A Data-sharing Model Based on Blockchain for Power Grid Big Data

Xudong Zhang1*, Linlin Zhao2, Xiaoxin Gao2, Xu Zhang2
1State Grid ZheJiang Electric Power Co., Ltd, Hangzhou 310007, Zhejiang, China
2Beijing China-Power Information Technology Co., Ltd, Beijing 102208, China
*Corresponding author’s e-mail: jiangtian_20@163.com

Abstract: In the existing smart grid, wireless sensor networks are widely deployed to monitor the entire grid's status to ensure the safe, reliable, and efficient operation of the grid. However, due to the storage mode of the infinite sensor network perception data based on the trusted central node in the smart grid, the smart grid exists information security issues such as the central node's vulnerability to malicious attacks and deliberate data tampering. In response to these information security issues, this paper proposes a data-sharing model based on blockchain technology and digital digest matching algorithm, which has the characteristics of decentralization, security and credibility, non-tampering, and protection of private data. The model uses the consortium blockchain technology to build a smart grid Data Storage Consortium Blockchain (DSCB). The data capability is taken as a unique token through the method of data capability to control the data access to achieve fine-grained access control with a resource as the unit. The capability update is convenient, and the capability withdrawal is supported to ensure the data security sharing between sensor nodes. Security analysis shows that the model can realize safe, effective, and privacy-protected data storage and sharing.

1. Introduction
The smart grid uses a wide range of wireless sensor networks to collect real-time data, monitor the power grid's operation, and realize the dynamic feedback and adjustment of the entire power grid's operation status through the rapid processing of collected data. The Wireless Sensor Network (WSN) has been widely deployed in the power grid due to its low cost, easy deployment, and embedded smart processing to ensure the power grid's safe and reliable operation [1]. Nevertheless, in the traditional smart grid, a large number of wireless sensor nodes are mainly used to monitor the operation status of power grid equipment in real-time, and the data collected through the adjacent data collection base station is uploaded to a trusted centralized node (Power Grid Data Center) so that it can realize data-sharing among units in the power grid [2], as illustrated in Fig. 1. This centralized data storage method faces some information security issues, such as the central nodes are more vulnerable to malicious attacks, and the collected data is more deliberately tampered with.

In response to these security issues, there is an urgent need to design a secure and reliable, decentralized data storage system to ensure the smart grid's smooth operation.
Fig. 1. Centralized data-sharing model of the power grid

Blockchain technology, as a new form of data structure organization, uses data encryption, timestamps, distributed consensus mechanism, economic incentives, and other means to realize peer-to-peer transactions based on decentralized credit, coordination, and collaboration in a distributed system where the nodes do not need to trust each other. It has the characteristics of decentralization, data security, reliability, traceability of historical information, etc., which provides solutions to high cost, low efficiency, and insecure data storage that are common in centralized institutions [2-3]. Blockchain technology has recently been widely used in the medical, electric, and financial fields. [4] proposed a medical data-sharing model based on blockchain, realizing the convenient, secure, and fast data-sharing by classifying the existing medical institutions and using an improved consensus mechanism. [5] presented a new method based on blockchain for the perception of distributed state and an optimization method for the distribution grid to replace the traditional optimization method based on reliable communication. [6] used blockchain to build a DSCB. In the DSCB, data-sharing was accomplished by smart contracts. Data owners set Data-sharing constraints. Computer language was used to replace the legal terms to regulate data access, ensuring collective maintenance for the safe and reliable data storage database in a decentralized manner. [7] provided a detailed analysis of business scenarios, technology construction, operation and maintenance challenges, and business standards for blockchain innovation in the financial sector. [8] presented a model based on the traditional blockchain for securing electrical energy transactions, data security verification, and storage. However, there is no blockchain-based sharing model explicitly designed for the smart grid in the current research. There are many wireless sensor nodes with limited energy in the smart grid, so it is impossible to deploy blockchain directly in the wireless sensor network. Otherwise, the vast computing cost would make the entire intelligent network not work usually.

Focus on blockchain's current application in the smart grid. This paper designs an intelligent grid data-sharing model based on blockchain. The model can solve centralized grid data storage problems, the difficulty in safe sharing and excessive dependence on data centers, and the pain of excessive network overhead caused by blockchain deployment across the entire network. The main design ideas of the model are as follows.

1) Data aggregator (data collection base station) consortium server group and audit server group

The sensor nodes in wireless sensor networks collect and integrate the power grid's big data and send them to the nearby data collection base stations. These base stations are called data aggregators, analyzing the collected data in real-time and responding to the power grid's operation. Different data aggregators communicate with each other and conduct collaborative analysis through a wired network connection. In this paper, instead of using the traditional way of implementing the consensus process among all the sensor nodes, the DSCB preselects some of the data collection base stations to establish
the consortium server group and the audit server group and then uses the consortium blockchain to implement the consensus process to realize data consensus among all of the network nodes.

2) Data block storage and distributed database system

The data aggregator audit server group is responsible for auditing all the perception data from sensors and compressing the data into the data blocks. The sensor nodes do not store data directly. Still, only a summary of the perception data so that a distributed database system is formed to solve centralized data storage.

3) Data access control method based on data capability

This model combines the data access control method based on blockchain's data capability to solve centralized management's data trust problem and ensures data supply security. Both sides of data demand can safely, reliably, and flexibly share data through smart contracts' automatic execution.

2. The structure of the data-sharing model based on blockchain

The data-sharing model based on blockchain contains three significant components as follows.

2.1. Preselected data aggregator

The DSCB in the data-sharing model is a consortium blockchain built on the part of the data aggregator. Consortium chain technology is used to execute the consensus process on the preselected data aggregator. The data aggregator has the right to control the consensus process and strive to obtain the qualification for data writing, thereby getting the rewards. The number of preselected data aggregators is related to the number of nodes in the entire network. From experience, the method to set the number is that if the number of nodes in the whole network is less than 500, the number of preselected data aggregators is 20% of the number of nodes, if the number of nodes in the entire network exceeds 500, the number of preselected data aggregators should be 101. The pre-selection of data aggregators can be obtained by voting. Each data collector votes to the trusted collector and pre-selects enough aggregators according to the number of high to low voices. The pre-selection aggregators acquire data blocks by accounting in turn.

2.2. Data block

All perception data acquired by sensors contain the encrypted pseudonyms of sensor nodes, data types, metadata tags, metadata indexes, records of perception data upload events, etc., stored in the DSCB. All the data is encrypted and signed to ensure its accuracy and verifiability. Meanwhile, in the DSCB, the sensor itself does not store the perception data but only stores the perception data's digital digest. The metadata's specific storage location can be determined from the digest containing the index information of the data. The preselected data aggregator is responsible for managing the jurisdiction's data sensors, including data acquisition and data management. The collected data are processed through the audit server group, the consensus process among the sensors is implemented, and then all the data is compressed into the data blocks. Each newly generated block contains encrypted hash values linked to the previous data block. This hash value can be used to trace and verify an individual data block.

2.3. Proof of work

In this model, the data aggregator's workload proof is similar to that of the bitcoin, and a new data block is generated every once in a while. The accounting power of the new data is obtained through the competition among the existing preselected data aggregators, while the data aggregator calculates the Merkle root of the data block by using the Merkle-Hash process and applies it to the bitcoin blockchain. In the chain, a transaction is formed similar to submitting a bitcoin. Each data aggregator competes to look for valid workload proof, and the one that finds first will be rewarded with credit points and responsible for reviewing the transaction records. The data that have passed the review are combined with data aggregators to form new data blocks in the DSCB.
3. The operation of the data-sharing model based on blockchain

The operation of the data-sharing model based on blockchain mainly consists of data storage and data-sharing. The following introduces mostly the steps of data storage and the critical control algorithm of data-sharing.

3.1. Data storage

Each wireless sensor node needs to be authenticated as a legal node for data transfer by the system administrator in the DSBC. After that, the sensor node will have access to the metadata index base of the current DSCB in the record pool of the neighboring data aggregator. Therefore, the sensor itself does not store the perception data, but only an index list of the perception data, which is used for finding the position of itself in the data block. The process of storing the perception data is shown in Fig. 2.

![Flow chart of the perception data storage based on blockchain](image)

**Fig. 2. Flow chart of the perception data storage based on blockchain**

3.1.1. Upload of the perception data

The sensing node (recorded as \(N_i\)) sends an upload request to the local data aggregator. The upload request contains the currently used pseudonym certificate and the node's digital signature to ensure the data source's reliability and authenticity. After receiving the request, the local aggregator verifies the node's request and identity information and responds to the node's upload request after confirming its legitimacy. In this process, the sensing node uses the current pseudonym's public key to encrypt the perception data and attaches the digital signature of the encrypted data. Then, the public key of the local data aggregator (recorded as \(DA_j\)) is used to encrypt the uploaded records to obtain the final uploaded record data.

3.1.2. Data collection of the data aggregator

The local data aggregator verifies the final uploaded record data. The safe data is stored in the local record pool while the unsafe data is ignored directly.

3.1.3. Proof of work

In the previous chapter, it is introduced that all data blocks are composed of several data sets. Within a fixed time interval (10 minutes), the local data aggregator integrates the security data collected within 10 minutes into a data set (recorded as \(Data\_set\)) and signs the data set to ensure the data set's source legitimacy and verifiability. All data aggregators actively look for sufficient proof of work to obtain the right of recording the data block and economic reward. The data aggregator, which obtains the proof of work, first gets the right to integrate the data sets into new data blocks, stores them in the
DSCB, and receives the corresponding system reward. So far, the bookkeeping management authority's determination in a certain period is completed through proof of work.

3.1.4. Block consensus
Consensus between data aggregators is accomplished by setting up the master node and slave ones. The data aggregator that calculates the proof of sufficient work at the fastest speed is set as the master node (recorded as $DA_{major}$) of the current consensus process. Then the remaining data aggregators are set as slave nodes. In this paper, the Byzantine Fault Tolerance (BFT) consensus mechanism is used for block consensus [9]. The steps are as follows.

1. The master node collects the data sets from slave nodes and integrates them into new data blocks. The master node's digital signature and the hash value of the new data block are attached for inspection and verification. The master node broadcasts the newly generated data block to each slave node for verification.

2. After receiving the data block, the slave node verifies the validity and correctness of the data block through the block's hash value and digital signature sent by the master node, then broadcasts the audit results with their digital signatures to other slave nodes to realize mutual supervision and standard inspection among them.

3. When a slave node receives and summarizes the audit results from other slave nodes, it compares them with its audit results and sends a reply to the master node containing the slave node's audit results, received all audit results, the conclusion of the audit comparison, and the corresponding digital signature.

4. The master node summarizes all audit replies from the slave nodes. If all data aggregators agree with the legality and correctness of the current data block, the master node will gather the data block and the certificates of the slave nodes participating in the audit and send the corresponding integrated digital signature to all slave nodes. The data block will then be stored in the DSCB in chronological order, and the master node will also get rewards from the system.

5. If some data aggregators do not agree with the current audit results, the master node will further analyze and check these data aggregators' audit results. The master node will resend the data block to these data aggregators for a second audit if necessary. If there are still some data aggregators that do not agree with, the principle that the minority is subordinate to the majority will be adopted. If more than a certain percentage of data aggregators agree with the data block, the data block will be loaded into the DSCB using the method mentioned in step 4. The master node further analyzes...
individual data aggregators' audit results and then judges whether these data aggregators have malicious behaviors and deal with those malicious data aggregators in time. This step can help detect and eliminate illegal and malicious data aggregators on time, thus ensuring the whole system's safe and stable operation.

To sum up, an operation framework of perception data storage system based on blockchain can be obtained (shown in Fig. 3).

3.2. Data-sharing between nodes

In this paper, the data access control method based on data capability is adopted to realize the data-sharing between nodes. This method is a typical resource management mechanism, which takes the capability as a unique token to achieve fine-grained access control taking resources as the unit. It is convenient to update permissions and supports permission withdrawal.

The functions of blockchain, smart contract, and capability-based access control method based on data capability are explained as follows (shown in Table1).

| Technology & Method                  | Function                                                        |
|--------------------------------------|-----------------------------------------------------------------|
| Blockchain                          | Open and transparent recording for all contract status and call records |
|                                      | Unalterable and encrypted authority information                  |
| Smart contract                       | Providing programmable business logic implementation             |
|                                      | Automatic execution                                             |
| A capability-based access control method | Providing a manageable and scalable approach for resource management |

The data-sharing process based on data capability is as follows, which is shown in Fig. 4.

Fig. 4. The schematic of data-sharing process based on data capability

(1) The data source (data aggregator $DA_j$) informs the data owner (sensing node $N_i$) about relevant information of the data set.
(2) The data owner (sensing node $N_i$) releases the data set.
(3) Data requester (sensing node $N_k$) sends a data request to the data owner (sensing node $N_i$).
The data access path and corresponding data access capability are packaged and sent to the data requester (the sensing node \(N_k\)) after the data owner (the sensing node \(N_i\)) passes the request verification.

(5) The data requester (sensing node \(N_k\)) applies for data from the data source (data aggregator \(DA_j\)) after obtaining the data access capability.

(6) The data source (data aggregator \(DA_j\)) provides data download after verification, and feeds back the information of the data requester (sensing node \(N_k\)) to the owner, and then updates the access record.

Data-sharing between nodes is performed automatically by the smart contract, and then the variable state changes as a result. The sharing process consists of the following five main steps, such as publishing data information, sending a data sharing request, executing a smart contract, sending a data share, and specifying data access.

1) Publish data information
The data aggregator acts as a data source and takes the data information, data source's address, data owner's address, data summary, and data entitlements as the input of the function. Then it generates the data object and adds it to the management list and finally puts it into the data retrieval list, and outputs the updated list of entitlements, data, and types.

2) Send data-sharing request
The sensing node sends a data-sharing request to the sensing node. The request includes the purpose, time, and the number of data accesses. When the node receives the data-sharing request, it first verifies the node's identity and then verifies the legitimacy of the data sharing request and authorizes access by specifying the constraints of the data access for the request. The constraints include the scope of the data-sharing effectiveness of the data-sharing and the number of data-sharing times, etc.

3) Execute a smart contract (authority interaction)
The data aggregator verifies the authorization information, executes the smart contract, locks the script according to the access constraints set by the node, decrypts the shared data according to the provided symmetric key, uses the public key of the access node to perform asymmetric encryption on the shared data and outputs the result.

4) Send shared data
If the data access node and the accessed node are within the coverage of the same data aggregator, the data aggregator can send the data to the data access node directly; otherwise, the node executing the smart contract sends the encryption result to the neighboring data aggregator of the data access node.

5) Access to designated data
The data access node receives the data and then decrypts the data with its private key and performs data access.

4. Security analysis on the performance of data-sharing model based on blockchain
The security analysis on the performance of the data-sharing model based on blockchain is mainly reflected in the decentralization, privacy protection of node identity information, and security verification of the stored data.

1) Decentralization
The data-sharing model for power grid big data designed in this paper uses the consortium blockchain technology combined with the data access control method based on data capability, which is different from the traditional centralized storage of data. It uses a distributed data storage method and does not rely on any third-party entity. The end-to-end communication method is adopted between sensor nodes, avoiding the risk of the traditional centralized storage of data being vulnerable to
centralized malicious attacks. This decentralized data storage system also has better scalability and reliability.

2) Privacy protection of node identity information

The data storage process encrypts the data collected at different times by using different asymmetric keys to ensure the most secure data storage. Through the execution of smart contracts, the data-sharing limits the preselected data aggregators' access to the data at will. Combined with the data access method based on data capability, the real owner of data, i.e., the sensor nodes are guaranteed to have the right to control data access and set data access conditions.

3) Security verification of the stored data

Using the proof of work mechanism, all the data forming the data blocks are validated and audited by the preselected aggregators. Simultaneously, the preselected aggregators contest through fair competition, which guarantees fairness during data validation and audit authority.

5. Conclusion

The data-sharing model for power grid big data designed in this paper mainly relies on the consortium blockchain technology and the data access control method based on data capability, which realizes a decentralized distributed data storage mode and a safe and reliable data sharing-mode with identity information privacy-protected function. Blockchain has an irreplaceable advantage during the applications on data storage and data-sharing of the power grid. However, there is still a risk that the stored data would be maliciously tampered with for the reason that the consortium blockchain technology performs the consensus mechanism by preselecting some sensor nodes. In the next step, the research will focus on improving the security performance of the consortium blockchain technology.

References

[1] Fadel, E., Gungor, V. C., Nassef, L., Akkari, N., Malik, M. A., Almasri, S., et al. (2015) A survey on wireless sensor networks for smart grid. J. Computer Communications, 71: 22-33.

[2] Shen, X., Pei, Q.Q., Liu, X.F. (2016) Survey of block chain. J. Chinese Journal of Network and Information Security, 2(11): 11-20.

[3] Dong, H., Zhang, C.Y., Yan B.F,. (2009) Research and prospect of block chain technology. J. China Internet, 13(11): 14-19.

[4] Xue, T.F., Fu, Q.C., Wang, C. (2017) A medical data sharing model via blockchain. J. Acta Automatica Sinica, 43(9): 1555-1562.

[5] Li, J., Zhou D.D., et al. (2020) Distributed state perception and optimization of distribution network based on block chain technology. J.Power Capacitor & Reactive Power Compensation, 43(9): 1555-1562.

[6] Wu, Z.S., Liang, Y.H., Kang, J.W., et al. (2017) Secure data storage and sharing system based on consortium blockchain in smart grid. J. Journal of Computer Applications, 37(10): 2742-2747.

[7] Wang, N.G. (2019) Financial applications of blockchain. J. New Business Weekly,8: 106-108.

[8] Xu, Y.R., Fan, Y., Dong, L., et al. (2020) Application of Blockchain in Power Trading. J. Electric Power Information and Communication Technology, 18(6): 42-48.

[9] Crain, T., Gramoli, V., Larrea, M., et al.(2017) Free Byzantine consensus for consortium blockchains.https://arxiv.org/pdf/1702.03068.pdf.