Research on Computer Assisted Collaborative Scheduling of Self-Optimization Collection by Artificial Intelligence

Zhongan Ding¹, *, Shian Zhan¹, Qian Chen¹, Dahai Zhuang²

¹State Grid Fujian Electric Power Co, Ltd. Marketing Service Center, Fuzhou Fujian 350013, China
²State Grid Info-Telecom Great Power Science and Technology Co., LTD, Fuzhou Fujian 350003, China

*Corresponding author: zhonganding@fj.sgcc.com.cn

Abstract. Based on the energy Internet, the paper studies the interest needs of energy Internet providers and the service needs of different users from the perspective of data collection "convenience, interaction, and automation", and realizes the interconnection of customer energy information and new collection systems. Based on the acquisition master station, communication channel, energy Internet equipment papers to achieve the existing acquisition system architecture design. Provide technical reference for the reasonable planning and collaborative scheduling of power big data wide-area concurrent computing task flow, and comprehensively improve the processing efficiency of power big data wide-area concurrent computing task flow.

Keywords: Artificial intelligence, power data collection, task scheduling, collection technology optimization, data collection system.

1. Introduction

In view of the increasing number and types of internal network equipment such as distribution network operation monitoring equipment and distributed power monitoring equipment, as well as external network equipment such as smart homes and mobile operating terminals, the data also shows massive, polymorphic, and relevance. The unique identification of the equipment, to achieve the purpose of efficient application integration and data sharing; research to meet the adaptation services of the server-side protocol, realize the conversion from the Internet communication protocol to the standard protocol, and realize the ubiquitous access and real-time perception of energy Internet equipment, Dynamic control and information services; research the high-concurrency communication technology architecture with CPS attributes, realize parallel processing of data, communication layer channel link monitoring, and flexible and scalable communication architecture; realize time-sensitive data transmission based on traditional Ethernet, from time to time Research on time-sensitive networks in synchronization, flow control, path control, management mechanisms, etc., to ensure the deterministic delay of important data transmission in the new collection system [1]. Study the communication media decoupling technology of the communication protocol, establish the object model of the application layer protocol for the data collection and business interaction of the client-side equipment, and improve the ubiquitous IoT capability of the communication protocol.
The article thoroughly investigates the real-time computing requirements of the Energy Internet, such as power outages and repairs and other scenarios that are sensitive to power outage data, and studies the online real-time processing technology of energy Internet big data, including distributed caching technology, distributed message queue technology, stream processing, memory computing, Parallel computing, load balancing technology, etc.; in-depth research on the batch processing requirements of the Energy Internet, such as theoretical line loss calculation, user power behaviour analysis, power theft inspection, and other scenarios that require massive data calculations. Research on the offline processing technology of energy Internet big data, including Distributed storage design, distributed parallel computing, multi-machine resource management and task dynamic management, HBase, Hive.

The thesis first analyses the business requirements of the new collection system for the energy Internet, builds the collection standards for the energy industry, analyses the problems caused by the new equipment, and determines the new system design requirements. Secondly, according to business needs, design the physical architecture, logical architecture and security protection architecture of the new collection system for the energy Internet [2]. Third, the research can be based on the plug-and-play distributed collection technology that can be dynamically deployed online based on the scale of access to realize the automatic expansion of the system scale; the research is based on distributed caching, distributed publish and subscribe messages, stream processing, and parallel computing frameworks. The system architecture of data technology for real-time calculation and query of massive data. Finally, study the collection task collaborative scheduling and collection strategy self-optimization technology, according to the collection priority and demand collection task coordination scheduling, and continuously self-learning and optimizing the collection strategy according to the actual operation situation, to achieve efficient and reliable collection of all kinds of energy equipment full data.

2. Principles of Distributed Data Collection

The distributed data collection method applies grid technology to the data collection of the integrated electricity consumption information collection system. Grid technology has powerful network resource integration capabilities. If applied to the power system, it can promote the sharing of information and resources between different dispatching systems, and can become a support platform for the calculation and simulation of the WAN distributed power system. Grid technology is used as a technical support platform to build a new power consumption information collection system for the energy Internet. Sharing resources and collaborative analysis can ensure the safe and stable operation and control of the power grid (Figure 1 shows the principle diagram of distributed data collection).

The introduction of power grid technology is of great significance to solve the data sharing and calculation analysis problems of China's super-large power grids [3]. The use of grid technology can dynamically share resources, including computing, data, storage, etc., without the need to predefine and maintain the data that needs to be shared, so that the current information "share when needed" model becomes possible. Transform to the "know when you need" model, strengthen the degree of power informationization, and make the tight coupling of information sharing loosely coupled.
When connecting, initialize the information of the data server collection area, obtain the communication connection between measuring point information and equipment-related commands, and obtain the judgment of these measuring points and the judgment of the communication and command of the equipment. The communication information in various states of this site, all communication equipment and communication states, measurement points, and medical reform commands do not belong to the collection area.

Collect and monitor real-time data in each data collection area, and obtain the electricity consumption information collection system application [4]. The database server is collected, processed and displayed, and the information of the corresponding tree collection area is obtained by using the query tool and the information data collection function without additional manual operation.

3. Collaborative collection task scheduling framework generation

Constructing a wide-area computing task flow processing framework for multiple data centres. Among them, the wide-area level mainly solves the problem of task decomposition and planning and scheduling; the regional level mainly solves the task scheduling problem of the local data centre. From a logical point of view, the multi-data centre-oriented computing task flow processing framework is divided into task layer, planning layer, and scheduling layer. The overall framework is shown in Figure 2.
Figure 2. Overall framework

1) The task layer refers to computing tasks initiated by users of power big data analysis. The data involved can be stored in spatially distributed data centers or logically distributed physical servers. For the various typical application functions of the new power consumption information collection system for the energy Internet, the application functions under open conditions are packaged in a unified manner. After the service is established, it must be registered in the format specified by the service registration center. Then, when there is a need for application functions when it goes live, the application functions under the system open condition are encapsulated as general services by default and organized in the service center to make the deployment of application functions more efficient and reliable.

The planning layer refers to the decomposition of computing tasks initiated by users of power big data analysis, and the optimization of computing tasks and data centers through multi-objective optimization methods.

The scheduling layer refers to the coordinated scheduling of tasks in the local data centre. Through scheduling algorithms and strategies, the infrastructure resource scheduling and computing task scheduling of the local data centre are realized. According to the classification of user types, there should be a unified portal and a unified authentication mechanism for general users, and for secondary developers, a unified interface service definition format and clear and checkable method descriptions should be used. If it is distinguished according to the internal users and external content users of the network adjustment centre, the invocation service with different permissions is provided for users of different regions and different levels.
3.1. Hardware facilities

The distributed data acquisition system we are involved in has 3 data acquisition areas, each data acquisition area has 2 front-end servers, and the corresponding terminal servers and routers are respectively used for serial channel communication and network channel communication, as shown in Figure 3.

![Figure 3. Distributed data acquisition hardware structure of the integrated system](image)

Each data collection area still collects data from the factory station through the channel resources dispatched by the original centre, and the three data collection areas jointly complete the data collection tasks of all the factory stations.

3.2. Two-channel communication mode

The main station and the sub-station communicate through the MIS network of the provincial power system with socket software. In terms of data, the slave station acts as a server listening port, and the master station acts as a client to connect and request data. In terms of communication settings, the master station acts as a server and monitors the network settings reports of each sub-station. The communication protocol uses a custom application layer protocol based on TCP, and TCP provides connection guarantee and transmission control for the upper layer data transmission. The custom application layer protocol defines the data, events and other information that need to be transmitted according to the functional requirements of online monitoring, and considers the differences of different monitoring equipment manufacturers to provide certain customization and self-description capabilities [6]. The concept of measuring points is used for the measured data, which is directly mapped to the database field. Generally, TCP-based data acquisition communication uses the slave station as the server and the master station as the client. The master station controls the on and off of the TCP connection, and the master station is the active inquirer in the communication protocol. In order to ensure the correct communication and connection of the master station, each sub-station must use a public IP address and port, which cannot be changed at will. However, IP address resources in the power network are currently in short supply, and sub-station communication devices and on-site monitoring devices are installed in substations of various cities, which need to obey the unified arrangement of each power supply bureau, and even use dynamically allocated addresses and ports, which makes the master station unable to obtain Substation address.
3.3. **Business Process**

The key steps of the adaptive performance acquisition process based on intelligent analysis are as follows.

**Step 1:** Periodic collection of equipment performance, and periodic collection of transmission equipment performance data according to default rules;

**Step 2:** Analyse the equipment performance trend and equipment performance early warning, and find the performance fault early warning, then enter the processing link, according to the early warning algorithm, calculate the early warning indicators of the monitored performance data, and according to the importance, scope, urgency of the fault, etc. Pre-defined levels, divided into warning levels;

**Step 3:** Adjust the equipment performance collection tasks. Based on the results of the equipment performance trend analysis, generate new performance collection tasks according to the pre-established rules for different levels of early warning, including the adjustment of various parameters: performance collection objects, collection Interval, performance granularity, coverage of performance values, etc.

**Step 4:** Execute a new collection task, adjust the actual collection parameters based on the automatically generated performance collection task, and perform performance data collection according to the new rules; if the warning is eliminated, the performance collection of the default rules will be restored. Through the design of adaptive collection process, it can avoid the shortcomings of poor timeliness, time-consuming and labour-consuming methods of traditional manual acquisition and processing of performance data, and can also avoid the disadvantage of simple periodic automatic collection that brings greater load pressure to transmission equipment and network management. Strong innovation and practicality.

4. **Data collection task scheduling algorithm**

Task decomposition refers to the process of dividing a complex task into multiple independent subtasks based on a certain method or principle, and the input and output of each subtask have a certain logical dependence. Common task decomposition methods include: classification decomposition method, hierarchical specification decomposition method, load balance decomposition method, subgraph decomposition method with the least connection, etc. This paper adopts the hierarchical normative decomposition method to fully consider the inherent spatial rules, element rules, timing rules, and functional rules of the task to be decomposed to decompose the task, and make the task granularity reach the optimal [7]. Affecting the calculation efficiency, the optimal task decomposition method based on the level specification is shown in Figure 4.
The continuous-time Markov return judgment process model used in this paper mainly uses approximate methods to calculate the steady-state probability, so CSRL can be used to describe the nature of the model in the choice of sequential logic. The grammatical form of the CSRL state formula:

\[
\phi ::= \text{true} | \neg\phi | \phi \land \phi | S_{\omega}(\phi) | P_{\omega}(\phi)
\]

(1)

Among them, \( q \in AP \) is the atomic proposition, \( p \in [0,1] \) is the probability value, \( \Delta \) is the comparison operator, and \( \varphi \) is the path formula. \( S_{\omega}(\phi) \) is the steady-state probability operator, which means that the probability of the final set of paths in the \( \phi \) state satisfies \( G \), and is the instantaneous probability operator, which means that the probability of all the path sets satisfying the path formula \( \varphi \) satisfies the \( \Delta \varphi \) CSRL. The syntax form of the path formula \( \varphi \) is:

\[
\varphi ::= \text{O} | \text{U} | \varphi \land \varphi
\]

(2)

Among them, \( \land \) and \( \lor \) represent the next step operator and the until operator, respectively, and \( I = [r, r'] \subseteq R_{\omega} \cup \{\infty\} \) and \( J = [r, r'] \subseteq R_{\omega} \cup \{\infty\} \) represents the time interval and the return value interval, respectively. For the continuous Markov model \( M \), in the initial state \( E \), the transition probability of transition to state \( s' \) at time \( t \) is expressed as:

\[
\pi^M(s_0, s', t) = pr_{s_0} \{\sigma \in \text{path}^M | \sigma(\sigma_1) = s' \}
\]

(3)

When the time \( t \) tends to infinity, the transition probability to any state in the state set \( S' \) is expressed as:

\[
\pi^M(s_0, S') = \lim_{t \to \infty} pr_{s_0} \{\sigma \in \text{path}^M | \sigma(\sigma_1) \in S' \}
\]

(4)

Based on the above description, the satisfiability judgment of CSRL state formula \( S_{\omega}(\phi) \) can be transformed into calculation of \( \pi^M(s_0, \text{Sat}(\phi)) \), where \( \text{Sat}(\phi) \) represents the satisfactory state set of formula \( \phi \). For CTMRM, in order to facilitate the calculation of \( \pi^M(s_0, \text{Sat}(\phi)) \), the model can be discretized before the corresponding calculation. In CTMRM, if the state \( s' \) has only one output transition, the probability of the state transition in time \( t \) satisfies the exponential distribution, and the value is \( 1 - e^{-R(s',a,n,x,t)} \); if the state \( s' \) has two or more output transitions, then the state \( s' \) will transition to one of them. The transition probability of a successor state \( s^* \) is:

\[
\frac{R(s',a,n,x^*)}{E(s',a,n)} (1 - e^{-E(s',a,n,t)})
\]

(5)

Among them

\[
E(s',a,n) = R(s',a,n,S') = \sum_{s^*} R(s',a,n,s^*)
\]

(6)

5. Analysis of Collaborative Scheduling Strategy for Collection Tasks

Construct a wide-area computing task flow collaboration model, including three parts: computing task collaboration management, computing task execution, and computing task monitoring. Among them, the collaborative management of computing tasks is mainly responsible for the establishment of the collaborative relationship between the various sub-computing tasks in the set of computing task flows,
and the establishment of the dependency relationship between each sub-computing task and computing resources; the execution of the computing task is responsible for the generation of processing programs for specific computing tasks, Release, etc.; computing task monitoring is mainly responsible for monitoring the execution of the entire computing task flow.

According to the resource load of each storage power big data and its replica nodes and the characteristic parameters of network transmission, such as the number of overall data storage nodes, the number of data copies on different storage nodes, and the current access requests of each storage node Execution and resource utilization ratio, number of concurrent tasks, network bandwidth, network delay, communication quality and other conditions, under the conditions of optimal utilization of parameters, construct a multi-objective optimization model for wide-area computing task flow and data storage nodes.

The weighted sum method is used to decompose the multi-objective optimization model for wide-area computing task flow and data storage nodes into several single-objective optimization sub-problems for sub-computing tasks and data storage nodes. On this basis, the evolutionary algorithm is used to solve the single-objective optimization problem at the same time.

6. Optimize simulation analysis of data acquisition tasks

6.1. Test environment setting
Set up system A, system B, and system C respectively in the experimental environment, a total of 3 sets of simulated power grid control systems, each set of simulation system deploys 2 acquisition servers, and system A deploys 2 application servers [8]. The network environment is the internal network bandwidth of the system is 1000 Mbit/s, and the network bandwidth between the systems is 100 Mbit/s. The environment deployment is shown in Figure 5.

![Figure 5. Schematic diagram of test deployment](image)

6.2. Performance test
First perform a low-load test. Start the subscription interface in System A, and subscribe to the "System B. Test Station" data. The station's data volume is 100 telemetry and 100 remote signalling. System B accesses the station's data message through the simulation protocol program, and view the time scale recorded when the data in system A and B flows through each link. Then perform a high-
load test, continue to subscribe to multiple sites of System B, increase the total subscription data to more than 100,000 telemetry and 100,000 remote signalling data points, simulate access to all plant data messages and view "System B . The time scale when the data of the “test station” flows through each link. Finally, the transmission rate is calculated according to the above-mentioned time scale. The test result is shown in 1. The simulation diagram of data acquisition efficiency is shown in Figure 6.

Table 1. Wide-area data collection performance test

| Test items | Time-consuming for simulation system B/ms | Network transmission time/ms | Time-consuming for simulation system A/ms | Total/ms |
|------------|-------------------------------------------|----------------------------|-------------------------------------------|----------|
| Low load test 1 | 13                                        | 8                          | 10                                        | 31       |
| Low load test 2 | 15                                        | 10                         | 8                                         | 33       |
| Low load test 3 | 12                                        | 8                          | 7                                         | 27       |
| High load test 1 | 48                                        | 24                         | 30                                        | 102      |
| High load test 2 | 40                                        | 30                         | 33                                        | 103      |
| High load test 3 | 45                                        | 32                         | 31                                        | 108      |

Figure 6. Simulation graph of data collection efficiency

It can be seen from the test results that from the message received by the system B to the wide-area data message received by the system A, the total time consumption at low load is about 30ms on average, and the total time at high load is about 105ms on average. The data transmission time between systems is less than 1s. The above test results show that wide-area data collection can ensure efficient and real-time transmission of wide-area data.
7. Conclusion
This article provides a real-time data collection method based on equipment monitoring in the field of power grid dispatching. Through distributed task scheduling technology, real-time streaming data processing technology and massive time series database technology, real-time collection and processing of equipment monitoring data of the Energy Internet is realized. Warehousing has further improved the monitoring of equipment and grid operation, and improved the intelligent level of equipment operation management. The data collection design plan is implemented on the integrated online monitoring platform of power grid electrical equipment. The system runs relatively stable. Although there are sometimes substation failures, the main station collection platform can automatically supplement data after it resumes work. The system can accurately reflect the production and operation data in time and realize the centralized display of the equipment's one-time online monitoring data, which is conducive to timely discovering equipment defects and improving the level of equipment safe operation.

Acknowledgments
Project Code: 5600-201955457A-0-0-00 "Research on Key Technology of Electricity Information Acquisition for Energy Internet

References
[1] Abdelfadeel, K. Q., Zorbas, D., Cionca, V., & Pesch, D. $ FREE $—Fine-Grained Scheduling for Reliable and Energy-Efficient Data Collection in LoRaWAN. IEEE Internet of Things Journal, 7(1) (2019) 669-683.
[2] Li, K., Yuen, C., Kusy, B., Jurdak, R., Ignjatovic, A., Kanhere, S. S., & Jha, S. Fair scheduling for data collection in mobile sensor networks with energy harvesting. IEEE Transactions on Mobile Computing, 18(6) (2018) 1274-1287.
[3] Li, K., Ni, W., Tovar, E., & Jamalipour, A. On-board deep Q-network for UAV-assisted online power transfer and data collection. IEEE Transactions on Vehicular Technology, 68(12) (2019) 12215-12226.
[4] Kaveh, A., & Vazirinia, Y. Smart-home electrical energy scheduling system using multi-objective antlion optimizer and evidential reasoning. Scientia Iranica, 27(1) (2020) 177-201.
[5] Nguyen, N. T., Liu, B. H., Chu, S. I., & Weng, H. Z. Challenges, designs, and performances of a distributed algorithm for minimum-latency of data-aggregation in multi-channel WSNs. IEEE Transactions on Network and Service Management, 16(1) (2018) 192-205.
[6] Osamy, W., El-Sawy, A. A., & Khedr, A. M. Effective TDMA scheduling for tree-based data collection using genetic algorithm in wireless sensor networks. Peer-to-Peer Networking and Applications, 13(3) (2020) 796-815.
[7] Zeng, F., Zhang, R., Cheng, X., & Yang, L. Channel prediction based scheduling for data dissemination in VANETs. IEEE Communications Letters, 21(6) (2017) 1409-1412.
[8] Foresti, R., Rossi, S., Magnani, M., Bianco, C. G. L., & Delmonte, N. Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. Engineering, 6(7) (2020) 835-846.