, b], Xixiang Zhang [1, c, *], Liyuan Zhang [1, d], and Xuhua Ai [1, e]

[1] Information Center, China Southern Power Grid Guangxi Power GridCo Ltd, Nan ning, China.

[a] dong_y.xt@gx.csg.cn, [b] chen_zl@im.gx.csg.cn, [c] zhang_xx.xt@gx.csg.cn,

[d] zhang_ly.xt@gx.csg.cn, [e] ai_xh.xt@gx.csg.cn

Abstract. With the increasing demand for electricity, efficient power data transmission is a reliable guarantee for the safe operation of the power grid. In order to effectively realize the fast and reliable transmission of data in the shared security interaction system, the k-Nearest Neighbor will be based on the nodes carrying information and use the opportunity to meet with neighboring nodes to accelerate data transmission. Simulation test results show that when the number of nodes is 600, the maximum delay time of blockchain technology is 24.2 ms, while the maximum delay time of the k-Nearest Neighbor is only 21.9 ms, which is 9.50% lower than that of blockchain technology. When the number of neighboring nodes is 9, the shared interactive system load reaches a minimum of 51.6%, which is 41.03% lower than the number of neighboring nodes 1. Therefore, it is recommended that the number of neighboring nodes of the shared interactive system should be controlled at 9 to ensure the efficiency of power data transmission speed. And the total packet loss of blockchain technology is 87.2 MB, while the packet loss of the data transmission mechanism based on the k-Nearest Neighbor is 5.3 MB, and the corresponding packet loss rates are 43.6% and 2.6%, respectively, which is a 41% saving over blockchain technology. In summary, the shared security interaction system based on the k-Nearest Neighbor can effectively improve the transmission and interaction efficiency of electric power data with the advantages of fast transmission speed and short delay time.

Keywords: k-Nearest Neighbor; delay time; power data; secure interaction; data transmission.

1. Introduction

With the increasing demand for electricity and the continuous growth of electricity data, it further leads to the problems of delay and poor stability of data transmission in the power sharing interactive system network. The power sharing interactive system can collect and transmit intelligent information about power [1], and provide users with power information management applications and value-added services. Therefore, it is very important to ensure the rapid transmission of data in power sharing systems[2].

In power sharing interactive system networks, encryption is usually used to ensure the integrity and security of data transmission. Therefore there is a need to ensure fast and reliable data transmission in order to reduce the cost of secure transmission control, thus reducing the energy consumption of nodes in order to reduce the resource consumption of the power sharing interactive system network [3-4].

One of the scholars, Xu Dong [5], proposed a BP neural network-based method for secure and fast data transmission over long distances. Under the premise of making full use of the limited computing resources of mobile terminals, the data transmission method based on BP neural network is studied, and the computing speed is effectively improved under the premise of ensuring information security. However, when data is transferred, problems such as excessive system load still occur. Scholar Zhu Tong [6] has proposed TEA cryptography is used to realize the security of information transmission in the network. On this basis, a encryption method based on TEA is proposed, which can obtain the encryption index, public key and private key of the user used for communication, so as to achieve the security of data transmission in the network. However, this approach results in a high rate of data loss. Scholar Yanliang Zhang [7] introduced the process of data acquisition and transmission in the environmental sensing application system, and then analyzed the data transmission from the environmental data source to the field control unit. Secure transmission is carried out by connecting to the cloud data center through the Internet.

And the electric power data sharing and interaction system needs to ensure data security and privacy protection. Electricity data involves national energy security and enterprise commercial secrets, and strict security measures must be taken to prevent data from being illegally acquired, tampered with and leaked. And the k-Nearest Neighbor is an instance-based learning method, which determines the most similar samples by calculating the distance between samples. In power data sharing and interaction systems, the k-Nearest Neighbor can be used to find the power data that is most similar to the given query data, thus realizing fast data retrieval and sharing.

Therefore, this paper proposes the use of the k-Nearest Neighbor to improve the secure and fast transmission of power data in shared interaction systems. Using the k-Nearest Neighbor, the data of neighboring nodes in the shared interactive system network is controlled to improve the success rate of data transmission and reduce the delay. And the closest neighbor algorithm is used to control the number of power data transmission by considering the node load on the basis of the connection strength of the shared interactive system network to further reduce the degree of data transmission congestion. At the same time, parameters such as the number of neighboring nodes, the number of lost packets, and the delay variation of the system network are used to evaluate the performance of the power data sharing and interaction system based on the k-Nearest Neighbor.

2. Shared Interactive System Technical Architecture

2.1 Changes in network node connectivity for data sharing and interaction systems

The data sharing system network is a physical mobile communication network consisting of wireless communication lines and communication devices, and its network location changes. As time goes by, the communication distance, connection and disconnection caused by wireless signals and communication modes will change [9-11]. The connection strength of the actual opportunity network is mainly affected by the signal strength, signal interference, communication distance and other factors, and the effect of network connection and node access mode. In the shared interactive system, the strength of connection between nodes is related to the number of encounters, geographical location and access frequency. The connectivity between the system and the communication network determines the connectivity of the entire network w_{ij}.

In evaluating the degree of node connectivity, the physical network of a data-sharing interactive system focuses on the communication coverage area of nodes and the persistence of link maintenance. In the network of shared interactive systems, the main concern is the interconnection attributes such as the frequency of node encounters and the rules followed by the encounters [12]. Generally speaking, in a data-sharing interactive system, the more frequently nodes encounter each other, the longer the connection time. The more fixed the regularity of node movement and encounter, the closer the connection between nodes will be, which will help further improve the efficiency of data transmission. Then the data security metric (SPM) can be used to measure the connection strength w between nodes.

$$SPM = \frac{\int_{t=0}^{T} f(t)dt}{T}, w_{ij} = \frac{1}{SPM_{ij}} \tag{1}$$

SPM_{ij} denotes the data security pressure between nodes i and j, T is the statistical time window, and SPM_{ij} satisfies $0 \leq SPM_{ij} \leq \frac{T}{2}$, $(SPM_{ij})_{max} = \frac{T}{2}$. $f(t)$ denotes the remaining time from the time of the next encounter between the two nodes, when the two nodes are connected, $f(t) = 0$, when two nodes are disconnected, $f(t) = t_2 - t_1$, where t_2 and t_1 denote the time of the next encounter and the current time, respectively [13].

The connection strength between nodes is inversely proportional to the data security index between nodes. A lower measure of data security allows the nodes to be more tightly linked together. When $SPMij=0$, w_{ij} tends to infinity, indicating a strong connection between the nodes; When $SPMij=1$ and $w_{ij}=1$, the link between nodes is weakest at this time. The effect of the value of the statistical window T on the calculation of the data security metric after normalisation is shown below:

$$SPM_{ij}^{'}=\frac{SPM_{ij}}{(SPM_{ij})_{max}}=\frac{2\int_{t=0}^{T}f(t)dt}{T^2},0\le SPM_{ij}^{'}\le 1 \tag{2}$$

After removing the effect of the statistical window T on the data security measure and normalising the data security measure value SPM. Quantify the connection strength w_{ij} as a range of 0-1, i.e., $w_{ij}=1$ indicates the strongest state of connection between nodes and $w_{ij}=0$ indicates the weakest state of connection between nodes. The strength of the connection between the nodes w_{ij} is given by the following equation:

$$w_{ij}=\frac{e^b(1-SPM_{ij}^{'})}{e^b-1},0\le w_{ij}\le 1 \tag{3}$$

where b is a positive coefficient.

2.2 Power Data Transmission Enhancement Based on k-Nearest Neighbor

In order to further improve the power data transmission efficiency, effectively increase the success rate of data delivery, reduce the delay, and further reduce the shared interactive system network overhead. The probability of power data leaving the current bearer node is estimated by the nearest-neighbour algorithm, and the expansion of the current shared interactive system network is also estimated to determine the replication and transmission strategy of power data [14]. The transmission method of power data is adaptively controlled to control the amount of power data. When the node strength w_{ij} and threshold w_{th} satisfy $w_{ij} \ge w_{th}$, the node transmits a single class of power data. And when $w_{ij} < w_{th}$, the node transmits multi-class power data.

However, when power data are securely transmitted for interaction, the transmission conditions should meet the requirements of interconnection between nodes, and the longer the node contact time to complete the correct transmission of information. Therefore, when assessing the possibility of transferring electrical energy data from a node to the next node, it is necessary to take into account the factors affecting the information transfer, such as node movement characteristics and link characteristics of the shared system [15-17].

Using the theory of k-Nearest Neighbor, the contact interval Δt_{ij} between nodes i and j in the network of shared interaction system is set to conform to the episodic exponential distribution with parameter λ_{ij}, and λ_{ij} is used to denote the contact probability between nodes i and j. t_{pre} and t denote the previous contact time and the next contact time of nodes i and j. Therefore the contact probability density function of nodes i and j at time t can be expressed as:

$$fij(t)=\lambda_{ij}e\text{-}\lambda ij(t-t_{pre}) \tag{4}$$

When nodes i and j are in contact, whether the information is transmitted successfully or is affected by factors such as the length of node contact time and the stability of the wireless signal of the shared interaction system. Where, B_0 delegates transferring information between nodes, when there is a wireless communication link between nodes, the probability of successful transmission of power data m in the opportunity communication network, Q_{ij}, is expressed as follows:

$$Q_{ij}=e-\lambda ij\frac{m}{B_0} \tag{5}$$

Starting from the current time t_1, in the future time period *C*, considering that node *i* has only one neighbour node *j*. The probability that the power data *m* successfully leaves node *i* and is transmitted to the other node *j* can be expressed as:

$$P_{ij}(m) = \int_{l_1}^{l_1+C} Q_{ij}(m) fij(t) dt = \int_{l_1}^{l_1+C} \lambda_{ij} e^{-\lambda_{ij}\frac{m}{B_0}(t-t_{pre})} dt \tag{6}$$

Assuming that node *i* has *n* neighbouring nodes, if the number of neighbouring nodes varies from 1 to *n*, then the probability of power data m successfully leaving the node within the time range of *C* is represented by the probability expectation that power data m is delivered to all neighbouring nodes. It is shown in equation (7):

$$Pi(m) = \sum_{j=1}^{n} P_{ij}(m) r_{ij} \tag{7}$$

Where the weight of neighbour node *j* in the list of all neighbour nodes of *i* is denoted by r_{ij} and combining with Eq. (3) yields the following result:

$$r_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \tag{8}$$

2.3 Reducing the probability of congestion in power data transmission

Multi-class power data can increase the rate of secure data delivery and reduce the delay in information delivery. Single class power data will further increase the burden on the shared interactive system network compared to multi-class power data. In addition, the k-Nearest Neighbor will speed up the data transmission based on the nodes carrying information and taking advantage of encounters with neighbouring nodes [18]. Therefore, power data sharing will be more prone to congestion in a shared interactive system network. In order to reduce the congestion probability of data transmission, the k-Nearest Neighbor is used to control the number of power data transmission by considering the node load based on the connection strength of the shared interactive system network. And the degree of congestion of nodes should be further considered when transmitting data to optimise the transmission queue of power data [19]. The congestion degree is related to the size of the remaining cache space and the service capacity of the node. The congestion degree of node *i* is measured by using the ratio of the node's data receiving speed s_i to the remaining memory capacity bi at time *Δt*, called Ci. i.e., it can be expressed in equation (9):

$$C_t = \frac{S_t}{b_t} \tag{9}$$

The node degrees of freedom are denoted as:

$$F_t = 1 - \frac{s_t}{b_t} \tag{10}$$

According to Eqs. (9) and (10), considering the probability of successful transmission of power data *m* and leaving node *i* . And pmi_st/CC denotes the transmission of power data m in the shared interactive system network, then the probability $p_{s\text{-}i}(m)$ of successful transmission of data *m* through node *i* at time *C* can be obtained.

When node *i* encounters a neighbour node *j*, it first sorts the $p_{s\text{-}i}(m)$ values of the *n* data to be transmitted in the cache of node *i* from the largest to the smallest to obtain the data sequence $List = \{m_1、m_2、\ldots、m_k, \ldots m_n\}$. Meanwhile, it uses the neighbour node *j* to compute its congestion level F_j and selects the $k = [n F_j]$ data as the preferred data to be transmitted from the list of power data sequences [20].

Define the utility value of the data in the list as $U_m = P_{s-i}(m) hOP_m$. hOP_m is the number of leap points through which the power data m passes. Based on the utility value of the power data, optimize the transmission order of the data and further solve the list of optimized data transmission sets.

$$\begin{cases} maxT = \sum_{k=1}^{z} u_k \chi_k \\ s.t. \sum_{k=1}^{z} u_k \chi_k \le \theta_t, \chi_k \in \{0,1\} \end{cases} \tag{11}$$

Where θ_j denotes the cache residual space - $\chi_k \in \{0,1\}$ of node j, and its values 1 and 0 indicate whether node i chooses to put the data into the transmission queue of the shared interactive system, respectively. In order to solve the optimized power data transmission order list, the utility value of each message in the alternative set needs to be calculated first. Based on the ratio of the utility value of power data to the length of power data, the power data in the alternative set is sorted in descending order and the power data m is viewed sequentially.

3. Experimental Results

3.1 Experimental environment setting

In order to test the availability, suitability and stability of the data transmission method of the nearest neighbor algorithm in practical applications, this study will set up an experimental platform according to relevant conditions, and verify its effect through data. In order to simulate the specific situation of the data transmission system in actual operation, MATLAB R2020b version is selected as the development environment of the simulation experiment.

The computing device used in the experiment had an Intel Core i7-10700 processor, a clock rate of 2.9 GHz, 16GB of RAM, and ran on a 64-bit Windows 10 operating system. In addition, the Depth Computing Unit (DCU) was introduced as an auxiliary device for performance enhancement. DCU is a graphics processor that runs on the Radeon Open Computer (ROCm) platform, supports OpenCL technology, and is compatible with major heterogeneous computing frameworks such as CUDA. In this architecture, the CPU acts as the main control unit and the DCU acts as the execution unit, and the two are interconnected through the PCIE bus. And in order to make the experimental results more in line with the requirements of safe transmission and protection of power data, when the data transmission facility is difficult to handle (processing delays or high system resource occupancy), the power data packets can be directly offloaded to the corresponding cloud servers in order to improve the operational efficiency of the system network.

3.2 Effect of the number of neighbouring nodes on the load of the data interaction system

Fig. 1 shows the effect of different number of neighbouring nodes on the load of shared interactive system. The simulation results show that when more neighbouring nodes are included in the data hierarchical transmission process, the load of the shared interaction system shows a decreasing and then increasing trend. When there is only one neighbouring node, the system load is about 87.5%, and the more neighbouring nodes, the lower the system load. When the number of neighbouring nodes is nine, the shared interactive system load reaches a minimum of 51.6%, which is 41.03% lower than when the number of neighbouring nodes is one. After that, as the number of neighbouring nodes continues to increase, the shared interactive system load stabilises, but the load increases compared to when the number of neighbouring nodes is 7. The average system load is 69.5% when the number of nodes is 11, 13 and 15. It decreases by 21.14%, 20.10% and 25.71% from the number of neighbouring nodes of 1. This is mainly due to the fact that the total number of packets in the network of the shared interactive system is fixed.

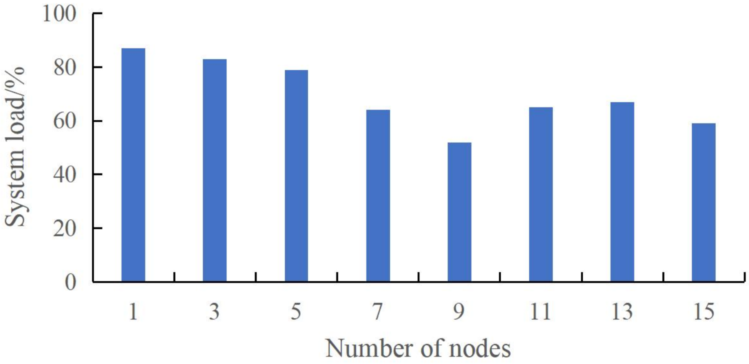

Fig. 1 Load variation of data sharing interaction system under different numbers of nodes

When the number of neighbouring nodes reaches a certain number, it can help the shared interactive system to speed up the transmission of data in the packets. And as the number of neighbouring nodes increases, the load of the system gradually decreases, and the lower system load will further improve the power data transmission speed and the stability of data transmission. Therefore, it is recommended that the number of neighbouring nodes of the shared interactive system be controlled at 9 to ensure the efficiency of power data transmission speed.

3.3 Change in number of lost packets

In order to evaluate the packet loss variations of the shared data interaction system in transmitting power data, the data transmission based on blockchain technology and the k-Nearest Neighbor proposed in this paper are used as a comparison. The two data transmission methods are performed in the same experimental environment with the same data type and data size. The packet loss performance statistics and calculations for the same data are shown in Table 1. The experimental results show that the packet rate of blockchain technology is 87.2 MB, while the packet loss rate of k-nearest-neighbour based data transmission method is only 5.3 MB, which is a larger decrease than the packet loss rate of blockchain technology.And it can be observed that the variation of data packet loss of both blockchain technology and k-Nearest Neighbor is not affected by the number of data transmissions, and the smallest data received size of blockchain technology is only 274.9 MB, while the k-Nearest Neighbor is as high as 298.4 MB.The main reason for the low number of packets lost by the k-Nearest Neighbor of this paper is that, the k-Nearest Neighbor, when it first accesses to the network of the shared interactive system, selects all the paths for the data transmission.

Table 1. Results of packet loss rate experiment

Number of data transfers	Data transfer size /MB	Data reception in blockchain technology /MB	Data reception for the k-Nearest Neighbor /MB
1	300	283.3	300
2	300	292.2	300
3	300	287.6	298.4
4	300	274.9	300
5	300	298.6	300
6	300	293.3	297.2
7	300	300	300
8	300	282.9	299.1

3.4 Latency Variation in Shared Interactive System Networks

Real-time data transfer is an important measure of network overhead for shared interactive systems. The average network latency for data transfer represents the average data transfer time from all nodes to Sink node. The k-Nearest Neighbor is also compared with blockchain technology in terms of average delay of data system network. The experimental results are shown in Fig. 2.

From the figure, it is clear that the average delay of the k-Nearest Neighbor is the lowest. The main reason is that using the k-Nearest Neighbor to control the number of power data transmission by considering the node load on the basis of the network connection strength of the shared interactive system. And the degree of node congestion should be further considered when transmitting data to optimise the transmission queue of power data and achieve point-to-point data transmission to improve the data transmission speed. When the number of nodes is 100-200, the delay time of the two algorithmic techniques increases more slowly, and when the number of nodes is greater than 300, the delay time of blockchain technology increases more, and when the number of nodes is 600, the maximum delay time is 24.2 ms. while the maximum delay time of the k-Nearest Neighbor is only 21.9 ms, which is 9.50% lower than that of blockchain technology, mainly because in the blockchain technology, as when the number of nodes in the system network increases, it further leads to an increase in the number of neighbouring nodes and the amount of data in the system. This leads to a corresponding increase in the data delay time. The delay time of the k-Nearest Neighbors varies more gently with an average delay time of 21.25 ms as compared to 22.75 ms in blockchain technology and it can be observed from Fig. 2 that the overall delay of all the algorithms increases with the increase in the number of nodes in the network of the shared interactive system. This is because the increase in the number of nodes leads to increase in data and sub-routing which in turn leads to increase in the delay of data transmission from source node to Sink. The probability of transferring data from the current node to other nodes is estimated using the k-nearest neighbour method with the current power sharing system, which determines the data transfer strategy, thus further reducing the data latency time in the sharing system.

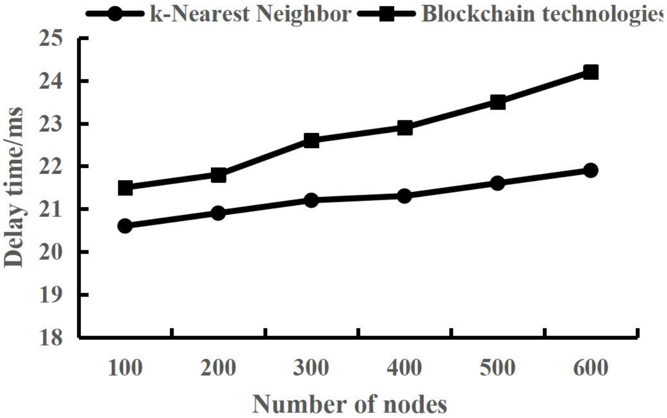

Fig. 2 Delay variation of shared interaction system network

4. Summary

When the data hierarchical transmission process includes more neighbouring nodes, the system load tends to decrease and then increase. When the number of nodes is greater than 9, the shared interactive system load tends to stabilize, but the load increases compared to when the number of neighbouring nodes is 7. The average system load is 69.5% when the number of nodes is 11, 13 and 15. It decreases by 21.14%, 20.10% and 25.71% from the number of neighbouring nodes 1, respectively. At a node count of 100-200, the delay time increases more slowly for both algorithmic techniques. When the number of nodes is greater than 300, the increase in delay time of the blockchain technique increases, and at a node count of 600, the maximum delay time is 24.2 ms, while the maximum delay time of the k-Nearest Neighbor is only 21.9 ms.

Acknowledgement

This research was supported by the Guangxi Power Grid Technology Project under Grant 046100KK52222001.

References

[1] Su Jiangwen, Huang Xiaoguang, Zhang Yao et al. Research and application of cloud computing-based artificial intelligence operation platform and edge terminal interaction mechanism[J]. Mechanical Design and Manufacturing Engineering,2023,52(08):130-134.

[2] Lan Yingbin. Wideband carrier data interaction method for low voltage distribution network stations[J]. Changjiang Information and Communication,2023,36(02):100-102.

[3] Wang Weixiong,Ling Youcheng. Application study of point-to-point VLC direct-connect broadband transmission system in 5G mobile network optical transmission[J]. Jiangxi Communication Science and Technology,2022,(04):13-16.

[4] Xueda Huang,Junjie Yang,Zhipeng Hou et al. A signalling and data transfer method based on Windows USB Mass Storage class[J]. Internet of Things Technology,2022,12(07):33-37.

[5] Xu D. Design of industrial internet test data interaction and integration system based on web interface[J]. Electronic Design Engineering,2021,29(17):40-44.

[6] Zhu Tong,Liu Yuan.Design development and application of DMZ bidirectional CS mode safety data interaction[J]. Automation Instrumentation,2021,42(04):106-110.

[7] Zhang Yanliang,Liu Yang,Zhang Guoliang et al. Analysis of factors affecting data remote transmission speed based on handshake interaction protocol[J]. Digital Technology and Application,2021,39(01):41-43.

[8] Zhao Jingming,Zhang Jie,SONG Peng et al. An efficient power data collection system based on cloud-side-end collaboration[J]. Grid and Clean Energy,2022,38(05):49-55.

[9] Qiang Tang. A method and design of Wi-Fi power data acquisition based on broadcast transmission network[J]. Electronic Technology and Software Engineering,2020,(24):13-14.

[10] Lei Shi, Youqi Jiang, Yong Lou. Research on composite material defect detection and transmission network congestion under bandwidth limitation[J]. Bonding,2021,46(05):54-57.

[11] Ren J, Wang TY. Secure transmission of power dispatch data network based on power communication network[J]. Science and technology wind,2019,(36):176.

[12] Wang Jinfeng,Wang Jin. Research and implementation of security of electric power data collection network for internet of things[J]. Information and Computer(Theoretical Edition),2011,(20):140-141+143.

[13] Yuanfang,Li Dong,Li Wencui et al. Design of power data integration system based on STM32[J]. Automation and Instrumentation,2022,(01):172-175+180.

[14] Zhang Yueyuan. Research on meta-modelling and data assurance based on big data[J]. Bonding,2021,45(03):96-100.

[15] Qiang Tang. A method and design of Wi-Fi power data acquisition based on broadcast transmission network[J]. Electronic Technology and Software Engineering,2020,(24):13-14.

[16] Y. Z. Song, L. Xi, W. Cui, et al. Research on multi-level encryption of power grid privacy data based on hidden classification algorithm[J]. Microcomputer Applications,2023,39(11):60-64+68.

[17] Jing Haobin, Yang Jie, SUN Jun et al. Three-dimensional LiDAR data processing method adapted to different distances[J]. Journal of Chongqing University of Technology(Natural Science),2020,34(08):22-28.

[18] Wand Pin, Zhao Yiqiang, Liu Yanjiang et al. Research on hardware Trojan horse electromagnetic information detection technology based on the fusion of Gaussian filtering and K-nearest neighbor algorithm[J]. Journal of Nanjing University(Natural Science),2020,56(02):264-269.

[19] Zhong Yunnan,Chen Xingbin,Zhan Tongping. Research on hierarchical clustering technology and data flow load balancing[J]. Bonding,2022,49(12):182-185.

[20] Xu Jiunyun,Sun Shan. Delay-protected transmission strategy for heterogeneous converged data in the Internet of Things[J]. Microelectronics and Computers,2021,38(07):42-47.