Application of Block Chain Technology in Distributed Power Transaction and Data Governance

Yujiang Long[1*], Xun Li[1], Wei Wei[1], Na Long[1], Rundong Gan[1], Ye Zhong[1]

[*] Corresponding author: longyujiang8876232@163.com

Information Center of Guizhou Power Grid Company Limited, Guiyang, China[1]

Abstract—Aiming at the problems of high transaction risk and insufficient data security governance in the practice of power data sharing, this paper explores the application method of block chain technology in distributed power transaction and data governance, collects and identifies the characteristics of distributed power transaction data, explores the consistency of distributed power data, and realizes data tracing and sharing encryption, block chain network and cryptography technology are used to encrypt distributed power transaction data. Finally, experiments show that the proposed application method of block chain technology in distributed power transaction and data governance can effectively improve the security of data governance and ensure the security of power transaction, which is of positive significance for power enterprises to grasp the opportunity of digital transformation and realize the goal of production, operation and service upgrading of traditional energy enterprises.

Keywords-Block chain; Distributed power; Electricity trading; Data governance

1. INTRODUCTION

Data governance is a series of work around data assets, aiming at serving the decision-making of all levels of the organization. It is a collection of data management technologies, processes, standards and policies. Improve data quality, consistency, availability, availability and security through the data governance process, and finally enable enterprises to manage and apply data as core assets[1]. Using its own characteristics, block chain technology can solve the problems of data governance results and application reliability, establish the traceability mechanism of wrong data, provide a consistent, credible and authoritative data platform for an information management system and business system, and meet the requirements of data distributed storage, high reliability and reliability of data application[2]. The general data governance platform with the characteristics of data security and traceability meets the data management requirements of the whole network consistency, credibility and authority of IT asset data. Power data also gradually plays a supporting role in decision-making in the monitoring and analysis of enterprise resumption of production and work, business climate index analysis, community personnel flow and prevention and control[3]. However, there are still many problems in the

ICEMME 2022, November 18-20, Nanjing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.18-11-2022.2327187

application of power data sharing, such as security and trust, security mechanism, inconsistent data, poor data traceability, uncertain sharing mode and so on. Therefore, according to the advantages or characteristics of block chain technology such as decentralization, transparency, smart contract, traceability and high data security, based on the research, exploration and practice of domestic and foreign scholars in the fields of contract management, data transaction and e-government, a power data sharing model based on block chain is constructed. Some application scenarios of power data sharing under the block chain are listed, which provides a new implementation path to further solve the problems of trust Island, data ownership, standard consistency control and so on.

2. APPLICATION OF BLOCK CHAIN TECHNOLOGY IN DISTRIBUTED POWER TRANSACTION AND DATA GOVERNANCE

2.1 Distributed power data identification management model

Data governance involves the comprehensive application of management methods and technical tools. A complete data governance project usually includes five key elements: objectives, organizations, systems, tools and standards[4]. At the same time, outside the data governance project, it is necessary to improve the awareness of data management of relevant personnel through effective training means, and cooperate with the construction of standardized management system and corporate culture to provide guarantee for the implementation of data governance project. Power data sharing is of great significance to improve the self-management of power enterprises, optimize customer service, reduce unit energy consumption, and promote the intelligent decision-making of the government[5]. At present, electric power enterprises actively respond to the national digital transformation strategy and explore in breaking data barriers and building a data center[6]. In order to promote the sharing application of power data across units and departments, a data resource sharing model based on block chain is built, as shown in Figure 1.

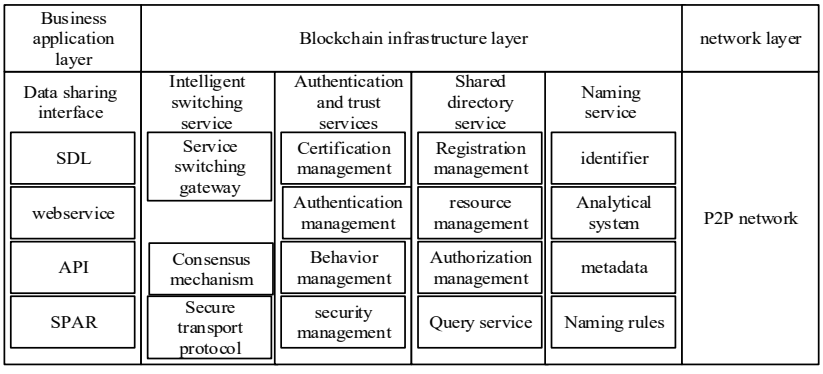

Figure 1 Power data sharing model architecture based on block chain

Block chain model mainly has two roles: data provider / exporter and data receiver / user. Government departments such as finance, taxation and public security, as well as enterprises in

the fields of finance, medical treatment and energy, are applicable to this model as block chain users. Block chain infrastructure layer: it is the core layer in the whole model and provides basic service capabilities for mutual trust, mutual recognition and interoperability in power data sharing.

2.2 Data transaction security evaluation algorithm based on block chain

The technical advantages of block chain technology make the information reach a consensus state. For block chain, it can be regarded as a complete information body, and the processed data is saved according to the processing time to become a block. Through the hash value, each block is effectively linked together to form a block chain, as shown in Figure 2.

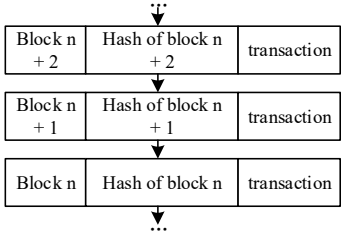

Figure 2 Chain structure of block

Block chain is a system born on the basis of the formation and existence of information transaction. Therefore, its system contains a large amount of data information about transaction. Huge data is displayed in the form of Merkle tree and stored in a tree structure, Comparison of typical block chain storage structures as table 1.

TABLE 1 COMPARISON OF TYPICAL BLOCK CHAIN STORAGE STRUCTURES

	Ethereum	Bitcoin	Super ledger
Root hash structure comparison	Merkle Patircia	Merkle	Merkle
Compared with the number of hashes	6	3	3
Comparison of data storage codes	RLP coding	Base60check code	Jason code

Although the business application layer can directly use the address information of the block, these address information data are long and the coding is complex, which is difficult for users to use in the actual data sharing process. Therefore, referring to the domain name resolution server in the network model, a block naming service system is built at the bottom of the infrastructure of the block chain. This paper sets the node that obtains all the demand data first in the whole network as the node that establishes the new block. Power CPS data prediction refers to the use of historical electrical quantity data and meteorological data such as temperature and season to predict the power data value in the future time period. Given the set

$(X_1, Y_1), (X_2, Y_2), ..., (X_k, Y_k)$ containing k training data, the prediction model based on limit learning machine is shown in the formula.

$$f = \min_b d + \frac{1}{k(X_k, Y_k)} b^2 \quad (1)$$

$$Y - Hb = d \quad (2)$$

Where, H is the hidden layer output matrix in the model; b Is the output weight matrix of hidden layer nodes and output neurons; Y is the output value matrix; d is the training error; κ is the adjustment coefficient to weigh the training error and output weight. The node forwards data L_s packets in a first in first out manner. The forwarding rate is m, which obeys the exponential distribution. δ is the service strength of the node. If the node cache capacity is configured as m, the node's processing process of packet data is set as k hybrid queuing model. The corresponding network performance parameters are obtained, including packet loss rate λ and delay time t in the node, as shown in the formula.

$$P_{loss} = \frac{1-\rho}{1-\rho^{m+1}} \delta^m \quad (3)$$

$$T_{delay} = \frac{L_s}{\overline{\rho}} = \frac{\frac{\rho}{1-\delta} - \frac{(m+1)\rho^{m+1}}{1-\rho^{m+1}}}{\lambda(1-P_{loss})} \quad (4)$$

Where L and P are the average waiting queue length and average service intensity in the node respectively. The time model of transmitting data from node to node not only considers the queuing of packet data at the node, but also considers the packet loss and retransmission mechanism, in which T forwarding nodes pass from the data transmitting node t_{waity}^{m} to the data receiving node D. By introducing the concepts of data transmission delay and network load balancing degree, and taking the minimum load balancing degree and transmission delay as the goal, the mathematical model of data demand driven transmission is established to meet the delay demand of data transmission according to power CPS business in the whole network, achieve load balancing, reduce the possibility of network congestion and improve the generation speed of data blocks. Calculation formula of P2P network load balancing degree.

$$l_{LoudBp2p?}=\sqrt{\left|\frac{\left(R_{Bi}-\overline{R}_{\mathrm{B}}\right)^{2}}{n_{1}}\right|} \tag{5}$$

Where R_{Bi} is the remaining bandwidth on the path from P2P source node to other destination nodes; $\overline{R}_{\mathrm{B}}$ is the average of the remaining bandwidth on all links; n is the number of all links used. The constraints from top to bottom are that the transmission delay found for different service demand strategies is not greater than the service demand time constraint, the remaining bandwidth constraint of the transmission link, the maximum upper bound of the number of hops of the transmission path, and the conservation constraint of the amount of data forwarded by the node.

2.3 Implementation of power transaction data management

When using k-means clustering algorithm, we need to pay attention to two main problems: the first is the calculation of the spatial distance between the cluster center and the nearest points in the target set. The calculation of spatial distance is usually selected according to the different types of data points. Generally, the European spatial distance calculation formula is selected as follows.

$$\mathrm{d}=\sqrt{l_{LoudBp2p}(X-X_{n})^{2}} \tag{6}$$

Similar to k-mcans algorithm, X_{n} is also a typical partition based clustering algorithm. The basic workflow of k-medoids clustering algorithm is to first randomly select K clusters in the predicted data set, in which the cluster center should be the real sample of the data set rather than the mean point, then select a cluster center point within the cluster center to calculate its consumption cost e, and compare the consumption cost e of cluster center point x_{j}, If o_{i} is less than k, assign the cluster center point to the non strip center point and continue to select the generation until the center does not change, it indicates that the clustering result has reached the best and the function converges. The clustering process ends, and the clustering consumption cost function is as shown in the formula

$$\mathrm{E}=\sum_{i=1}^{k}\left|x_{j}-o_{i}\right|^{2} \tag{7}$$

According to the knowledge of probability and statistics, if the samples are independent of each other, its classification accuracy and efficiency will be very high. The main idea of naïve Bayesian algorithm is closely related to probability. If $P(\mathrm{A})$ is used to represent the probability of event a, then $P(\mathrm{AB})$ represents the probability of event a when event $P(\mathrm{B})$ has occurred. Its calculation formula is as follows:

$$\mathrm{P}(\mathrm{A}|\ \mathrm{B}) = \frac{P(\mathrm{AB})}{P(\mathrm{B})} = \frac{P(\mathrm{A}) * \mathrm{P}(\mathrm{B}|\ \mathrm{A})}{P(\mathrm{B})} \tag{8}$$

The main idea of the general data classification algorithm is to know the training samples and the samples to be classified, and judge which kind of training samples the samples to be classified belong to after training. If we consider it from the perspective of naïve Bayesian algorithm, the result can also be transformed into $\left(\mathrm{X}|\ Y_i\right)$ which probability of the samples to be classified is greater in the training samples. By converting the above probability formula, the calculation formula of data classification can be obtained as follows:

$$\mathrm{P}\left(Y_i|\ \mathrm{X}\right) = \frac{P\left(Y_i\right) * \mathrm{P}\left(\mathrm{X}|\ Y_i\right)}{\sum_{i=1}^{k} P\left(Y_i\right) * \mathrm{P}\left(\mathrm{X}|\ Y_i\right)} = \frac{P\left(Y_i\right) * \mathrm{P}\left(\mathrm{X}|\ Y_i\right)}{P(\mathrm{X})} \tag{9}$$

Not all data resources are data assets. The process of realizing data resources is called data resource capitalization. Through the analysis of this type of data resources, it can optimize production and management methods and indirectly improve the income of existing products. This is an indirect way to realize data resources.

3. ANALYSIS OF EXPERIMENTAL RESULTS

The development environment for local testing is: operating system mac0s10.14.6, six core cpu3.0ghz, ram16gb. The single machine is configured with multiple block chain network nodes, one sorting node orderer, and two organizations org0 and org1. Each organization has two nodes peer0 and Perl. The power data storage scheme is verified by simulating nodes on the virtual machine. The virtual machine version is Ubuntu 16.04. To start the nodes in the fabric network, first install the tools and environment for building the network: docker, docker compose and go language environment, then pull the relevant docker image, and then start each node by configuring the docker compose file. The tools for building the experimental environment are shown in Table 2.

TABLE 2 DESCRIPTION OF TEST ENVIRONMENT

Test environment or tool	Configuration
Operating system	Win7
Computer model	Lenovo g480
Processor	Intel Core 2 dual core p7450
Memory	16G
Hard disk	500g hard disk

Block chain network	Ethereum
Smart contract test	Remix
Ethereum client	Go Ethereum

Test whether the block chain network can run successfully, that is, the availability of basic block chain functions (starting nodes, generating blocks, establishing channels, installing chain codes, etc.). Then, limit learning machine (ELM) prediction model and support vector machine (SVM) prediction model are used to predict the whole day power load data, and the power load prediction results are shown in the figure 3.

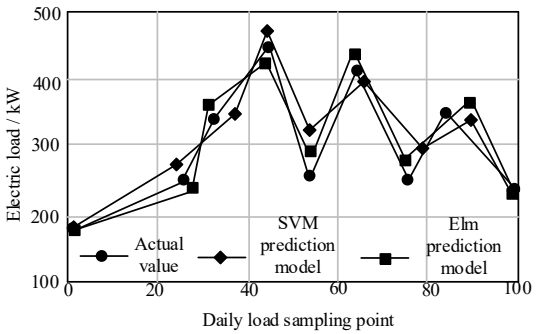

Figure 3 Load forecasting results

According to the prediction error result, the error threshold of the actual sampling value and the prediction data is set σ is 8.28, and then compare and analyze the threshold of the measured data collected at the next time, and judge whether the data is abnormal. Under the five scale power grid scenarios, the imbalance is reduced by 34.12% compared with the traditional algorithm and 23.82% compared with the infectious algorithm. The balanced use of network resources will reduce the probability of network congestion and packet loss under heavy load, and improve the overall performance of the network.

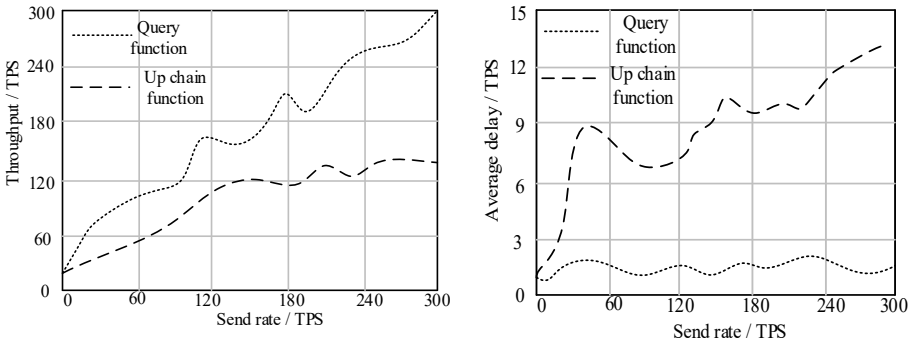

Figure 4 Data governance performance test results

The performance test results are shown in Figure 4, the throughput of the query function is about equal to the transmission rate within the test range, and the average delay remains small. Because the query function does not need to change the state of the chain through consensus. The uplink function keeps stable after the throughput reaches the peak value with the increase of the transmission rate, and the average delay increases faster than the average delay of the query function. This is because the uplink function involves the process of transaction simulation execution, endorsement, sorting and synchronization.

4. CONCLUSION

With a series of technological innovations such as "distributed storage" and "smart contract", block chain overcomes the problems of "information island" and "data right confirmation" of the traditional Internet. It is a "subversive" technological innovation. At the level of government governance structure, block chain reshapes government, market and social boundaries and operation mechanism. At the level of government governance model, block chain promotes innovation in many fields, such as economic regulation, economic and social supervision, public service supply, data governance system, social credit investigation system and so on. Of course, there are limits and challenges for block chain to innovate the government governance system, including the challenges of value ethics, technology empowerment and capital role.

REFERENCES

[1] KF A, ZB B, ML A, AVVC D, WS E. 2020, Dredas: Decentralized, reliable and efficient remote outsourced data auditing scheme with block chain smart contract for industrial IoT. Future Generation Computer Systems, 110(4):665-674.

[2] Rathore, Mohamed, Guizani. 2020, A Survey of block chain Enabled Cyber-Physical Systems. Sensors, 20(1):282.

[3] Cernian A, Tiganoaia B, Sacala I, Pavel A, Iftemi A. 2020, PatientDataChain: A block chain-Based Approach to Integrate Personal Health Records. Sensors, 20(22):6538.

[4] Costa D, Neto A, Pinheiro B, Cordeiro W, Araujo R, Abelem A. 2019, Securing light clients in block chain with DLCP. International Journal of Network Management, 29(3):e2055.1-e2055.21.

[5] Thai QT, Yim JC, Yoo TW, Yoo HK, Kim SM. 2019, Hierarchical Byzantine fault-tolerance protocol for permissioned block chain systems. The Journal of Supercomputing, 75(11):7337-7365.

[6] V Chamola, V Hassija, V Gupta, M Guizani. 2020, A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, block chain and 5G in Managing its Impact. IEEE Access, PP(99):1-1.