Creativity and Innovation
2025, VOL. 9, NO. 1, 44-49
DOI: 10.47297/wspciWSP2516-252709.20250901

Design And Implementation Of A Distributed Network Storage System Based On Blockchain Technology

Zhenxiang Guan, Hongyan Sun

Liaoning University of Science and Technology, Anshan, Liaoning 114051

ABSTRACT
With the rapid development of information technology, the data storage demand is increasing, and the traditional centralized storage system faces many challenges, such as data security, privacy protection and single point of failure. Blockchain technology provides new ideas for the design and implementation of distributed network storage system by virtue of its decentralization, non-tampering and high transparency characteristics. This paper firstly introduces the basic principle of blockchain technology and the demand background of distributed network storage system, and then elaborates the architecture design of distributed network storage system based on blockchain technology, including data storage mechanism, consensus mechanism and security mechanism. The performance and security of the system are verified through experiments, and the results show that the system has significant advantages in data storage efficiency, data security and privacy protection. Finally, this paper analyzes the challenges faced by the current system and looks forward to the future development direction, including technological innovation, application expansion, standardization and legal regulation.

KEYWORDS
blockchain technology; distributed storage; data security; privacy protection; system design that

1 Introduction

In recent years, domestic scholars and research institutions have conducted extensive research on the combination of blockchain technology and distributed storage. For example, a research team from Sun Yat-sen University, in the paper "Blockchain Storage: Technology and Challenges", reviewed many aspects of the combination of blockchain and storage technology, including decentralized storage system built on blockchain, optimization of the storage performance of the existing system based on blockchain, and the storage method and optimization of the blockchain system itself. They pointed out that blockchain technology can effectively improve the data security and scalability of the storage system, but also faces challenges such as query performance bottlenecks and increased storage space overhead. In addition, some domestic enterprises are also actively exploring the application of blockchain technology in distributed storage, such as Huawei, Aliyun, etc. By combining blockchain technology with cloud computing, big data and other technologies, they have developed a series of innovative distributed storage products and services.

2 The foundations of blockchain technology

2.1 Blockchain Fundamentals

Blockchain is a distributed ledger technology, the core of which lies in recording and verifying transaction information in a decentralized manner. Each block contains a certain number of transaction records and is linked to the previous block through cryptographic algorithms, forming a chain structure that cannot be tampered with. This structure ensures data integrity and tamperability, and provides a reliable trust basis for distributed network storage systems. The decentralized nature of blockchain eliminates the risk of a single point of failure in traditional centralized systems and enhances the reliability and availability of the system. In addition, the distributed ledger mechanism of blockchain allows each node to store a complete copy of the data, which further improves data redundancy and security.

2.2 Cryptography of the blockchain

Encryption technology is the key guarantee of blockchain security. Blockchain mainly adopts asymmetric encryption algorithms, such as elliptic curve cryptography (ECC), to generate a pair of public and private keys for each user. The public key is used to encrypt the data, while the private key is used to decrypt the data, ensuring that only users with the corresponding private key can access and modify their data. In addition, blockchain also makes extensive use of hash algorithms, such as SHA-256, to hash the transaction data in the block and generate a unique hash value. The hash value has one-way and collision resistance, any small changes to the data will lead to significant changes in the hash value, thus ensuring the integrity of the data and tampering. Through these encryption techniques, blockchain can effectively protect the privacy and security of user data, and provide strong security for distributed network storage systems.

2.3 The consensus mechanism of the blockchain

Consensus mechanism is a key mechanism for nodes in a blockchain network to reach agreement on the validity of transaction records and blocks, ensuring that all nodes agree on the validity of transaction records and blocks. Common consensus mechanisms include Proof of Work (PoW), Proof of Stake (PoS), and Byzantine Fault Tolerance (PBFT), etc. PoW verifies transactions and generates new blocks by calculating complex mathematical puzzles, which is highly secure but consumes a lot of computational resources, while PoS selects block verifiers based on the number of tokens held by the nodes and the amount of time spent on it, which is relatively energy-efficient and highly effective. PBFT mechanism verifies transactions and generates blocks through voting and consensus among nodes, which can ensure the normal operation of the system in the case of malicious nodes. Different consensus mechanisms are suitable for different application scenarios and needs, and choosing the appropriate consensus mechanism is crucial for the performance and security of distributed network storage systems.

3 Overview of distributed networked storage systems

3.1 Key Technologies for Distributed Storage

The core technologies of distributed storage systems include the following.

Data slicing and distribution: the data will be divided into multiple small pieces, and scattered storage in different nodes. Data slicing can be realized through hash algorithms, consistent hash and other technologies to ensure that the data is evenly distributed among the nodes. For example, consistent hash algorithm can effectively reduce the node addition and deletion of data migration, improve the scalability of the system.

Data Redundancy and Fault Tolerance: Ensure data redundancy and backup through data copy or code correction technology to improve the fault tolerance of the system. Data copy strategy is simple and reliable, but will increase the storage space overhead; corrective code technology in the storage space savings at the same time, increase the computational complexity of data recovery.

Consistency protocols: distributed storage systems need to use consistency protocols to solve the problem of data consistency between multiple nodes. Common protocols include Paxos, Raft and so on, these protocols through the election, voting and data synchronization mechanism to ensure that all copies of the data to maintain consistency.

Load balancing: By dynamically adjusting the distribution of data between nodes, load balancing is achieved to improve the overall performance of the system. Load balancing can monitor the load of the node, automatically migrate data to the node with lower load.

Metadata management: Metadata is the information that describes the attributes of the data and is used to support the storage location, access rights and other functions of the data. Distributed storage systems need to efficiently manage metadata to ensure system performance and reliability. Metadata management can be achieved in a centralized or distributed way, and different management methods have their own advantages and disadvantages.

Data encryption and security: In order to protect the privacy and security of data, distributed storage systems usually use encryption technology to encrypt data. Data encryption can be carried out on the storage node or in the process of data transmission to ensure the security of data in the storage and transmission process.

3.2 Application Scenarios for Distributed Storage

Distributed storage systems are widely used in a variety of fields and have important practical applications.

Big data storage: Distributed storage can efficiently store and process large-scale data, such as Internet data, sensor data, video data and so on. For example, HDFS is a core component in the Hadoop ecosystem, which is widely used in big data processing and analyzing tasks, supporting efficient storage and access to large-scale data.

Cloud storage services: Distributed storage can be used to build a cloud storage platform, providing cloud storage, cloud backup, cloud archiving and other services. For example, Amazon S3 is a kind of cloud storage service based on object storage, supporting massive data storage and access, with high availability and high scalability.

Data analysis and mining: Distributed storage systems can be used in combination with distributed computing frameworks (such as Hadoop, Spark, etc.) for large-scale data analysis and mining, such as machine learning and data mining. For example, Ceph's high performance and high availability make it an ideal storage solution for big data analysis.

Content Delivery Network (CDN): Distributed storage can be used to build CDNs, storing data in locations close to users to improve the efficiency and speed of content distribution. For example, FastDFS is a lightweight distributed file system for file storage and distribution, which can effectively improve file access speed.

Internet of Things (IoT) applications: Distributed storage can be used to store and process massive amounts of sensor data, such as smart home, smart transportation, smart health, and so on. For example, data generated by IoT devices can be stored and analyzed in real time through distributed storage systems to support intelligent decision-making and automation control.

Enterprise applications: Distributed storage systems can be used for internal data storage and management to meet the requirements of enterprises for data security, reliability and performance. For example, Ceph enterprise storage solutions support block storage, file storage and object storage, which can meet the diverse storage needs of enterprises.

4 Design of a blockchain-based distributed network storage system

4.1 System architecture design

The architecture design of blockchain-based distributed network storage system needs to take into account the decentralization and tampering characteristics of blockchain technology and the high availability, scalability and flexibility of distributed storage. The system architecture mainly includes the following key modules, and the structure is shown in Figure 1.

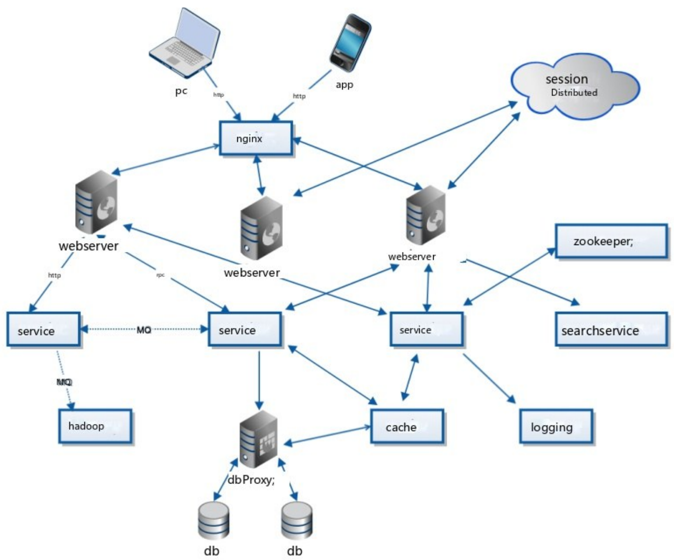

Figure 1 Architecture of blockchain-based distributed network storage system

Node management module: responsible for node join, exit, authentication and permission management. Each node is not only a data storage node, but also a verification node in the blockchain network. Through the consensus mechanism of blockchain, it ensures the trust relationship between nodes and prevents the attack of malicious nodes.

Data slicing and distribution module: the use of consistent hash algorithm for data slicing, and data fragments evenly distributed to the various storage nodes. Consistent hash algorithm can effectively reduce the amount of data migration when nodes are added or deleted, and improve the scalability of the system. At the same time, combined with the smart contract technology of blockchain, the automated management and dynamic adjustment of data slicing can be realized.

Data storage module: each storage node is responsible for storing data fragments and providing data read and write operations. The data storage uses encryption technology to ensure the security of the data in the storage process. In addition, the storage node is also responsible for maintaining copies of the data, through data redundancy technology to improve the fault tolerance of the system.

Data Access Module: Provide a unified data access interface to support users to read and write data. Through the blockchain smart contract technology, it realizes the automated management and access control of data access to ensure the security and privacy of data.

Blockchain Module: As the trust core of the system, the blockchain module is responsible for recording and verifying the transaction information of all data operations, ensuring the non-tampering and traceability of the data. The blockchain module adopts appropriate consensus mechanism, such as PBFT, to ensure the performance and security of the system.

4.2 Data storage and management mechanisms

Data storage and management mechanism is the core part of distributed network storage system, which directly affects the performance and reliability of the system. The blockchain-based distributed network storage system adopts the following data storage and management mechanisms.

Data encryption and hash algorithm: Before data storage, asymmetric encryption algorithm is used to encrypt the data to ensure the privacy and security of the data in the storage and transmission process. At the same time, a hash value is generated for each data segment, and the hash value is stored on the blockchain to ensure the integrity and consistency of the data through the uniqueness and non-tampering of the hash value.

Data slicing and indexing structure: large files are split into multiple small segments, and indexing information is generated for each segment. The index information is stored on the blockchain, and the automated management and query operation of the index is realized through smart contracts. This mechanism not only improves the data storage efficiency, but also accelerates the data retrieval speed and improves the overall performance of the system.

Data replica and fault tolerance mechanism: using data replica strategy, each data fragment to store multiple copies, distributed in different storage nodes. When a node fails, the system can recover the data through other replica nodes to ensure data availability and reliability. In addition, combined with the tamper-proof characteristics of the blockchain, it can effectively prevent the data from being maliciously tampered with.

Data consistency maintenance: Through the consensus mechanism of blockchain and smart contract technology, data consistency maintenance between multiple nodes is realized. When the data is updated, the consensus mechanism ensures that the data of all replica nodes are synchronously updated to ensure data consistency. At the same time, appropriate caching strategy and load balancing technology are adopted to improve the system's read/write performance and data retrieval efficiency.

5 Case studies

5.1 Introduction to typical application scenarios

Distributed network storage systems based on blockchain technology have a wide range of application prospects in many fields, and the following are several typical application scenariosFigure 2.

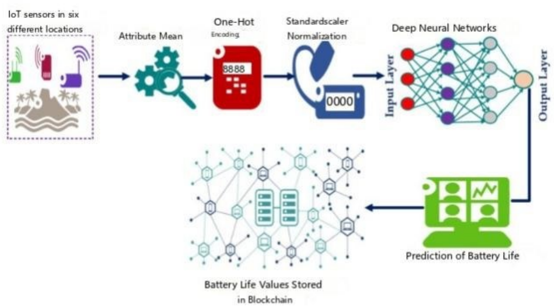

Figure 2 Typical application scenarios

Cloud Storage Service: Storj is a decentralized cloud storage platform based on blockchain. It divides and encrypts user data and stores it on multiple nodes, so that only users with the correct key can access and recover the data. This storage method improves data security, reduces storage costs, and realizes automated management and billing of data access through smart contract technology.

Big data storage and analysis: In the financial field, banks and financial institutions need to store and analyze large amounts of transaction data and customer information. Blockchain-based distributed storage systems can securely store these data and ensure the authenticity and integrity of the data through tamper-proof features. Smart contract technology can also automate the analysis and processing of data, improving efficiency and accuracy.

IoT data storage: In the smart home system, the data generated by various smart devices can be stored in multiple nodes through distributed storage system to ensure data security and privacy. The smart contract technology of blockchain can realize automated interaction and data sharing between devices and improve the intelligence level of IoT system.

Enterprise-class data storage: IBM's Hyperledger Fabric project is an enterprise-class distributed storage platform based on blockchain. It introduces smart contracts, consensus mechanisms and other technologies to provide an efficient and secure distributed storage solution for enterprise-level applications. Enterprises can use the platform to store and manage various business data, ensuring data security and consistency.

5.2 Evaluation and analysis of the effectiveness of the application

Improved data security: In cloud storage services, blockchain-based decentralized cloud storage platform can effectively prevent data tampering and leakage through data encryption and distributed storage. In big data storage and analysis, blockchain's non-tampering characteristics ensure the authenticity and integrity of data. In IoT data storage, blockchain technology protects user privacy and prevents illegal access and tampering of device data. In enterprise-level data storage, blockchain's encryption technology and consensus mechanism ensures the security and consistency of enterprise data and reduces the risk of data leakage and tampering.

Enhanced data consistency and reliability: Blockchain-based distributed storage system effectively solves the data consistency problem in distributed storage through consensus mechanism and smart contract technology. In cloud storage services, the consensus mechanism of blockchain ensures data synchronization and consistency among multiple nodes. In big data storage and analysis, smart contract technology realizes automated data management and consistency maintenance. In IoT data storage, blockchain's non-tampering characteristics ensure the consistency and credibility of device data. In enterprise-level data storage, the consensus mechanism and data redundancy technology of blockchain improves the fault-tolerance of the system and ensures the reliability and consistency of enterprise data.

System performance optimization: Although blockchain technology introduces a certain performance overhead, the blockchain-based distributed storage system performs well in terms of performance through reasonable optimization strategies. In cloud storage services, by adopting appropriate data sharding and caching strategies, the system can realize efficient read/write operations and data retrieval. In big data storage and analysis, smart contract technology can realize automated data processing and analysis and improve the overall performance of the system. In IoT data storage, by optimizing the consensus mechanism of the blockchain and the efficiency of smart contract execution, the system can quickly respond to the read and write requests of device data. In enterprise-level data storage, through load balancing and network optimization strategies, the system can achieve efficient storage and access performance.

Cost-benefit analysis: From the cost-benefit point of view, blockchain-based distributed storage system has advantages in many aspects. In cloud storage services, decentralized storage reduces storage costs, while smart contract technology realizes automated billing and management and improves operational efficiency. In big data storage and analysis, blockchain technology reduces the cost of data management and security protection, and improves the efficiency and value of data analysis. In IoT data storage, blockchain technology reduces the cost of device data management and improves the intelligence of IoT systems. In enterprise-level data storage, blockchain technology improves the security and reliability of data storage and management, and reduces the operational risks and costs of enterprises.

Improved user experience: Blockchain-based distributed storage systems have also achieved significant improvements in user experience. In cloud storage services, users can securely store and access data through a decentralized platform without worrying about data leakage and tampering. In big data storage and analysis, users can obtain more accurate and credible data analysis results, improving the scientific and effective decision-making. In IoT data storage, users can manage and share device data more conveniently, which improves the user experience of IoT system. In enterprise-level data storage, users can store and manage enterprise data more securely, which improves the operational efficiency and competitiveness of enterprises.

6 Concluding remarks

With the arrival of the digital era, the demand for data storage is growing explosively, and the traditional storage system faces many challenges in security, privacy protection and reliability. Distributed network storage system based on blockchain technology provides new ideas and methods to solve these problems by virtue of its decentralized, tamper-proof and highly transparent features. However, the system still faces many challenges in terms of performance bottleneck, security and privacy, technology compatibility, law and regulation. Nevertheless, its future development prospects are still broad, and the trends of technological innovation and optimization, application expansion and integration, standardization and normalization, as well as the adaptive adjustment of laws and regulations will inject a strong impetus to its development. In the future, with the continuous maturity of blockchain technology and the continuous expansion of application scenarios, the blockchain-based distributed network storage system is expected to play an important role in more fields and provide solid technical support for the development of digital economy.

References

[1] Li Man,Zhou Huachun,Qin Yajuan.Two-Stage Intelligent Model for Detecting Malicious DDoS Behavior[J]. Sensors.2022:66-68.

[2] Xie Bin, Dong Xinyu, Wang Changguang. An Improved K-Means Clustering Intrusion Detection Algorithm for Wireless Networks Based on Federated Learning[J].WIRELESS COMMUNICATIONS & MOBILE COMPUTING.2021:23-29.

[3] Papadopoulos Pavlos, Abramson Will, Hall Adam J., Pitropakis Nikolaos. Buchanan William J. Privacy and Trust Redefined in Federated Machine Learning[J].Machine Learning and Knowledge Extraction.2021:235-236.

[4] Kairouz Peter; McMahan H. Brendan. Avent Brendan; Bellet Aurelien; Bennis Mehdi; Bhagoji Arjun Nitin; Bonawitz Kallista; Charles Zachary; Cormode Graham; Cummings Rachel; D'Oliveira Rafael G. L.; Eichner Hubert; El Rouayheb Salim; Evans David; Gardner Josh; Garrett Zachary; Gascon Adria;Ghazi Badih;Gibbons Phillip B.;Gruteser Marco;Harchaoui Zaid;He Chaoyang;He Lie; Huo Zhouyuan; Hutchinson Ben; Hsu Justin; Jaggi Martin; Javidi Tara; Joshi Gauri; Khodak Mikhail; Konecny Jakub; Korolova Aleksandra; Koushanfar Farinaz; Koyejo Sanmi; Lepoint Tancrede; Liu Yang; Mittal Prateek; Mohri Mehryar; Nock Richard; Ozgur Ayfer; Pagh Rasmus; Qi Hang; Ramage Daniel; Raskar Ramesh; Raykova Mariana; Song Dawn; Song Weikang; Stich Sebastian U.; Sun Ziteng; Suresh Ananda Theertha; Tramer Florian; Vepakomma Praneeth; Wang Jianyu; Xiong Li; Xu Zheng; Yang Qiang; Yu Felix X.; Yu Han; Zhao Sen. Advances and Open Problems in Federated Learning[J]. FOUNDATIONS AND TRENDS IN MACHINE LEARNING,2021(1-2).17-19.

[5] Junjun Guo, Le Wang. Learning to upgrade internet information security and protection strategy in big data era[J]. Computer Communications.2023.23-29.

[6] Lu Yunlong,Huang Xiaohong,Dai Yueyue,Maharjan Sabita,Zhang Yan.Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT[J].IEEE Transactions on Industrial Informatics.2021:564-569.

[7] Xiaojun Chen, Shengbin Jia, Yang Xiang. A review: Knowledge reasoning over knowledge graph[J]. Expert Systems With Applications. 2022: 89-90.

[8] Haoye Chai, Supeng Leng, Yijin Chen; Ke Zhang. A Hierarchical Blockchain-Enabled Federated Learning Algorithm for Knowledge Sharing in Internet of Vehicles[J].IEEE Transactions on Intelligent Transportation Systems.2023:45-46.

[9] Shiva Raj Pokhrel, Jinho Choi. Federated Learning with Blockchain for Autonomous Vehicles: Analysis and Design Challenges[J]. IEEE Transactions on Communications.2024:33-39.

[10] Ruijie Zhao, Yue Yin, Yong Shi, Zhi Xue.Intelligent intrusion detection based on federated learning aided long short-term memory[J]. Physical Communication.2024.:77-78.