Research on Distributed Network Data Storage Security based on block chain

Qiuxiang Li*, Zhiyu Liu, Yanru Chen

The First Research Institute of the Ministry of Public Security of P.R.C, 1 Shouti Nan Lu, Haidian District, Beijing, China

* Corresponding author: 15010189805@163.com

Abstract. With the advantages of mass storage and easy expansion, distributed storage architecture has become the main way of user big data storage, but the confidentiality, integrity and availability of data are facing serious challenges. As a new distributed shared ledger, blockchain technology has become an important technology to ensure data security since it was proposed by virtue of its characteristics of decentralization, tamper-proof and anonymity. Based on blockchain technology, this paper studies the key technology and application scheme design of data distributed storage, and verifies the feasibility of the scheme.

Keywords: Blockchain, storage security, distributed storage, decentralization, privacy protection, big data, encryption, tamper-proof

1 INTRODUCTION

The Internet era has come, and network data is also produced synchronously. After data mining and analysis of network data, the development law of things can be obtained and applied to the actual production and life to produce social value. At the same time, the security problem of network data is getting worse and worse, and the privacy protection and data storage security in the process of data transmission cannot be guaranteed, such as the theft of sensitive user information, distributed denial of service attacks, etc. As a result, the access permission of data is confused, the risk of information leakage is increased, and the difficulty of security protection is increased. Therefore, how to ensure the security of network data storage is an urgent focus.

With the continuous expansion of the overall scale of the computer system, the global storage of data information surges, the emergence of massive data makes the data storage, processing and other aspects have a huge pressure. Faced with such a large amount of data information, traditional centralized storage is prone to data loss, personal information and privacy disclosure, network theft and other problems, and centralized storage occupies a large network broadband, resulting in the operation of the storage system difficulties [1]. Distributed storage can facilitate broadband expansion, and the security of data storage is relatively high, is currently the main applica-

A. Rauf et al. (eds.), *Proceedings of the 3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, Atlantis Highlights in Engineering 20,
https://doi.org/10.2991/978-94-6463-262-0_44

tion mode of data storage, by all walks of life. In the distributed network environment, users can store data more securely and transparently, and the storage cost is lower. In addition, the time of data query and invocation is shortened. Although distributed storage has many advantages, the data security strategy adopted by distributed storage still has obvious limitations, and the failure of a node may cause the loss of key data. Therefore, finding a secure and stable network data storage method is still an important target of data security [2]. Blockchain technology, as a special distributed database, has the characteristics of decentralization. It is an emerging electronic bookkeeping system. There is no centralized controller in the blockchain system to control the whole system, and all nodes store all or part of the data. If the data of a node is lost, the node will automatically synchronize data to other nodes, so as to not affect the safe operation of the whole system and greatly improve the reliability of data storage [3]. This paper combines distributed storage and blockchain technology to form a new storage mode, which has the advantages of distributed data storage, encryption algorithm, consensus mechanism and point-to-point transmission. It effectively solves the trust problem between multiple points and data security problems, and has a bright prospect in the field of network data storage.

2 Technologies related to network data storage security

2.1 Blockchain technology

The storage of data by blockchain technology is mainly achieved through the block structure, and its data maintenance is completed with the joint participation of multiple parties. The security of data storage is guaranteed by cryptographic technology. Under the synergistic effect of encryption algorithms, transaction information will be recorded in the blockchain system according to the time sequence of information generation, and the corresponding time stamp will be attached. All organization member nodes in the network system realize data sharing, recovery and consistency among each other through the distributive ledger. The distributive network structure ledger is recorded and stored forever in a continuous chain of encrypted hash blocks. This can be done by all participant nodes in a blockchain network system by recording updates in a peer-to-peer network structure ledger or by consensus in the absence of a central organization or trusted third party. But the books cannot be updated or withdrawn unless each network member has reached a consensus during subsequent transactions. If the digital block is to be updated, it must be realized with the consent of all transaction participants. Therefore, under the blockchain technology, network data is difficult to be modified, intercepted or deleted by attackers. Moreover, the security hash calculation of the core technology of blockchain, also known as the security hash algorithm, can be used to construct the tree architecture of data and information security. Nodes in the tree structure can be used to detect network information statistics and more effectively ensure the sense of reality and integrity of data. Therefore, blockchain technology has a series of advantages such as security, traceability, high credibility, time stamp and decentralization [4].

2.2 Distributed storage technology

Distributed storage technology usually adopts a two-level architecture, in which operation, maintenance and management functions are mainly distributed in the upper architecture. Distributed storage system provides multiple nodes, and storage requests can be distributed to multiple machines for simultaneous processing, so as to achieve high-performance data storage [5]. Data nodes in this layer are scattered and sinking, providing service access and data storage functions. Distributed storage technology applies the cluster construction mode. After the data load is written, it will be shared to each node in the cluster, and then data slices will be formed for data storage. Therefore, this data storage technology has high writing efficiency, supports data reconstruction, and improves the security of written data [6].

2.3 Distributed network data storage model based on blockchain

According to the structural characteristics and technical advantages of blockchain technology and distributed storage technology, this paper builds a distributed network data storage model based on blockchain, as shown in Figure 1. Relying on the consensus rules between blockchain nodes, it can effectively solve the problem of multi-point mutual trust, improve the security of data information storage, improve the confidentiality and access of data, and solve the core problem of big data storage.

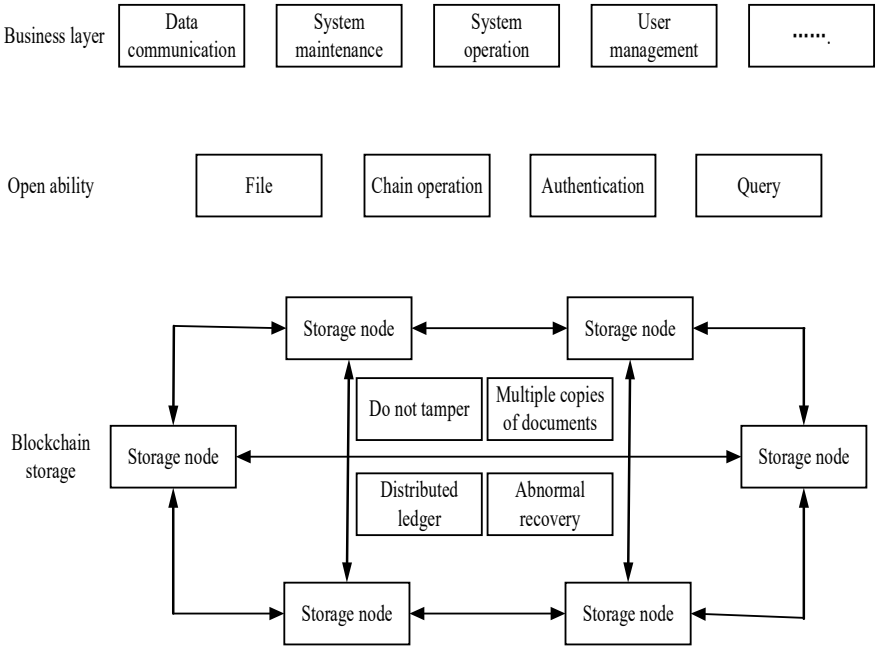

Fig. 1. Distributed network data storage model based on blockchain

3 Design of distributed network data security storage scheme

3.1 Storage security analysis

Combined with the above technologies and the threats faced by network data storage, this paper analyzes the storage security of the designed distributed network data security storage scheme, and obtains the corresponding scheme according to the following requirements, so as to solve the potential storage security threats.

(1) Anonymity: Data is stored in the form of strings. In blockchain, the identities of users and all nodes are represented by different hash strings, through which the user's identity information can be obtained. It is crucial to ensure the anonymity of the system.

(2) Recoverability: Distributed storage systems often achieve system reliability through data redundancy strategy, which is one of the good choices for the current technology. Data is replicated in blocks. When a data block is damaged, the corresponding redundant blocks can be selected for recovery, ensuring that the original data file is always available.

(3) Encryptibility: In the cyber world, openness and security have always been at odds. Security threats to data are everywhere, whether it is hackers in the network world, or system security loopholes, may lead to data security damage. One of the basic technologies in blockchain technology is asymmetric encryption. Different implementations of blockchain networks use different encryption algorithms, which are currently resistant to quantum computer deciphering. Splitting a data file into blocks and storing the blocks in different places makes the data secure both logically and physically.

(4) Fault tolerance: Distributed storage system can effectively avoid the "single point of failure" problem. When a node fails, the system needs to find another node to replace it. Ensure that the user client does not feel the data inconsistency problem, improve the fault tolerance of the delivery system.

(5) Scalability: Good scalability often requires that P2P storage systems can add peer hosts transparently and openly to achieve system expansion. In addition, you can safely delete peer hosts to reduce capacity. In the blockchain network, the entry and exit of nodes are very simple and convenient. The horizontal expansion of the system can be realized through the addition of more nodes, so as to linearly improve the overall capacity and performance of the system and reduce the dependence on specific devices.

3.2 Specific scheme design

From the advantages of the above combination mode of blockchain technology and distributed storage, we can know that the data storage mode combined by the two is more conducive to ensuring the security of data information storage. Therefore, this paper designs the application scheme of the combination mode in data storage, so as

to ensure the security of data storage. The specific application scheme designed in this paper is shown in Figure 2.

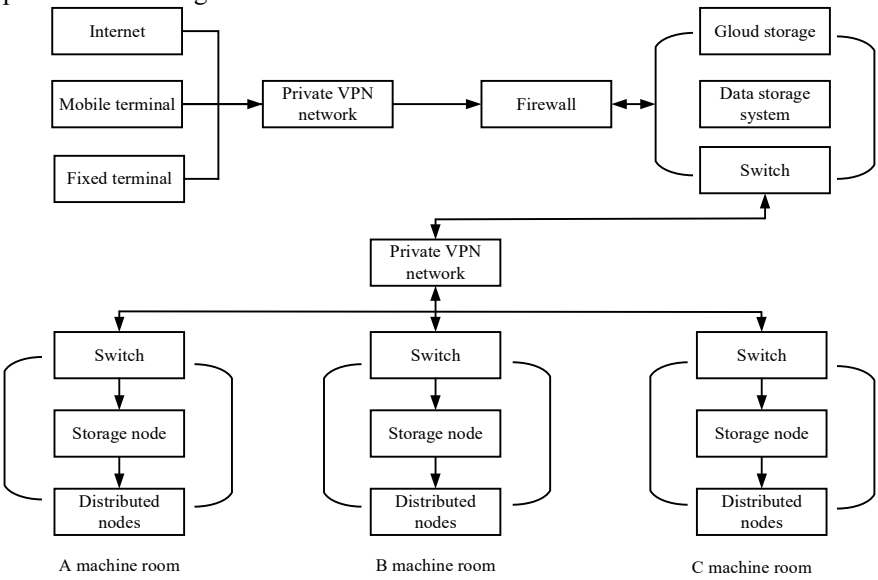

Fig. 2. Distributed network data storage application scheme based on block chain

Firstly, the blockchain service platform is deployed in the cloud storage. Users can upload important data and data files to the deployed blockchain service platform using the Internet or Virtual Private Network (VPN). Subsequently, blockchain function nodes are deployed on each distributed storage node to avoid the risk of direct exposure of stored data in the public network. Each node of the blockchain interacts with the main node of the service platform deployed in the cloud resource platform and VPN network, thus enhancing data confidentiality. When deploying the blockchain service platform, storage space should be allocated according to the actual needs of users, and the Internet Protocol (IP) blacklist and whitelist of the public network should be restricted according to the minimum access policy. At the same time, the required copy storage should be configured. The processing of network data requires writing into the terms of the smart contract and waiting for execution, and receiving network data executes the smart contract. The contract represents the storage space of the storage service provider in the form of code. When the contract reaches the trigger conditions and implements automatic execution, the storage space is dynamically updated. Because the execution of smart contracts is implemented based on a consistent consensus algorithm, the results are publicly searchable. Therefore, it can fundamentally avoid the connection between the data owner and the service provider with insufficient storage capacity, and the user can obtain the right to access the network data stipulated in the smart contract terms through the asymmetric encryption technology.

Then, considering the user's authority on the blockchain platform, the important data files will be uploaded to the blockchain business platform, which can encrypt the data file information onto the chain, and then asynchronously store it to each storage node according to the user's storage requirements, usually adopting the mode of multi-copy remote storage. The reliability and security of data storage can be greatly improved. At this time, all nodes in the blockchain can be used as forensic data to jointly maintain the security of data stored in the system, so that the obtained network data cannot be deleted or tampered with once on the chain. By periodically scanning data information files, the system analyzes whether the data stored by each node is missing or damaged. If there is an error, the data file will be verified, and the correct data information can be retrieved from other blockchain nodes. Finally, the damaged or lost data can be repaired, so as to ensure the security of the whole process of network data storage.

3.3 Application effect

In this scheme, the attacker cannot open the encrypted power data information system by brute force cracking in a short time. At the same time, the scheme will add the time stamp into the data information, once the attacker enables the replay attack, it will effectively resist the attack. In the whole communication process, generally, the communication node will resist the fake information attack formed by the attacker through the digital signature technology, but in this scheme, this method cannot realize the protection of power data information through the verification mechanism.

This scheme can protect the identity and privacy of nodes. Virtually every node in the system adopts "pseudonym protection" to complete the communication, and the communication parties will not obtain the data information of the communication nodes. Meanwhile, the data collected at various periods will be encrypted by asymmetric keys during the data storage, so that the data can be stored safely. And the scheme will implement data sharing through smart contracts, so that the data aggregator has arbitrary access to the data.

For ordinary nodes, even if a node colludes with the attacker to falsify data, this scheme can discover the attacked data in time through examination through consensus mechanism and workload proof. Only when the attacker controls more than half of the nodes can the data be tampered with, effectively ensuring the authenticity and legitimacy of the data. The data tampering attacks launched by aggregators can also be effectively defused, and the probability of success of data tampering is very small. Finally, data cannot be falsified under this scheme. The essential attribute of the application scheme after the combination of blockchain technology and distributed data storage is that it can be combined with digital signature technology, so as to ensure that attackers cannot interfere with the network data storage process by entity impersonation.

4 summarize

As a symbol, network data has important epistemological significance. Distributed storage based on block chain technology creates a new path for network data security. The security of data storage is related to the development quality of digital economy. Ensuring the security of data storage can improve the stability of system operation and promote the sustainable development of industry. At present, the disadvantages of centralized storage are very prominent, while distributed storage technology also has some shortcomings. Therefore, this paper combines blockchain technology and distributed storage technology to design a feasible scheme conducive to improving the security of network data storage, which can provide certain guarantee and reference for the development of data storage field.

References

1. Z. Chi, F. Zhang, Z. Du and R. Liu, "Cloud storage of massive remote sensing data based on distributed file system," 2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013), KunMing, China, 2013, pp. 1-4, doi: 10.1109/ICSPCC.2013.6663922.
2. K. Renuga, S. S. Tan, Y. Q. Zhu, T. C. Low and Y. H. Wang, "Balanced and Efficient Data Placement and Replication Strategy for Distributed Backup Storage Systems," 2009 International Conference on Computational Science and Engineering, Vancouver, BC, Canada, 2009, pp. 87-94, doi: 10.1109/CSE.2009.27.
3. G. Si, Y. Sun, W. Chen and L. Chen, "Node Switching Method in Power Distribution Internet of Things based on Blockchain," 2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC), Chongqing, China, 2020, pp. 291-295, doi: 10.1109/ICCEIC51584.2020.00062.
4. R. Wang, J. He, C. Liu, Q. Li, W. -T. Tsai and E. Deng, "A Privacy-Aware PKI System Based on Permissioned Blockchains," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 928-931, doi: 10.1109/ICSESS.2018.8663738.
5. T. Wang, "An Analytical Model of Distributed Energy Storage Systems in Power Distribution Networks," 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Wuxi, China, 2018, pp. 1-4, doi: 10.1109/DCABES.2018.00011.
6. X. -Y. Yang, Z. Liu, W. Zhang and D. -T. Guo, "A High-efficiency Data Distribution Algorithm in Distributed Storage," 2009 Fifth International Conference on Information Assurance and Security, Xi'an, China, 2009, pp. 627-630, doi: 10.1109/IAS.2009.225.