Research on Scientific Research Data Platform Based on Big Data

Ji Yuxiang
State Grid Jibei Electric Power Company Limited Skills Training Center, Baoding Electric Power VOC.& TECH. College, Baoding, China

Abstract. In order to realize the fault tolerance of real-time tasks in scientific research data platform information platform, based on adaptive feedback equalization and symbol modulation technology, a fault-tolerant scheduling model of real-time tasks in scientific research data platform information is proposed in the paper. What is more, the platform transmission channel model of the scientific research data platform information under routing conflicts is first constructed to optimize the information transmission protocol of the scientific research data platform. Secondly, the fuzzy C-means clustering method is applied to perform the inspection robot information fusion. Meanwhile, the symbol modulation method is used for fault-tolerant scheduling of real-time tasks in the information platform. Finally, a simulation experiment is carried out. The results show that the method proposed in the paper has better fault tolerance for platform real-time task scheduling of scientific research data platform information, and the channel balance of the information platform is stronger, which improves the platform real-time task scheduling capability of scientific research data platform information.

1. Introduction

Science and technology information has played an important role in the formulation and implementation of national, social, and corporate strategies and plans, and the prediction of scientific research hotspots is a relatively new application requirement in the field of scientific and technological information [1]. Therefore, scientific research workers and scientific research project managers must be forward-looking when selecting topics and initiating projects, that is, based on the current status of science and technology as well as social development, judgments are made on new theories that may be generated in the future or new technologies that generate application value [2]. The current research hotspot prediction models established at home and abroad still have the following problems in terms of application effects: 1) When a new theory and technology is born, its related application fields still need a lot of work to explore. 2) In the field of scientific research, there are a huge number of scientific research hotspot words, and the trend of each hotspot is not the same. The accuracy of fitting hot trends with standard machine learning models is not high. Therefore, it is urgent to design a way to ensure the independent operation of key partitions through resource division at the system layer, which determines the minimum resource extreme value according to the partition utilization and improves the reliability and accuracy of task scheduling in the safe area, so that the efficiency of rational use of operating system resources can be improved [3-4].

In terms of the problems mentioned above, a fault-tolerant scheduling model for scientific research data platform information based on adaptive feedback equalization and symbol modulation technology is proposed in the paper. The scientific research data information transmission channel model is first constructed to establish the transmission optimization protocol of scientific research data platform
information. Meanwhile, the fuzzy C-means clustering method is used to fuse scientific research data performance testing when designing simulation experiments for verification. The experimental results show that the method designed in the paper is conducive to real-time task scheduling of scientific research information and has superior performance in fault-tolerant scheduling.

2. Research Data Information Transmission Channel Model and Transmission Protocol Optimization

2.1. Research Data Platform Information Transmission Channel Model

In order to realize the fault-tolerant scheduling of real-time tasks of scientific research data platform information, a transmission channel model is established, and the scientific research data platform is constructed in the routing conflict mode, so that the task transmission link set of scientific research data platform information can be analyzed[5]. Meanwhile, the matched filter detection method is used for the real-time task scheduling and interference suppression of scientific research data platform information, and the transmission channel model is obtained as well[6].

The fault-tolerant scheduling model of real-time task in scientific research data platform information is used for output conversion control, and the transmission channel equalization method is used for output conversion control of the real-time task in the scientific research data platform information. The transfer function of the construction task scheduling is expressed as follows:

\[ f_{ij} = w_{ij} \delta_i + w_{ij} \delta_j + w_{ij} \delta_2 + w_{ij} \delta_4 \]  (1)

Among them, \( w_{ij} + w_{ij} + w_{ij} + w_{ij} = 1 \). Besides, \( t \) refers to the interval distribution of the time required for the real-time tasks of the scientific research data platform information, namely, time; \( c \) represents the storage overhead cost in the task scheduling and delivery process, namely, cost; \( q \) indicates the task scheduling quality result, namely, quality; \( s \) is the security performance characteristics of information real-time task scheduling of scientific research data platform. In addition, considering the factors of network intelligence effect, the implementation strategy of channel balanced allocation is proposed to build a diversity and convergence model for fault-tolerant scheduling of real-time detection tasks under the condition of maximizing constraints[7]. Moreover, given the comprehensive distribution status of task traffic characteristics, association rule algorithm is used to perform fault-tolerant scheduling of real-time task in scientific research data platform information, the iterative equation is obtained as:

\[ x_{id}^{t+1} = x_{id}^t + aV_{id}^{t+1} \]  (2)

A channel model for information transmission of scientific research data platforms is constructed under routing conflicts. Besides, considering the factor of the task scheduling information data transmission performance in the information intelligent management system, the conversion result control needs to be performed during the output process, and the least square planning model expression of the task scheduling is obtained as follows:

\[
\begin{align*}
\varphi_1 &= \varphi_1 c_1 \\
\varphi_2 &= \varphi_2 c_2 \\
\varphi &= \varphi_1 + \varphi_2 \\
p &= \frac{\varphi_1 p_1 + \varphi_2 q}{\varphi_1 + \varphi_2}
\end{align*}
\]  (3)

The ZigBee sensor networking design method is applied to design the Internet of Things for the information real-time task scheduling of the scientific research data platform, and the priority attribute distribution of the data center task scheduling is as follows:

\[
\begin{align*}
\nu(t+1) &= \nu(t) + \varphi(p - x(t)) \\
x(t+1) &= x(t) + \nu(t+1)
\end{align*}
\]  (4)  (5)

According to the analysis mentioned above, a transmission channel model of scientific research data platform information is established to realize scientific research information transmission optimization and fault-tolerant scheduling under the optimal transmission control protocol[8].
2.2. Task Transfer Protocol Optimization

The real-time task transmission model of the scientific research data platform information is analyzed to optimize the information transmission protocol of the scientific research data platform. Combining the adaptive feedback equalization control method, the transmission channel equalization of the scientific research data platform is designed. The forwarding control protocol of the real-time task fault-tolerant scheduling is described as follows:

\[ x(t+1) + (\varphi - w)x(t) = \varphi y \]  

Under the control of maximizing network throughput, the optimal solution set evaluation of the information real-time task scheduling in the scientific research data platform is performed, and the optimal solution set evaluation feature quantity is obtained to satisfy \( u(t) \in L^2(\mathbb{R}) \). Then the evaluation function of \( u(t) \) is defined as:

\[ \phi_{ss}(s, \tau) = \left| \chi_{ss}(s, \tau) \right|^2 \]  

\[ X_{ss}(s, \tau) = \sqrt{\int_{\tau}^{\infty} u(t) \mu \left[ s(t-\tau) \right] dt} \]  

The fuzzy association rule scheduling method is used to design the optimal performance scheduling structure of scientific research data platform information transmission under routing conflicts, and the associated simulation feature is selected from the big data. Meanwhile, the data transmission of scientific research data information is optimized in a uniformly distributed vector set \([9]\). \( t/a \) is used to replace \( t \) to perform random equilibrium control of scientific research data information output, and the adaptive function of task fault-tolerant scheduling obtaining scientific research data information is defined as:

\[ \theta(t) = 2\pi \int_{t-\tau}^{t+\tau} \left( \frac{K}{t_s - t} \right) dt' = -2\pi \ln(1 - \frac{t}{at_s}) + \theta_i \]  

Among them, \( \theta_i = -2\pi K \ln(1 + \frac{T}{2t_s}) \). Under routing conflicts, the statistical characteristic quantities of real-time task scheduling obtained from scientific research data platform information are:

\[ \mu = \sum_{i=1}^{L} \lambda_i / \sum_{j=1}^{H} \lambda_j (L \leq H) \]  

Among them, the subscript \( i \) represents the task transmission sequence, which judges the priority of the task transmission sequence \([10]\). To calculate the queuing delay, the constraint relationship of the priority attribute of the task flow is:

\[ \rho(y|\alpha, \theta) = \sum_{i=1}^{K} \alpha_i p_i(y|\mu, \sum_i\lambda_i) \]  

As a result, a transmission optimization protocol of scientific research data platform information is constructed, and task scheduling is performed according to the optimized transmission protocol.

3. Simulation Experiment and Result Analysis

3.1. Experimental Environment and Experimental Indicators

In order to verify the feasibility of the method designed in the paper as well as the application performance when realizing fault tolerant scheduling of real-time tasks in scientific research data platform information, the simulation experiment analysis is carried out with Matlab. Moreover, the experimental data is selected from iResearch data, and the database website is http://www.iresearch.cn. Besides, the information transmission delay of the scientific research data platform is 12.6ms, and the bandwidth of real-time task fault-tolerant scheduling is 12dB~20dB. In addition, the frequency of sampling information on the scientific research data platform is 16KHz, and the amount of data for fault-tolerant scheduling in real-time tasks has increased from 100 MB to 1 GB.

The experimental indicators are as follows:
Delay in task fault-tolerant scheduling: Delay in task fault-tolerant scheduling refers to the time required for a message or packet to be transmitted from one end of a network to another during the information scheduling process, which sends task data packets to a website terminal to see whether the receiver meets the restriction requirements so as to achieve the performance of network scheduling.

The delay calculation formula for task fault-tolerant scheduling is as follows:

\[ f = \frac{l}{v} \]  

In formula 12, \( f \) represents the fault-tolerant scheduling delay, \( l \) refers to the data frame length, and \( v \) indicates the scheduling rate.

Scheduling task scale: The scale of the scheduling task refers to allocating the qualified network allocation to the reasonable range of the specified task under certain constraint conditions, which is used as an index to judge the scheduling accuracy.

Fault tolerance success rate: The fault tolerance rate is the probability of allowing mistakes to occur. If the success rate of fault tolerance is guaranteed, the reliability of scheduling will be increased to a certain extent. Additionally, fault tolerance success rate is calculated by inputting relevant data into EXCEL function.

3.2. Comparative Analysis of the Delay in Task Fault-tolerant Scheduling

The method proposed in the paper is compared and analyzed with the heuristic algorithm of dynamic real-time fault-tolerant scheduling in literature [5], the one-time fault-tolerant scheduling algorithm for critical data in literature [6] and the safety-critical task scheduling method in literature [7]. According to the simulation environment and parameter settings mentioned above, the real-time task of fault-tolerant scheduling in scientific research data platform information is performed, and the scheduling sequence is shown in Figure 1.

The feedback equalization method is used for the channel equalization design of scientific research data information transmission, and the symbol modulation method is applied for the fault-tolerant scheduling of information in real-time tasks. According to the optimized scheduling results, the output mode is shown in Figure 2:
According to Figure 2, the method proposed in the paper can effectively realize the real-time task scheduling of scientific research data platform information, and the scheduling has better fault tolerance. Meanwhile, the scheduling delay is tested, and the comparison results of the fault-tolerant scheduling delay of real-time task are shown in Table 1.

| Task size/MBit | Method of this article | Dynamic real-time fault-tolerant scheduling heuristic algorithm | One-time fault-tolerant scheduling algorithm for critical data | Safety-critical task scheduling method |
|---------------|------------------------|-------------------------------------------------------------|------------------------------------------------------------|--------------------------------------|
| 100           | 0.14                   | 0.54                                                        | 0.69                                                       | 0.66                                 |
| 200           | 0.21                   | 0.79                                                        | 0.98                                                       | 0.89                                 |
| 300           | 0.37                   | 0.80                                                        | 1.12                                                       | 0.96                                 |
| 400           | 0.48                   | 1.21                                                        | 1.34                                                       | 1.25                                 |

It can be seen from Table 1 that the method proposed in the paper has a shorter delay for real-time tasks of fault-tolerant scheduling in scientific research data platform information and better scheduling performance, while the average scheduling delay of dynamic real-time fault-tolerant scheduling heuristic algorithm, critical data one-time fault-tolerant scheduling algorithm, and safety-critical task scheduling method is 4 to 5 times that of the method proposed in the paper. The main reason is that the method proposed in the paper constructs a transmission channel model of scientific research data platform information to realize scientific research information transmission optimization and fault-tolerant scheduling under the optimal transmission control protocol, which ensures the efficiency of task fault-tolerant scheduling and reduces the scheduling delay.

3.3. Comparative Analysis of Test Scheduling Task Scale

The scale of the scheduling task is tested, and the comparison result is shown in Figure 3. The analysis shows that the method proposed in the paper has a higher accuracy for the fault-tolerant scheduling of the information real-time task in the scientific research data platform.

It can be seen from Figure 3 that when increasing task scale, the scheduling accuracy of the method proposed in the paper is close to 100%, and the scheduling accuracy of dynamic real-time fault-tolerant scheduling heuristic algorithm, critical data one-time fault-tolerant scheduling algorithm, and safety-critical task scheduling method has decreased trend. The main reason is that the method proposed in the paper has better fault tolerance for real-time task scheduling of scientific research data platform information and stronger information channel balance. Moreover, experiment with the method proposed in the paper is performed. It is proved that the method proposed in the paper has strong scheduling performance on scientific research data platform. Meanwhile, the real-time task control is effective, and
task scheduling accuracy is better, which controls the scale of test scheduling tasks within a reasonable range.

4. Conclusion
Scientific research data platform information is designed to improve the information processing and integrated scheduling capabilities of scientific research data platform by optimizing the scheduling of tasks of scientific research data platform information. Therefore, based on adaptive feedback equalization and symbol modulation technology, a fault-tolerant scheduling model of real-time task in scientific research data platform information is proposed in the paper, where fuzzy association rule scheduling method is used to design the optimal scheduling of scientific research data platform information transmission under routing conflicts. Meanwhile, the associated feature quantity of the task to be scheduled is extracted, and combining the adaptive feedback equalization method, the channel equalization design and fault-tolerant scheduling of scientific research data information transmission are performed. Besides, the innovations of the research algorithm in this paper are as follows: One is that a scientific research data information fusion model is constructed in the paper; the other is that the task priority of the virtual tasks is controlled in the scientific research data platform information. According to the research, the method in this paper has better real-time and fault tolerance for information task scheduling of scientific research data platform.

References
[1] Hua Songyi, Zhang Yuchen, Ji Pengfei. Research on the construction of a unified scientific and technological information sharing platform for scientific research institutes [J]. Information Exploration, 2020(5): 53-58.
[2] Wang Zhenhuang, Chen Siming, Yuan Xiaoru. Research on Visual Analysis of Weibo Topics [J]. Journal of Software, 2018, 29(4): 1115-1130.
[3] Wu Lulu. Analysis of research hotspots based on topic model [D]. Chengdu: University of Electronic Science and Technology of China, 2018.
[4] Wang Shengyi, Lin Xiaokun, Yu Zhen, et al. Comprehensive intelligent prediction method of transmission line project cost based on clustering [J]. Shandong Electric Power Technology, 2019, 46(5): 12-16.
[5] Zhu Liang, Meng Xianxue, Zhao Ruixue, et al. Analysis on the resource construction of the National Agricultural Science Data Sharing Center [J]. Digital Library Forum, 2017(11): 15-20.
[6] Zhao Hongwei, Cui Yunpeng. Construction of agricultural science and technology work data collection system [J]. Journal of Inner Mongolia Agricultural University (Natural Science Edition), 2019, 40(4): 85-93.
[7] Wang Miye, Zheng Tao, Li Nan, etc. Medical big data integration and application platform system construction [J]. Journal of Medical Informatics, 2019, 40(8): 37-42.
[8] Liu Jingyu, Huang Jinxia, Wang Fang, et al. Research on the Description Framework of the Rights and Interests of Digital Resources [J]. Digital Library Forum, 2019(9): 9-15.
[9] Liu Bosong. Research on Academic Big Data Knowledge Organization and Service Model. Hangzhou: Zhejiang University Press, 2018: 22-23.
[10] Qian Li, Xie Jing, Chang Zhijun, etc. Research and Design of Intelligent Knowledge Service System Based on Big Data of Science and Technology [J]. Modern Library and Information Technology, 2019, 3(1): 4-14.