2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) | 978-1-6654-1596-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/AEMCSE51986.2021.00064

Research on Framework of Smart Grid Data Secure Storage from Blockchain Perspective

Sujin SHEN*
Nanjing University of Posts and Telecommunications
Advanced Technology Research Institute
Nanjing China
794790333@qq.com

Chuang SUN
Nanjing University of Posts and Telecommunications
Advanced Technology Research Institute
Nanjing China
390862120@qq.com

***Abstract*—With the development of technology, the structure of power grid becomes more and more complex, and the amount of data collected is also increasing. In the existing smart power grid, the data collected by sensors need to be uploaded and stored to the trusted central node, but the centralized storage method is easy to cause the malicious attack of the central node, resulting in single point failure, data tampering and other security problems. In order to solve these information security problems, this paper proposes a new data security storage framework based on private blockchain. By using the improved raft algorithm, partial decentralized data storage is used instead of traditional centralized storage. It also introduces in detail the working mechanism of the smart grid data security storage framework, including the process of uploading collected data, data verification, and data block consensus. The security analysis shows the effectiveness of the proposed data storage framework.**

Keywords-component;Blockchain; Smart Grid; Data Storage.

I. INTRODUCTION

Smart grid combines the traditional power system and advanced intelligent communication system, sensing and measurement technology, control technology, with comprehensive and perfect security strategy to realize the interaction between power grid users and power suppliers, so as to ensure the smart, reliable, safe, friendly and efficient operation of power grid. However, the openness and inclusiveness of smart grid has led to new security vulnerabilities in communication, authentication, data collection and other aspects, which will lead to a large number of information security issues [1].

In the existing smart grid, the operation of the grid is monitored in real time through the widespread deployment of sensor networks, and the sensor nodes collect grid status data and upload them to a trusted central node for storage and sharing on a regular basis. This centralized data storage mode is faced with information security problems such as centralized malicious attacks, single point failure of central nodes, malicious tampering of stored data in data centers, etc[2]. At present, many methods have been proposed for this centralized data communication and storage mechanism to defend against network attacks[3-5]. However, centralized data storage methods are always accompanied by high risks. In response to these security challenges, there is an urgent need to design a safe and reliable decentralized data storage system to ensure the normal operation of the smart grid.

Blockchain is a chained data structure that generates blocks in chronological order and combines them sequentially. It uses consensus algorithms to generate and update data, and uses encryption technology to ensure that data is not tampered with or forged. According to its characteristics of decentralization, joint maintenance, non-tampering, encryption security, etc., scholars have introduced blockchain technology into power grid-related fields and conducted research on it. Reference [6] proposed a grid monitoring model that allows users to monitor power usage without the need for third-party management, and realizes the efficient operation of the grid system through smart contracts on the blockchain. Reference [7] designed a new distributed blockchain protection framework to enhance the self-defense ability of modern power systems against network attacks. They discussed how to use blockchain technology to enhance the robustness and security of the power grid. Reference [8] combines traditional blockchain technology and digital signatures to ensure safe electrical energy transactions and safe verification and storage of data. Reference [9] proposed a blockchain-based privacy protection and data aggregation scheme to protect the privacy of users in the smart grid.

Based on the application scenarios of smart grid data security storage combined with blockchain technology, this paper proposes a new private blockchain-based grid data protection system, and introduces the working mechanism of the system in detail.

II. BLOCKCHAIN-BASED GRID DATA SECURE STORAGE FRAMEWORK

A. Choice of Blockchain Technology

Block chain technology is the use of block chain to verify the data structure and data storage and use of distributed node consensus algorithm to generate and update the data, the use of cryptography way to ensure the security of data transmission and access, the use of automated script code intelligent contracts to programming and operating data of a new kind of distributed infrastructure and computing paradigm [10]. To put it simply, in the blockchain system, the transaction data generated by each participant will be packaged into a data block after a period of time. The data blocks are arranged in chronological order to form a chain of data blocks. Each participant has the same data chain and cannot be tampered with unilaterally.

In different application scenarios or design systems, block chains are generally divided into public chains, consortium

978-1-6654-1596-5/21/$31.00 ©2021 IEEE
DOI 10.1109/AEMCSE51986.2021.00064

chains and private chains according to the access mechanism [11]. The public chain is a completely decentralized blockchain, such as Bitcoin and Ethereum. On the public chain, each node can freely join and exit the network, participate in the reading and writing of points on the chain, and there is no centralized server node in the network. Alliance chain is mainly built for specific companies or organizations, which is composed of multiple institutions to jointly maintain a blockchain, and most of them use the algorithm based on PBFT. TPS can exceed ten thousand, but they are still vulnerable to the influence of PBFT algorithm. When the number of nodes is too large, consensus is difficult to reach and efficiency is greatly affected. Private chain is similar to alliance chain, but private chain is managed by a single institution or organization. The authority of each node in the system needs to be allocated by the organization, and the amount of data open to each node should be determined by the organization according to the situation. In this paper, the private chain will be used to build a secure storage blockchain for grid data, and the improved RAFT algorithm will be applied to the private chain, which not only preserves the efficiency of the RAFT algorithm, but also ensures the system stability and other nodes are not affected by malicious node attacks.

B. Data storage private chain system framework

From the perspective of private block chain, the technical framework of smart grid alliance chain data protection system can be divided into three core levels , which are: power grid data acquisition layer, power grid data network layer, and power grid data consensus layer. Each level has its own core functions, which cooperate with each other to form the basic framework of data storage private chain system (DSPC).

1) Acquisition layer: In smart power grid, sensor network uses sensor nodes to monitor the power distribution, transmission and power generation equipment in the power grid in real time and collect monitoring data, including voltage, current, circuit breaker status, transformer wiring position and so on.

2) Network layer: The network layer of the DSBC includes the networking mode of block chain and data authentication protocol and other technologies to ensure that all block nodes can participate in the transmission and verification process of power grid data information. The network layer of the DSBC adopts the P2P network for data transmission. In the network layer, all block nodes have the same function module, but the initial function of each block node in the network is defined by the staff. According to the different permissions of block nodes, all block nodes are divided into two parts: master node and secondary node. All nodes are responsible for the transmission of grid data and the authentication of block information as well as the network routing protocol. The data storage private chain proposed in this paper is a data storage system of "partial decentralization and distributed storage". The advantage of this mode is to ensure data security and improve the efficiency of data uploading and storage in the whole system. The architecture diagram of the data storage private chain system is shown in Fig. 1.

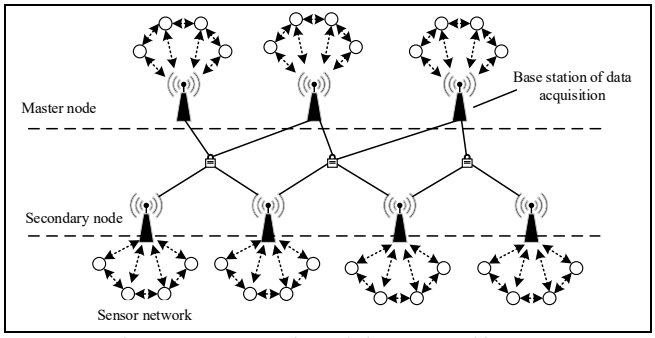

Fig. 1. Data storage private chain system architecture

3) Consensus layer: Consensus mechanism in the DSBC is the most important part, which enables distributed nodes to reach consensus quickly for different types of grid data. Common consensus mechanisms include: proof of work, authorized share proof, proof of interest, practical Byzantine fault tolerance algorithm, Ribbo consensus mechanism and so on. These consensus mechanisms all have their own characteristics. For example, POW mechanism ensures that the information in the power grid can be quickly authenticated by making the blocks in the power grid system maintain the power grid data together based on the competitive computing power. The POS mechanism can set the rights and interests of the block accounting rights of the grid data, which improves the efficiency of the block nodes to reach a consensus. The PBFT is capable of accommodating nearly one-third of grid error node errors, avoiding blockchain forking. There are various consensus mechanisms, but according to the demand of smart grid data storage private chain, these consensus mechanisms cannot meet the requirements. Once there is too much node data, the speed of reaching consensus among nodes will be greatly reduced. RAFT algorithms are leader-based algorithms that use leader election as an important part of consensus protocols. Compared with PBFT, RAFT algorithm is efficient and simple. However, RAFT does not consider Byzantine fault tolerance, but only considers non-human problems such as system node downtime and network failure. In this paper, the improved RAFT algorithm is used to improve the consensus speed between block nodes and prevent malicious node attacks.

III. WORKING MECHANISM OF DSBC

In the framework proposed in this article, the collected data is ultimately stored in the ledger as blocks, which exist as chains in each block node. Before storage, in order to ensure the accuracy of the data, each of the following procedures needs to be performed: data encryption; Data broadcasting; Data verification; Accumulation of data content and data synchronization. Unlike a typical blockchain system, we remove the process of mining. In the process of data linking, the mining mechanism causes a lot of time loss. Removing the mining mechanism can improve the speed of data linking. In this section, we will describe in detail the working mechanism of the proposed framework, which includes data transfer, data validation, and data storage.

A. Data collection and uploading

The sensor node becomes a legitimate node of the sensor network after being authenticated by the system manager, and obtains the pseudonym set and certificate used for encrypting data, which is represented as $\{PK_{Ni}, SK_{Ni}, Cert_{Ni}\}$。The sensing node N_i first sends the data storage request to the local data base station DS_j, in which the data storage request contains the pseudonymous certificate $Cert_{Ni}$ and digital signature Sig_{Ni} of the node, so as to ensure the reliable request and authenticity of the data source. After receiving the storage request, the local data base station verifies the request and identity information of the node, and responds to the storage request of the node after confirming the validity of the request. The sensor node uses the public key PK_{Ni} of the current pseudo name to encrypt the stored data $Data_{Ni}$ and attach the digital signature of the encrypted data. Then it uses the public key PK_{DS_j} of the local data base station DS_j to encrypt the storage Record to get the final stored data $Record$. The process is as follows:

$$Record_{(N_i \to DS_j)} = E_{PK_{DS_j}}(Data_{E_{N_i}} || Cert_{N_i} || Sig_{Sign_{N_i}} || timestamp) \quad (1)$$

$$Data_{E_{N_i}} = E_{PK_{N_i}}(Data_{N_i} || timestamp) \quad (2)$$

$$Sig_{Sign_{N_i}} = Sign_{SK_{N_i}}(Data_{\{}E_{N_i}) \quad (3)$$

Among them, PK_{DS_i} is the public key of the data base station DS_j, $E_{PK_{DS_j}}$ represents the use of PK_{DS_j} to encrypt information, PK_{N_i} is the public key of entity N_i, $E_{PK_{N_i}}$ represents the use of PK_{N_i} to encrypt information, $Data_{N_i}$ is the original data collected by the sensor node N_i, and timestamp is the timestamp. SK_{N_i} is the private key of sensor node N_i, and $Sign_{SK_{N_i}}$ represents the signature data after hashing $Data_{E_{N_i}}$ with the private key of sensor node N_i.

B. Data verification

The local data base station DS_j receives the data $Record$ uploaded by the sensing node and verifies it. It decrypts the data using $DS_j's$ own private key SK_{DS_j}, extracts the $Cert$ field, verifies the identity information and determines that the data comes from the sensing node N_i. The Sig field is extracted and decrypted with $N_i's$ public key PK_{N_i}, and then the $Data$ is encrypted with hash to get the hash value, which is verified with the hash value decrypted from the digital signature sent by N_i from the sensing node. If the results are the same, the verification is passed, and the data is safe and effective. If the results are different, the verification is not passed and the data is directly ignored. The process is as follows:

$$D_{SK_{DS_j}}(Record) = (Data || Cert || Sig || timestamp) \quad (4)$$

$$D_{PK_{DS_j}}(Sig) ? = Hash(Data) \quad (5)$$

Among them, $D_{SK_{DS_j}}(Record)$ represents the decryption process of Record using the public key of DS_j, and $D_{PK_{N_i}}(Sig)$ represents the decryption process of Sig using the public key of N_i.

C. Consensus process of data blocks

All the data base stations are divided into the data master base station (MBS) and the data secondary base station (SBS). MBS is the master node in the consensus process, and the election Leader node in the master node is set as DL_j according to RAFT consensus algorithm, and SBS is the secondary node in the consensus process. In this paper, RAFT consensus algorithm is adopted for block consensus, and the specific consensus process is as follows:

Step 1: The Leader node collects and sorts the data set of the other nodes. When the data volume reaches the threshold, the Leader node integrates the data set into a new data block, with the digital signature of the Leader node and the hash value of the data block attached. The Leader node broadcasts the newly generated data block to all others. The process is as follows:

$$Record_{DL_j \to All} = (Data_{DL_j} || Data\_hash || Cert_{DL_j} || Sig_{DL_i} || timestamp) \quad (6)$$

$$Data\_hash = Hash(Data_{DL_j} || timestamp) \quad (7)$$

$$Sig_{DL_j} = Sign_{SK_{DL_j}}(Data_{DL_j} || Data\_hash) \quad (8)$$

Step 2: After receiving the data block, the other master nodes verify the validity and correctness of the data block through the information such as the block hash value and digital signature sent by the Leader node, and feed back their audit results (Result) with their digital signatures to the Leader node, and temporarily store the data block in the local log.

Step 3: The Leader node receives and aggregates audit feedback from the remaining master nodes. If more than half of the master nodes agree on the validity and correctness of the current block, the Leader node will send the data block together with the certificate set ($\{Cert_{DL}\}$) of the master node participating in the audit and the corresponding digital signature to all the nodes, and write the pending block in the local log into the blockchain. The master node that receives the feedback result consistent with the audit result writes the block of the local log pending chain into the blockchain, and the master node that receives the feedback result contrary to the audit result updates the local blockchain according to the data block in the feedback result. After receiving the incoming data

packet from the master node, the secondary node updates the data block to the local blockchain and records the nodes with audit errors in this round of the master node. The nodes with continuous audit errors will be demoted to the secondary node, and a node will be elected from the secondary nodes to upgrade to the master node. The process is as follows:

$$Data\_block_{DL_j \to All} = \left(Data\_2 || Sig_{DL_j} || timestamp\right) \quad (9)$$

$$Data\_2 = (Data_{DL_j} || Data\_hash || \{Cert_{DL}\} || timestamp) \quad (10)$$

$$Sig_{DL_j} = Sign_{SK_{DL_j}}(Data\_2) \quad (11)$$

Step 4: If the Leader node receives the audit results, more than half disapprove of the legitimacy of the current block and correctness, Leader nodes will lose Leader status, and from the master node to run for the Leader node, at the same time deputy node records to the master node for errors in the audit process, audit will be demoted to deputy error of node in a row, and from the pair of nodes for upgrade to give priority to a node.

IV. Safety analysis

Asymmetric encryption technology is used in the smart grid data storage alliance chain proposed in this paper, which has a good ability to resist the traditional security attacks. Through the encryption and verification mechanism, the attacker can not break the encrypted information in a short time. And in the process of communication, we use digital signature technology to resist the attack of attackers disguised as legal entities or forged false information; No entity can forge another entity's digital signature without the signer's private key; And legitimate entities can verify that the information they receive has not been changed through digital signatures.

For ordinary sensor nodes, even if a sensor node is invaded by an attacker and uploads forged data to the data base station, through the authentication mechanism, data verification mechanism and consensus mechanism, these attacked data will also be found by other data base stations in the audit and inspection of problems. Data can only be tampered with if the attacker controls more than 50% of the nodes. For the data base station that elects the master node, RAFT consensus mechanism is adopted among nodes. If the attacker attacks the master node of the Leader, the other master nodes will refuse to give feedback to the Leader when problems are found in the audit and inspection. The incumbent Leader will be forced to step down, and the other master nodes will re-elect as Leader. Only when the attacker controls more than 50% of the master nodes can the data tampering attack be successfully launched.

V. Conclusion

In view of the information security problems in the centralized storage of grid status data in smart grid, such as the central single point attack and the data being tampered with intentionally, the block chain technology is used to respond to the decentralized distributed storage. In this paper, a smart grid data protection framework based on alliance block chain is proposed. Combined with smart grid data storage scenarios, block chain technology is applied to solve the problem of secure tamper-proof storage of grid status data. The working mechanism of identity authentication mechanism, data collection uploading, data verification, data consensus, data block on the chain and so on are introduced in detail. The author hopes that this study can provide an innovative reference for blockchain technology in the technical framework of smart grid data secure storage.

References

[1] TANG Yi, CHEN Qian, LI Mengya, et al. Overview on Cyber-attacks Against Cyber Physical Power System [J]. Automation of Electric Power Systems, 2016,40(17):59-69.

[2] LIU Ting, TIAN Jue, WANG Jiazhou, et al. Integrated Security Threats and Defense of Cyber-physical Systems [J]. Acta Automatica Sinica, 2019, 45(1): 5-24.

[3] YAN Ye, QIAN Yi, SHARIF Hamid, et al. A Survey on Cyber Security for Smart Grid Communications [J]. IEEE Communications Surveys & Tutorials, 2012, 14(4):998-1010.

[4] ASHOK A , GOVINDARASU M , AJJARAPU V . Online Detection of Stealthy False Data Injection Attacks in Power System State Estimation [J]. IEEE Transactions on Smart Grid, 2018, 9(3):1636-1646.

[5] KURT M N, YILMAZ Y, WANG X. Real-Time Detection of Hybrid and Stealthy Cyber-Attacks in Smart Grid [J]. IEEE Transactions on Information Forensics & Security, 2019, 14(2):498-513.

[6] J. Gao et al., "GridMonitoring: Secured sovereign blockchain based monitoring on smart grid", IEEE Access, vol. 6, pp. 9917-9925, Mar. 2018.

[7] G. Liang, S. R. Weller, F. Luo, J. Zhao, Z. Y. Dong, "Distributed blockchain-based data protection framework for modern power systems against cyber attacks", IEEE Trans. Smart Grid, vol. 10, no. 3, pp. 3162-3173, May 2019.

[8] N. Z. Aitzhan, D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures blockchain and anonymous messaging streams", IEEE Trans. Depend. Sec. Comput., vol. 15, no. 5, pp. 840-852, Sep./Oct. 2018.

[9] Z. Guan et al., "Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities", IEEE Commun. Mag., vol. 56, no. 7, pp. 82-88, Jul. 2018.

[10] M. Sharples and J. Domingue, “The blockchain and kudos: A distributed system for educational record, reputation and reward,” in Proc. 11th Eur. Conf. Technol. Enhanced Learn., Lyon, France, Sep. 2016, pp. 490–496.

[11] M. Pilkington, “Blockchain technology: Principles and applications,” in Research Handbook on Digital Transformations, F. X. Olleros and M. Zhegu, Eds. Cheltenham, U.K.: Edward Elgar, 2016.