

RESEARCH ON DATA EXCHANGE TECHNOLOGY BASED ON THE RAILWAY SIGNAL TEST

Qiao LIPING¹, Liu XIA², Chu JIANLI³

The railway industry has been developing rapidly in recent years, though, it has solved the huge demand of domestic transportation, and however, it has revealed the weakness of all aspects on the based management of the railway. Especially signal security is facing a huge test. Within this context, this paper presents the technology of interactive signal tests data processing based on railway network organization, through the study of the railway signal test business structure and signal data management process. This technology using RTADE (Real Time Automatic Data Exchange Algorithm) algorithm can report a large number of test data scattered at each station to the partners of the higher quickly and accurately, so the higher-level units can statistic and analyze the data. According to this study, they can grasp the current work of the station signal equipment accurately, and find the existence of security risks in time. Based on these, the traffic safety of railway transportation industry can be effectively guaranteed.

Keywords: Signal test; Data exchange; Data analysis; Railway Safety

1. Introduction

With the development of the railway transportation industry, the safety of railway signal is also facing a severe challenge. The railway equipment signal test is not only the guarantee of the running safety of high speed trains, but also an important precondition to improve the efficiency of railway transport. How to obtain the signals of each site test data quickly and accurately, then analyze the data deeply, and make the fault prediction is the urgent problem in the process of signal testing [1-3].

Electricity section is the basic unit of railway signal system; it is responsible for the management and maintenance of the train ground signal, locomotive signal and the normal operation of the railway [14-15]. Electricity section rules hundreds of stations. Each station must be tested according to the test cycle of the signal equipment, then, the test data will be reported to the higher authorities. Finally, the higher units do the overall statistical analysis of the test data, and deal with the hidden trouble of equipment failure and equipment safety.

¹ Department of information Engineering, Xingtai Polytechnic College, Xingtai, Hebei, China,
e-mail: qiaoliping1992@126.com, cheast@163.com

² Department of information Engineering, Xingtai Polytechnic College, Xingtai, Hebei, China

³ Department of information Engineering, Xingtai Polytechnic College, Xingtai, Hebei, China

So far, most of the current electricity sections have to use manual filling and artificial interaction to collect report and analyze the signal data. This model not only affects the speed of data processing, but also results in data loss and confusion.

Some electricity section has achieved preliminary information management for the signal tests service; there is no more effective management method for uploading and storing a large number of signal data between hundreds of stations [13]. To solve these problems, the author has conducted a lot of research for the organizational structure and signal test business process of the railway section, and proposed the method of the Interactive data processing, in order to solve the problem of railway information management, improve management efficiency, though all of the methods above to promote, develop and perfect the management of railway information [5]. Starting from the information management of the signal data, this paper wills deep the study of signal tests data interactive processing.

2. The information management of the signal data

Normally the railway electricity section rules a number of workshops, each workshop rules some work area which is composed of the station and section [4,6]. Therefore, the work area is usually used as the basic management unit of the signal test. The organizational structure of electricity section is showed in Fig. 1.

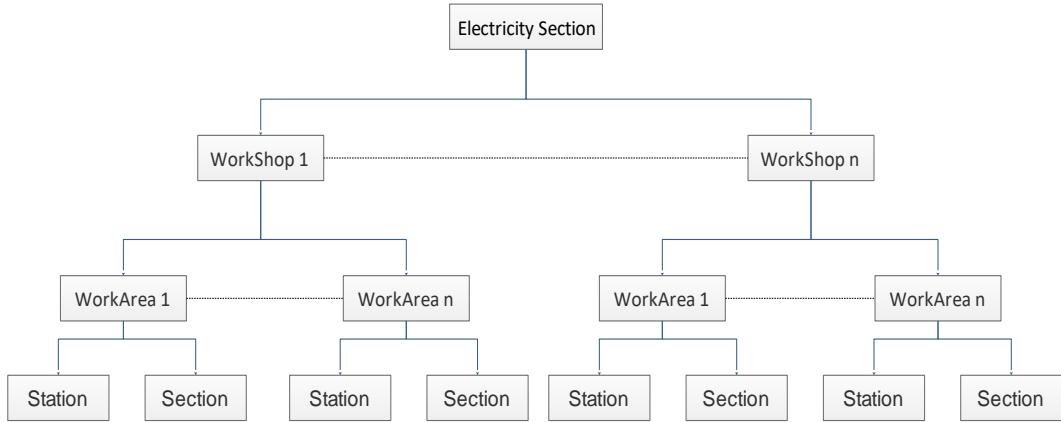


Fig.1. The organizational structure of all units in electricity section

When testing the signal of the electricity section, the testers of the work area firstly record signal test data of the station, and then input the signal data in the spreadsheet manually for the different signal equipment. In addition, calculate and save the spreadsheet data manually, afterwards, the data will be uploaded to the higher-level units step by step. Lastly, the electricity sections analyze and statistic the signal data of all the stations below. The entire process is complicated and tedious, so the way of data filling manually and artificial calculation, will directly

affect the level of mastery of the operation about signal equipment in time by the superior management unit.

So, before realizing the data interaction processing we must carry on the information management of the signal data. The design flow of information management is illustrated in Fig. 2.

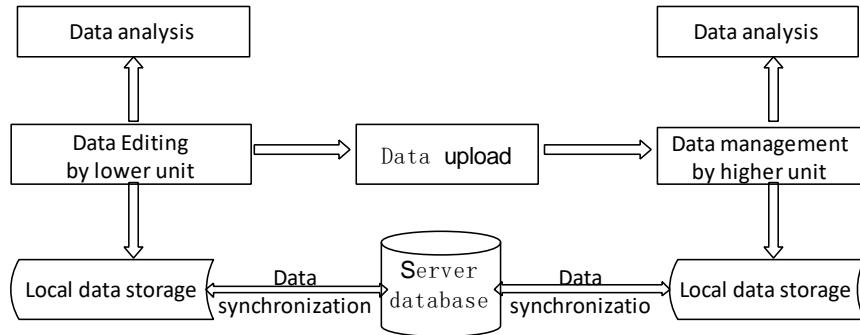


Fig. 2. The design of information management flow

From the above figure, we know that completing the overall signal tests information management requires three steps: Firstly, we provide the information system for the data editing of the work areas which are the lower unit. The testers can input data quickly into this system, and store the data locally, so that we can analyze and review the local data [16]. Secondly, the data can be automatically uploaded to the higher-level units by the work area, after the completion of data acquisition. The higher-level units, such as the work shop and the electricity section, can grasp the signal data onto the work area in time, meanwhile, the data will be stored locally. Thirdly, we establish the server database by this system which can store the reported data from all the work areas and workshops, and synchronize the local data from all levels of units. Lastly, we can achieve the data sharing services among the stations.

3. The design of the signal data flow of the information management

In the railway signal testing service, the signal data is a kind of scattered data. In order to determine the probability of different kinds of signal data exchanges and the order for the data exchanges list, we use the way of Information entropy to calculate the above result. Information entropy is used to represent the average information provided by a signal source. The lower the value of information entropy, the stronger the system order is [17]. We design the following formula as the entropy calculation model:

$$H(X) = -E[\log_2 p(x_i)] = -\sum_{x_i \in X} p(x_i) \log_2 p(x_i) \quad (i = 1, 2, \dots, n) \quad (1)$$

In the formula, x represents the signal source, x_i indicates the possible value of the random variable of the signal source x , $p(x_i)$ is the probability value of the x_i . It can be known that the value of the information entropy is always positive, from its calculation model. It means that the signal source is always able to provide a certain amount of information, in any case. We use the discrete random variable ξ to represent various values of the signal source x , and assume that its distribution follows the matrix:

$$\begin{bmatrix} x_1, x_2 \dots x_n \\ p_1, p_2 \dots p_n \end{bmatrix} \quad (2)$$

In this premise, we order $x_i = \log_2 1/p(x_i)$, and then, we get the result:

$$H(\xi) = E(\xi) \quad (3)$$

If it is the random independent distribution sequence obeyed the probability distribution of the formula 2.

And when n tends to infinity, the value $\frac{1}{n} \log_2 p(\xi_1, \xi_2, \dots, \xi_n)$ will tend to be $H(\xi)$, it is the follow formula:

$$\frac{1}{n} \log_2 p(\xi_1, \xi_2, \dots, \xi_n) \xrightarrow{p} H(\xi) \quad (4)$$

On the basis of the information entropy calculation model, we can design the signal data flow management to realize the interactive processing technique of signal data. As seen above, we know that the organization structure of the railway signal test includes the work area, the workshop and the electricity section. The signal data flow of each unit is showed in Fig. 3.

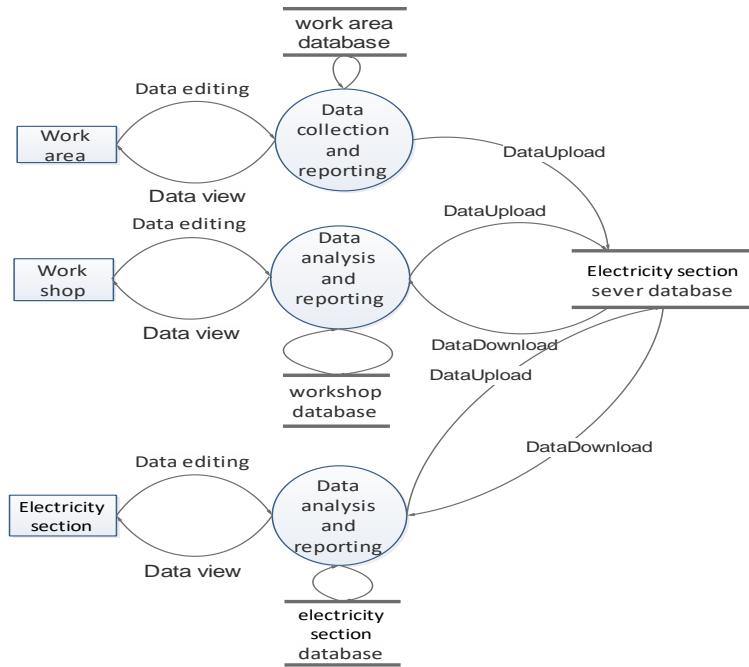


Fig. 3. The flow of data exchange among all units

In the data process, the signal data is acquitted from the work area, and then uploaded to the server database [16].

The workshops firstly receive the data onto the work area belonged to this workshop, and then execute the data view and the data analysis, finally upload results to the server database.

The electricity section can obtain the whole data from the server database, on this basis; it can view and analyze the data, and upload the analysis results to the server database. So, the server database is the core of the data interaction of units at all levels [7-9]. Units at all level can view and edit the unit data according to different permissions. In order to upload and download data onto the server database, the local data storage is retained at all units.

4. Signal data exchange processing technology

From the above analysis, we know that the railway signal test data need to be reported, analyzed and downloaded between different units. For example, the equipment failure data of the work area needed to be report onto the workshop for approval, and then reported to the electricity section by the workshop [12-13]. When the electricity section receives the equipment failure data, it will inform the technical personnel to deal with the fault. We know that there is a large amount of data interaction between the units at all levels, in order to improve the efficiency of

data interaction, we not only need to design a reasonable structure of the signal data, but also need to design and improve the data interaction process.

4.1 The design of the Signal data structure

Signal data structure can not only save real-time data, record the state of the data, but also can identify the data flow. In order to keep the uniqueness, integrity and real time of the data during the data interaction, we design the basic structure of the signal data which is shown in Fig. 4.

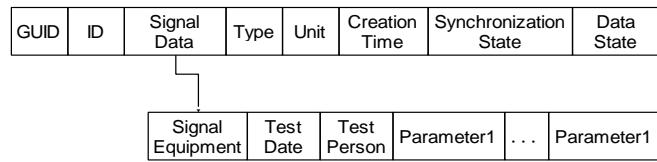


Fig. 4. The signal data structure

From the figure above, we know the GUID can guarantee the uniqueness of data synchronization between multiple databases, the ID represents the serial number of the local data, the Signal Data is the real-time data of signal test, and the Synchronization State and the Date State can record the process of data interaction between units at all levels.

4.2 RTADE algorithm

RTADE algorithm is the Real Time Automatic Data Exchange Algorithm, the algorithm design of the data exchange process as shown in Fig. 5:

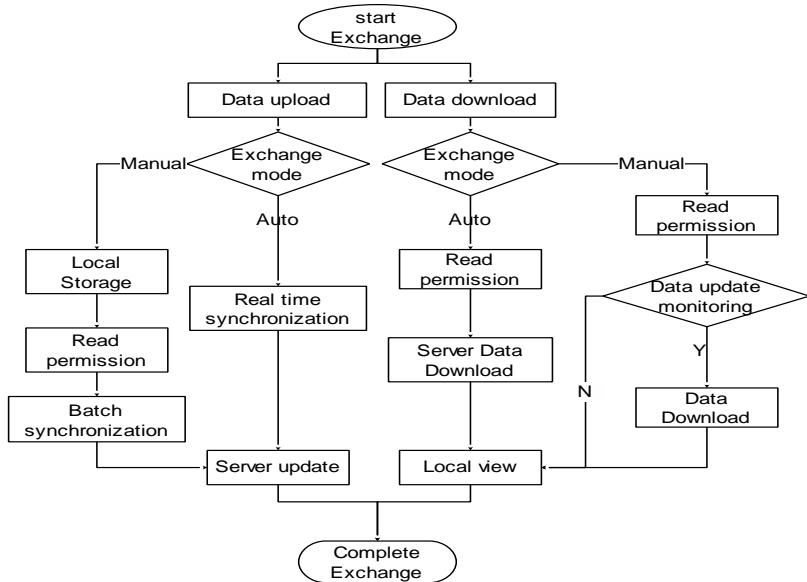


Fig. 5. The design of the Data Exchange

From Fig. 5, we know that data upload contains two kinds of exchange mode: Manual and Automatic. If what we need to exchanged is the real-time data and small batch data, we will choose the Automatic mode. It will update the local data and remote server data after the data editing; but if what we need to exchanged is the delayed data and big data, we will choose the Manual mode. In this way, we store the data onto the local, and set the synchronization status of the data to be not synchronized, after the completion of editing, we update the server data by using batch upload method.

When downloading data, we choose different exchange mode, according to the characteristics of the data. If the interactive mode is Automatic, the data will be downloaded directly from the server under different conditions, and viewed locally; Otherwise if the interactive mode is manual, we need to compare the difference between the server and the local data, if the server data has been updated, it will be downloaded to local, and viewed locally. But if the server data is the same as the local data, we will directly open the local data to view. In accordance with the above data exchange process, the specific algorithm steps are as follows:

Step 1 : keyword sort

Before the data exchange, we will sort the data in accordance with the keyword, the keyword can be data storage time, the sorting algorithm is as follows:

```
List<DateTime> itlist={2016-1-2:8:20,...}  
// Defining the storage time of the uploaded data  
for (int i=0;i<itlist.count;i++)  
// Multiple traversal for Data List  
{  
    DateTime max=itlist[i]; // Latest storage time  
    foreach(DateTime dt in itlist) // One by one comparison  
    {  
        if(dt>max) // Executing data exchange  
        {  
            value=dt;  
            dt=max;  
            max=value;  
        }  
    }  
}
```

Step 2: GUID code upload

On the basis of the first step, we can encode and upload sorted data, because the data from more than one area has the probability of the same value, so, when uploading data, we encode the multiple area data by the GUID way, encoding algorithm is as follows:

```

struct UploadedData // Defining the Upload data structure
{
    GUID dataId; //GUID Code
    String syncState; // Data synchronization status
    DateTime insertedTime; // Data storage time
    String unitName; // organization name
    InfoData value; // Signal data packet
    string dataEditState; // Data editing status
    string type; // data type
}
Switch (syncState)
{
case "Synchronized":
    RemoveList (GUID); // Remove sync List
case "Not synchronized":
    AddList (GUID); // Add sync List
Case "To be synchronized"
    SyncData (GUID); //Executing data synchronization
}

```

Step 3: The method of the data exchange optimal calculation

The implementation of data interaction requires consideration of several factors, such as: the big data, frequent interaction, complex network, redundant data, so we establish the data interaction mathematical model, which can calculate the minimum amount of data, during the interaction. The model includes two parts, the objective function and constraint conditions.

The objective function:

$$\min W = \sum_{i=1}^t \sum_{j=1}^c R_{ij} Q_{ij} + \left(\sum_{j=1}^c X_j D_j - \sum (D_j + E_j) E_j \right) V$$

Constraint conditions:

1. $\sum_{j=1}^t X_j \leq T$
2. $t \geq 2, c \geq 1, c \leq t$
3. $X_j = 1, 0; j = 1, 2, \dots, c$
4. $Q_{ij} \geq 1$
5. $D_j \geq 0, D_j < E_j$
6. $V > 0$

The symbolic meanings of the above functions and conditions are as follows: the symbol "W" is the total data for the exchange; the symbol "t" is the number of all the exchange centers; the symbol "C" is the number of centers needed for this data exchange; the symbol "Rij" is the unit data size from the exchange center i to the exchange center j; the symbol "Qij" is the number of all the data from the exchange center i to the exchange center j; the symbol "Xj" is used to mark

whether the exchange center j is needed for this data exchange process, if needed, the X_j will be 1, otherwise, it will be 0; the symbol “ D_j ” is the delay time for completing the fixed amount of data exchange conversion at the exchange center j ; the symbol “ E_j ” is the exchange time of the unit data for the exchange center j ; the symbol “ V ” is the average speed of railway network bandwidth. This mathematical model can choose the best path for data conversion, and calculate the minimum amount of interactive data, so that, it will be easy to predict the time limit of data exchange at the basis of the mathematical model.

We can complete a large number of data interactions from all the units belonged to the electricity section more effectively and accurately according to the design method above, which is on the basis of information management of signal data.

4. 3 The design of the data update mode

The Core operation of realizing data interaction is data update. During the interaction process, the data will be modified and deleted according to the actual situation by the units that have operated authority, and the new data will be generated by editing and computing [10-11]. The data has often been invalid after the completion of the data interaction above, even the result of non-validity, such as the data loss and data duplication will happen. Therefore, the design of a reasonable data update mode directly affects the accuracy and real time of data, during the whole data interaction process. Fig. 6 is the design flow of data update.

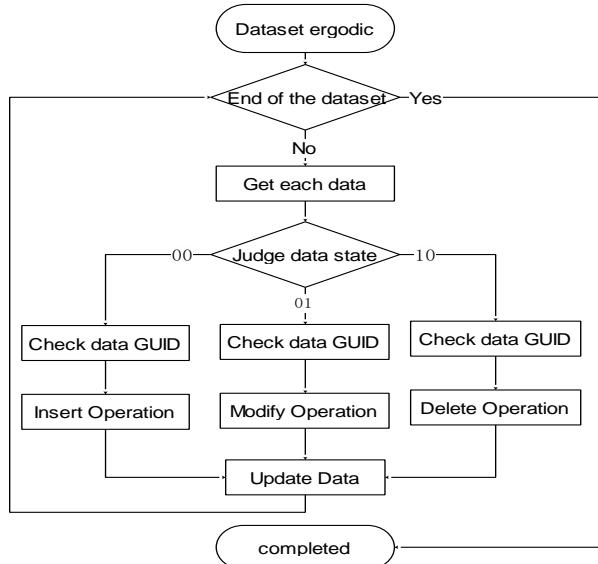


Fig. 6. The design of Data updates process

The figure above shows the whole process needs to traverse data collection, and update data one by one. In order to perform data updated operations accurately,

we add a data state represented in binary in the signal data structure, where the "00" means the data has been inserted, the "01" means the data has been modified, and the "10" means the data has been deleted in a unit which has operated a permission. When updating data, the algorithm will read the data state, and perform different operations in the database needed to be updated and finally complete the process of updating the entire data collection.

The advantage of data updated based on the data state is that the data state can record the operation performed by the updated data. So the system does not need to compare the differences between the two sides of the data collection, just re-executing the operation according to the data state which can ensure the data synchronization between the two sides.

4.4 The recovery processing of Data synchronization breakpoint

The main task of data interaction is to keep the data synchronization between the two sides, so the interaction process is based on the network to transmit data. Railway data transmission uses railway network, the network does not normally happen abnormal phenomenon. But during the data synchronization, the network is abnormal, the system must have the capacity of solving the problem in time, and record which data has been synchronized successfully, and which data has not been synchronized. After the network fault repaired, the system can continue to carry out data synchronization. This is the recovery processing technology of Data synchronization breakpoint, as showed in Fig. 7

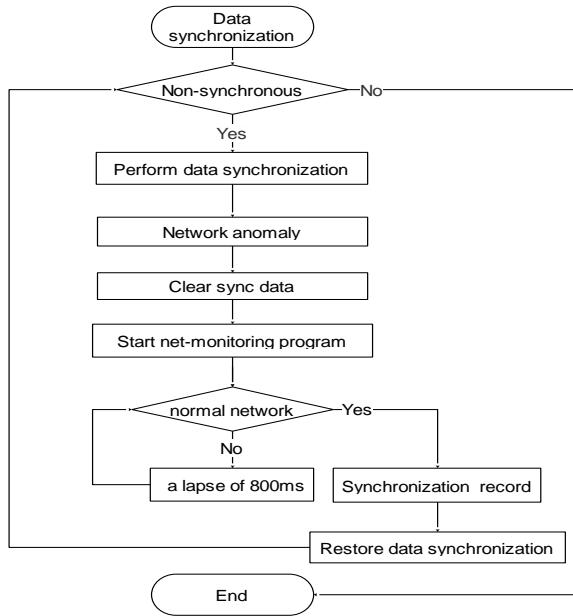


Fig. 7. The recovery technology of Data synchronization

The initial operation of the data synchronization is to detect whether there is non- synchronization data according to the synchronization state of the data structure. If there is no synchronization data, it will trigger the data synchronization operation, but if the network exception occurs during the synchronization, the algorithm will clear the synchronized data from the synchronization data list, and start up network monitoring program. This program which is background service won't affect the system operation by users, it monitors the state of the network every 800ms. If the network exception has been repaired, the algorithm will recover the data synchronization operation according to the synchronization schedule before the broken network, until the entire data synchronization process is completed. This technology can ensure data synchronization which will automatically recover from where it has been interrupted, so that users don't have to worry about the problem of data loss or duplication caused by the network exception or system error.

The data interactive technology is based on the information management of railway signal test. We firstly designed a reasonable data storage structure, which can not only save the parameters of the data but also record states of the data synchronization and the operation. In order to keep the data synchronization, we design multi interactive data updated mode according to ideas of the data process. Finally, we implement the breakpoint recovery processing technology for data synchronization exception to ensure the integrity and accuracy of the data.

4.5 Experimental analysis of data exchange technology

Data exchange is more effective to ensure the interaction of signal data through the use of RTADE algorithm and the recovery processing technology of Data synchronization breakpoint. This experiment includes two sets of experimental data before and after using the RTADE algorithm. The experimental objects include: the same amount of data upload time, transmission delay, resolution time and storage time. Table 1 is the experimental data before using RTADE algorithm, and table 2 is the experimental data after using RTADE algorithm.

Table 1

The experimental data before using RTADE algorithm

Data size(w)	data upload time(s)	transmission delay (s)	resolution time(s)	storage time(s)
1.2	5.46	0.13	2.41	1.57
4.6	14.47	0.29	5.36	4.79
7.3	25.13	0.56	9.82	7.52
8.7	31.52	0.71	12.13	9.32
10.4	37.24	0.85	14.64	10.83
14.3	49.76	1.09	19.58	15.02

Table 2

The experimental data after using RTADE algorithm

Data size(w)	data upload time(s)	transmission delay(s)	resolution time(s)	storage time(s)
1.2	4.12	0.08	2.14	1.41
4.6	12.26	0.21	4.81	4.52
7.3	21.37	0.43	8.56	6.83
8.7	25.49	0.52	10.63	8.19
10.4	30.01	0.61	12.84	9.86
14.3	42.46	0.84	21.27	13.75

It can be seen from the data, after using RTADE algorithm and Data interactive optimal calculation method, the upload time and the transmission delay of the same data decreased by 24.54% and 38.46% respectively. According to the design of the above signal data structure, the data resolution time of each platform is reduced by 11.2%, the time of storing data onto the server database is reduced by 10.19%, during the process of data exchange. In order to describe the differences clearly in the above time, we make a column chart and a line chart, for the above three sets of data before and after the algorithm, as shown in Fig. 8 and Fig. 9. In the horizontal axis of the graph, the symbol whose suffix is A represents the data after using RTADE algorithm, and the symbol whose suffix is B represents the data before using RTADE algorithm.

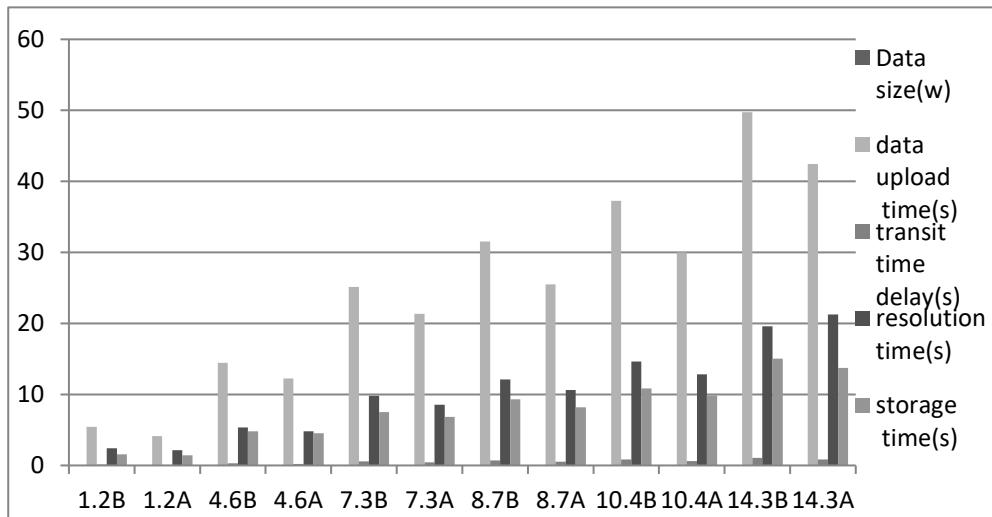


Fig. 8. Data analysis column chart before and after using the algorithm

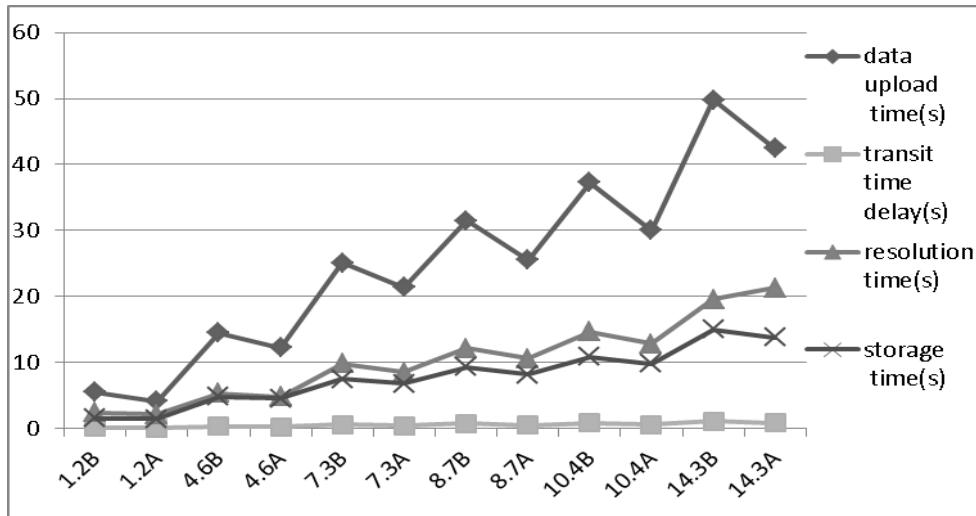


Fig. 9. Data analysis line chart before and after using the algorithm

5. Conclusion

The data exchange technology uses the automatic batch processing mode for the two sides of the data synchronization. This method does not need to care about the data difference between the two sides, but directly use different update modes to complete the data synchronization according to the data synchronization status. Especially when the network exception occurs, it will start up the breakpoint recovery processing function and the real-time monitoring network which can be enabled too. Once the network exception has been repaired, the system will continue to carry out data synchronization, starting from the last synchronized data breakpoints. So, this technology is more effective to reduce the data error, greatly improve the efficiency of data synchronization, and provide a strong guarantee for the signal data test and analysis.

This paper applied the data interaction technology to the railway signal information management system, and carried out the function test and the performance test, by using the railway signal simulation data. The experimental results show that the technology can effectively solve the problem of railway signal data industry. Especially in the complex case of batch data and multi-unit interaction, the performance advantage of the technology is more prominent. This technology not only solves the problem of the operation complexity and time lag of the manual filling and manual reporting methods for the railway in the past, but also provides a guarantee for the integrity and accuracy of the signal data. The rationality of signal data structure design, not only can accurately obtain the data, but also can track the source of data and changes of the data flow state. It is fully prepared for the railway follow-up data analysis and data feedback.

R E F E R E N C E

- [1] Hwang, Jong-Gyu; Baek, Jong-Hyun, MSC-based Test-case Generation Module for Railway Signaling Software Testing. The Transactions of the Korean Institute of Electrical Engineers (2015), Vol. 64, pp.138-142
- [2] Ilhan Mutlu*, Design of a Hardware and Software based Test Bed for Railway Signalization Systems, IFAC Proceedings Volumes, (2012), Vol.45, No. 24, pp.185-190
- [3] Marek Pawlik, Control Command Systems IMPACT on The Railway Operational Safety, Nauka ta Progres Transportu, (2015), Vol. 56, pp.58-64
- [4] Dincel, E.;Eris, O.;Kurtulan, S. Automata-Based Railway Signaling and Interlocking System Design [Testing Ourselves], IEEE Antennas and Propagation Magazine, (2013), Vol. 55, No. 4, pp. 308-319
- [5] Dain, Stephen J.; Casolin, Armand; Long, Jennifer; Hilmi, Mohd Radzi Color Vision and the Railways: Part 1. The Railway LED Lantern Test, (2015), Vol.92, No.2, pp.138-146
- [6] Lopez-Gordo, M.1; Padilla, P.1; Pelayo Valle, F.2, A proposal for bio-synchronized transmission of EEG/ERP data, EURASIP Journal on Wireless Communications and working, (2016), Vol. 2016, No. 1, pp. 308-319
- [7] DATE, SACHIN, Should You Upload or Ship Big Data to the Cloud? ; Communications of the ACM,(2016), Vol.59,NO.7,pp. 44-51
- [8] Muhsin Shodiq;Rini Wongso;Rendy Setya Pratama;Eko Rhenardo;Kevin, Implementation of Data Synchronization with Data Marker Using Web Service Data, Procedia Computer Science, (2015), Vol. 59, pp.366-372
- [9] Milosavljevic, Nebojsa;Pawar, Sameer;El Rouayheb, Salim;Gastpar, Michael;Ramchandran, Kannan, Efficient Algorithms for the Data Exchange Problem, Information Theory, IEEE Transactions on Information Theory, (2016), Vol.62, pp.1878-1896
- [10] Sekhavat, Yoones A.1; Parsons, Jeffrey2, SEDEX: Scalable Entity Preserving Data Exchange, IEEE Transactions on Knowledge & Data Engineering, (2016), Vol. 28, No. 7, pp. 1878-1890
- [11] Donovan, Kevin, Exchanging translated HTML data (almost) painlessly. MultiLingual, (2016), Vol. 27, No. 3, pp.56-61
- [12] Hwang, Jong-Gyu;Jo, Hyun-Jeong;Baek, Jong-Hyun, Interoperability Test and Testing Tool for Railway Signaling System, Journal of the Korea Academia-Industrial cooperation Society,(2015), Vol. 16, No. 9, pp. 5802-5809
- [13] Jacek Mocki; Ljubo Vlacic, Railway interlocking process - formal method for documenting railway junction interlocking and signalling processes, International Journal of Automation and Logistics,(2015), Vol. 1, No. 2, pp.176-208
- [14] Li Yan; Xuming Fang; Yuguang Fang, Control and data signaling decoupled architecture for railway wireless networks, IEEE Wireless Communications Magazine, (2015),Vol. 22, No. 1, pp. 103-111
- [15] Pushpa Sharma;G. Geetha, Data Obtainer: an application for digital acquisition , International Journal of Wireless and Mobile Computing,(2016), Vol. 11, No. 1, pp.1-9
- [16] Elena BALMUS, Aircraft Data Acquisition, INCAS Bulletin, (2016), Vol.8, No. 1, pp. 141-151
- [17] P. G. Sankaran; S. M. Sunoj, Quantile-based cumulative entropies, Communications in Statistics - Theory and Methods, (2017), Vol. 46, No. 2, pp. 805-814