Blockchain-based Decentralized Storage Scheme

IOP Conf. Series: Journal of Physics: Conf. Series **1237** (2019) 042008 doi:10.1088/1742-6596/1237/4/042008

Blockchain-based Decentralized Storage Scheme

Yan Zhu[1], Chunli Lv[1], Zichuan Zeng[1], Jingfu Wang[1], Bei Pei[2]

[1]College of Information and Electrical Engineering, China Agriculture University, 100083 Beijing, China

[2]Third Research Institute of Ministry of Public Security, Shanghai 210204, China

lvcl@cau.edu.cn

Abstract. Different from the current cloud storage solutions, which are mostly centralized storage providers, this paper proposes a decentralized storage system based on blockchain technology, which can make full use of the remaining space of personal hard disks of users around the world. Storage provider performs a data integrity certificate to the user, and after verifying that the verification is passed, the user pays the storage fee to the storage provider through the lightning network technology. All proofs and payment information are stored in the blockchain, which guarantees the security and credibility of the system. Compared with the current mainstream distributed storage systems, this scheme has been improved in terms of system access and payment methods.

1. Introduction

In recent years, with the continuous improvement of information technology, people's demand for computing and resource storage has also shown a rapid growth trend. People are constantly exploring new ways of computing to satisfy the pursuit of higher computing power and larger storage space. After the emergence of computing models such as peer-to-peer computing, grid computing, utility computing, and distributed computing, cloud computing has become a focus of academic and industrial circles[1]. Cloud computing distributes computing tasks across resource pools of a large number of computers, enabling various application systems to acquire computing power, storage space, and software services as needed[2].

Since 2008, Nakamoto published "Bitcoin: a peer-to-peer electronic cash system"[3], proposed blockchain technology and implemented and applied it in Bitcoin. Blockchain technology has been receiving attention from academic circles. Blockchain has the characteristics of decentralization, non-tampering, unforgeable, privacy protection, and automatic execution of intelligent contracts, which makes blockchain technology have a wide range of application scenarios, such as: digital currency, data storage, voting elections, product traceability, financial transactions and other scenarios[4].

Cloud storage systems enable users to access massive amounts of storage at a cheaper price. At present, most of the cloud storage companies at home and abroad still provide their own centralized storage space, and do not meet the requirements for storage resource integration and distributed storage. Paper[5] compares and analyzes the system performance of centralized storage scheme, multi-center storage scheme based on genetic algorithm, network delay and security analysis of distributed P2P storage scheme based on blockchain, and draws conclusions: based on blockchain, the p2p storage scheme has significant improvements in system performance over the two conventional storage schemes described above. Distributed storage systems are an important research area of p2p systems[6].

Nowadays, distributed storage systems have become the key issues in the application technology of blockchains at academic circles. Filecoin[7] based on IPFS[8] proposes a completely decentralized distributed storage network where customers and storage miners request services and submit orders to the storage and retrieval markets. And the miner provides a service to view matching quotes to initiate a transaction. The protocol guarantees the integrity of data storage by copying proofs and space-time certificates. The Filecoin protocol writes the order book, token transactions, and integrity challenge response records to the blockchain. Siacoin[9] is a simple decentralized storage platform that uses smart contract technology to create a document intelligence contract between the storage provider and the customer. The contract requires that the storage provider provides the customer with a data storage certificate during the certification window. If the proof is legal, the smart contract will automatically pay the customer's provider the negotiated storage fee. Storj[10-11] is a p2p cloud storage network. Customers send encrypt data to the network. File block encryption protects the confidentiality of files. Data direct encryption is a secure and efficient way to protect data confidentiality in distributed storage[12]. The storage provider provides proof of data recoverability, and after the verification is passed, the customer pays the storage provider a storage fee. However, literature[13] proposed the current bitcoin transaction delay problem. Their analysis pointed out that 43% of transactions can be written to the block after more than one hour of release. Distributed storage solutions with transaction latency problems do not have a competitive advantage in practical applications. At the same time, it also shows the importance of updating the system agreement and upgrading the middleman to the vitality of the system. Also, the block generated by the new protocol node is invalid in the node, which means that it will inevitably introduce a hard fork. Ethereum[14] has experienced a hard fork process and splits into two numbers: ETH and ETC Currency[15]. The decentralized system brings the advantages of high reliability and high security, but at the same time brings problems such as slow update and difficult maintenance.

This paper proposes a distributed storage scheme based on blockchain. The user uploads the encrypted data to the middleman, and the middleman sends the data to the storage provider and informs the user of the data storage location. After the data integrity certificate is completed between the user and the storage provider, the user uses the lightning network technology to pay the storage fee to the storage provider.

2. Cryptography in the scheme

In this section, the cryptographic techniques we use including blockchain, lightning network technology, and remote data integrity proof technology for a brief introduction.

2.1. Blockchain

Blockchain is a characteristic data structure formed by combining data blocks in a chain order in chronological order[16], and cryptographically guarantees decentralized, non-tamperable, unforgeable distributed shared ledger system.

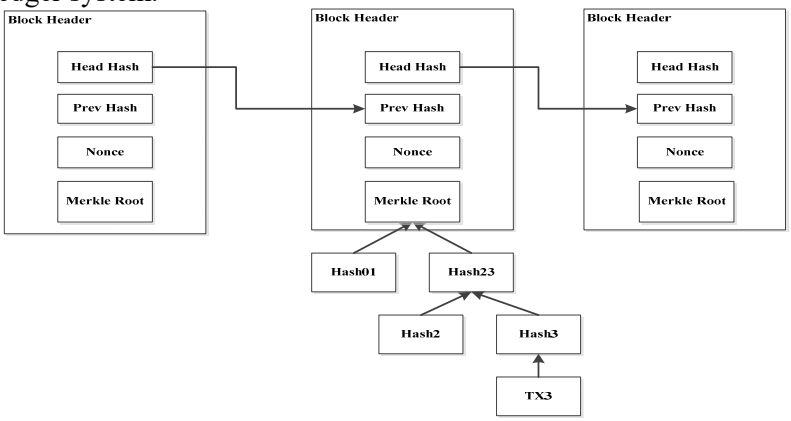

Fig.1 Blockchain Structure

The process of verification, accounting, storage, maintenance and transmission of the blockchain is based on the distributed system structure, instead of adopting a centralized mechanism to build the trust relationship between the nodes, thus forming the characteristics of decentralization. The underlying data layer of the blockchain is supported by techniques such as hashing, asymmetric encryption [17], Merkle tree [18], and timestamp. The basic structure model of blockchain technology is shown in Figure 1.

2.2. Lightning Network

The Lightning Network [19] is mainly composed of two contracts: Sequence Expiration Revocable (RSMC) and Hash Time Locked Contract (HTLC). The expiration of the sequence can solve the problem of fast two-way transfer between the two sides of the channel. The hash time locking contract solves the problem of transfer between nodes. These two types of trading portfolios form the lightning network. The payment method in this system is mainly one-way payment between two users, so the system mainly adopts the sequence expiration revocable contract to realize unlimited fast offline transfer between two users.

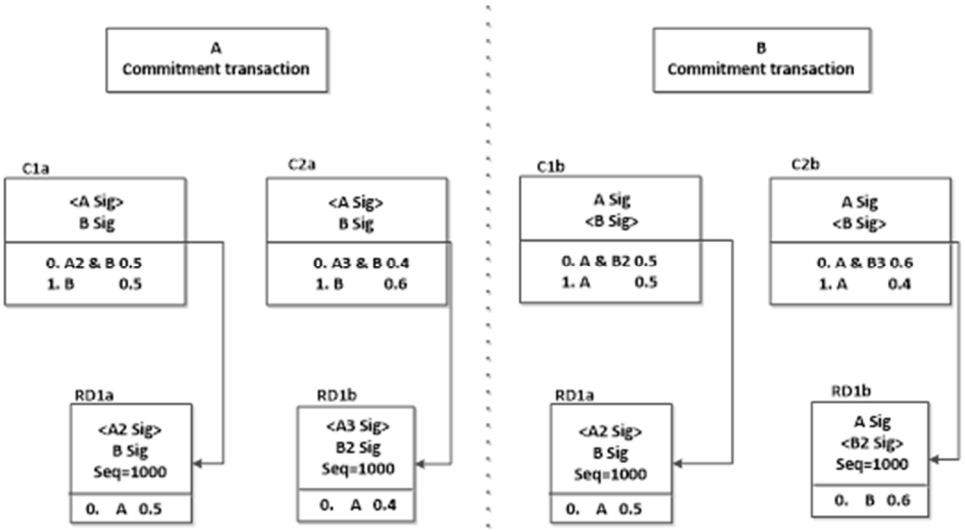

Fig.2 Lightning network transaction update

2.3. Data Integrity

The blockchain-based distributed storage system uses a Merkel tree-based data integrity certification scheme. After the user encrypts the file data, the user generates and saves a random challenge: These random challenges correspond to the block data one by one. The block data and the random challenge are hashed together to become the Merkel leaf node, and constructed into a Merkel tree. The user saves the node information of the Merkel tree leaf and the height of the Merkel tree. During the verification phase, the user randomly selects a challenge in a random challenge: and sends it to the middleman. The middleman sends the challenge to the storage provider.

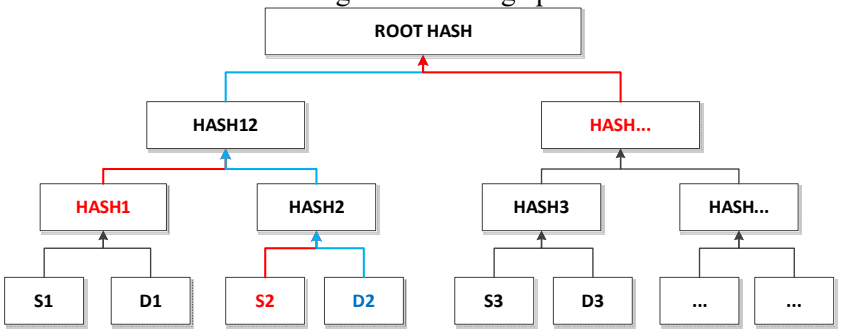

Fig.3 Merkle tree proof mechanism

3. Scheme design

As shown in Fig.4, a blockchain-based distributed storage system includes a user, an intermediary, and a storage provider. The system flow is mainly composed of Register, Setup, Challenge-proof and Pay phases. The Register phase is the system middleman's access phase to the user and the storage provider. The interaction information is stored in the blockchain. The Setup phase is distributed to the system in the early stage. The user encrypts the file data and hands it to the middleman. The middleman encrypts the file. The data block is distributed to the storage provider; the Challenge-proof phase is the storage provider's remote integrity certification phase for storing data, the user issues a data integrity challenge to the storage provider, and the storage provider proves. If it proves successful, enter the Pay phase. If the proof is unsuccessful, the user will no longer pay for the storage. The following is an introduction to the system flow at each stage.

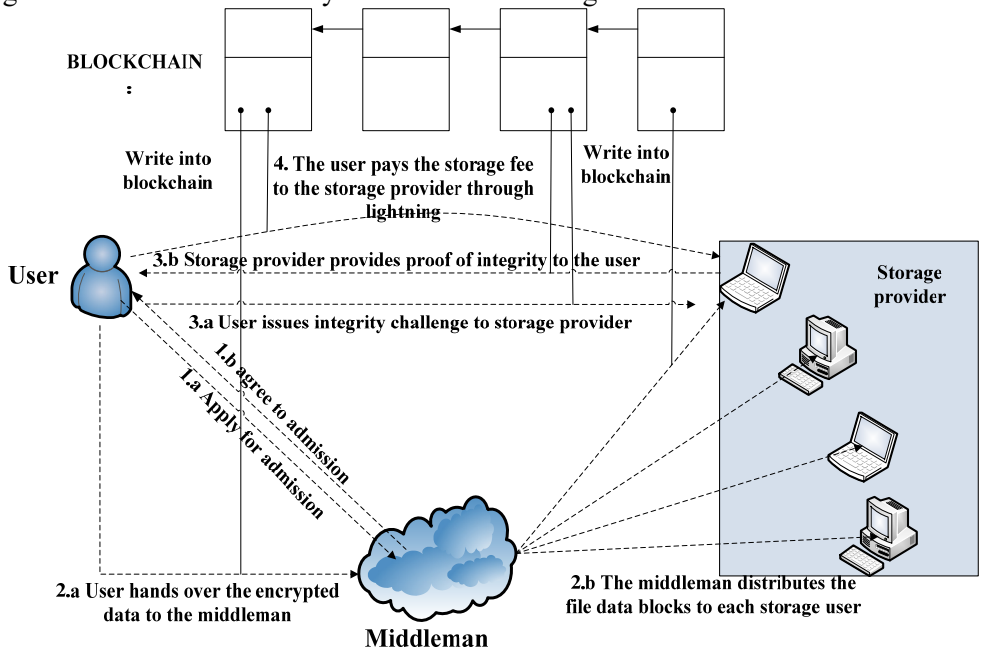

Fig.4 Architecture of distributed storage system based on Blockchain

3.1. Register Stage

1. The user issues a system access application to the intermediary, and the user provides basic information to the intermediary.

2. The system intermediary checks the user's qualifications.

3. The audit result is written into the blockchain. After the network node is confirmed, if the audit is passed, the intermediary issues the access certificate to the user. If the audit fails, the access certificate will not be issued.

3.2. Register Stage

1. The user sends the admission certificate and the storage request to the intermediary.

2. After the middleman confirms the identity of the user, the user encrypts the file by its own private key and sends it to the middleman.

3. After the user encrypts the data, a series of challenge factors are randomly generated and the challenge factors are saved. The user hashes the challenge factor with the data: and constructs the file Merkel tree as shown in Fig.3.

4. The user saves the file Merkel tree locally, and sends the block file data to the file intermediary. The intermediary distributes the file data to the storage provider, and informs the user of the address of the file data after the distribution is completed. A lightning network payment channel is established between the user and the storage provider to prepare for the pay phase.

3.3. Challenge-proof Stage

1. The user randomly selects a challenge factor randomly among them and sends it to the storage provider.
2. The storage provider produces a data integrity certificate.
3. The storage provider sends evidence of data integrity to the user.
4. The user verifies that the proof of completeness is correct.

3.4. Pay Stage

In the Setup phase, a sequence expired revocable contract channel has been established between the user and the storage provider. When the Challenge Phase Data Integrity Certificate is verified, the user pays the storage provider a storage fee.

4. Analysis

4.1. Centralization of Middleman Problem

In order to maintain the vitality of the system and increase the access mechanism of the system users, this program introduces the role of system middleman. But the core feature of the blockchain is decentralization. The centralization of system intermediaries is a problem that this program needs to consider. In this system, the intermediary is responsible for system access, protocol upgrade, and data distribution. The user's access interaction with the system middleman and the intermediary's interaction with the user data block need to be recorded in the blockchain, and can be written into the blockchain after the whole network node confirms. The irrevocable nature of the blockchain will make these interactive information always preserved after being written into the block, and will be open to the entire network and subject to network-wide supervision. In all, there will be no middleman centralization in the system.

4.2. Comparison of distributed storage schemes based on blockchain

At present, the mainstream blockchain-based distributed storage solutions include Filecoin, Sia, and Storj. As shown in Table 1, the data scheme is compared with the three storage protocols from six aspects: data integrity verification, data confidentiality protection, system maintenance, system access, fee payment method, and storage provider.

Tab.1 Comparison of distributed storage protocols based on blockchain

Storage Scheme	Data Integrity	Data Confidentially	System Maintenance	System Acces	Fee Payment method	Avoid Provider cheating
File coin	√	×	×	×	One-time Payment	Pay the Deposit
Sia	√	×	×	×	Payment after proof	Contract
Storj	√	√	×	×	One-time Payment	Contract
This Scheme	√	√	√	√	Payment after proof	Lost Storage Fee

5. Conslusion

This paper proposes a distributed storage scheme based on blockchain technology, and introduces the system design in detail. The system uses cryptography techniques such as blockchain technology, lightning network technology, remote data integrity certification, and remote data confidentiality protection technology. Make full use of the user's remaining disk space, solve the waste of resources,

and encourage nodes to join the cloud storage service. The system proposes a storage protocol with middleman, which can realize the system update and upgrade, system access and other functions. In the future, the author will focus on the implementation of the model in the Ethereum test network, while optimizing and upgrading the above key technologies.

Acknowledgments
This research was financially supported by the National Science Foundation (61202479) and Information Network Security Key Laboratory Development Project of the Ministry of Public Security (C16611).

References

[1] Armbrust M, Fox A, Joseph A D, et al.Above the clouds: A Berkeley view of cloud computing.EECS Department, University of California, California, USA: Technical Report UCB/EECS-2009-28, 2009.
[2] Liu P. Definition and characteristics of cloud computing. Cloud Computing in China. 2009 (in Chinese).
[3] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Consulted, 2008..
[4] Liu AD, Du XH, Wang N, Li SZ. Research Progress of Blockchain Technology and its Application in Information Security. Journal of Software,2018,6,14:1-24(in Chinese).
[5] Li J, Wu J, Chen L. Block-secure: Blockchain based scheme for secure P2P cloud storage[J]. Information Sciences, 2018, 465: 219-231(in Chinese).
[6] Xu F, Yang GW, Ju DP. Design of distributed storage system on peer-to-peer structure. Journal of Software, 2004,15(2):268-277. (in Chinese).
[7] Protocol Labs. Filecoin: A Decentralized Storage Network. https://filecoin.io/ filecoin.pdf, 2017.
[8] Benet J. IPFS-content addressed, versioned, P2P file system[J]. arXiv preprint arXiv:1407.3561, 2014.
[9] David Vorick and Luke Champine. Sia: Simple Decentralized Storage. 2014.
[10] Wilkinson S, Lowry J, Boshevski T.Metadisk a blockchain-based decentralized file storage application[J]. Technical Report. 2014.
[11] Shawn Wilkinson et al. Storj: A Peer-to-Peer Cloud Storage Network. 2016.
[12] TIAN Hong-Liang, ZHANG Yong, LI Chao, XING Chun-Xiao, A Survey of Confidentiality Protection for Cloud Databases, 2017,Vol. 40, Online Publishing No. 78. (in Chinese).
[13] Pappalardo G, Di Matteo T, Caldarelli G, et al. Blockchain inefficiency in the Bitcoin peers network[J]. EPJ Data Science, 2018, 7(1): 30. https://arxiv.org/pdf/1704.01414.pdf
[14] Wood G. Ethereum: A secure decentralised generalised transaction ledger[J]. Ethereum project yellow paper, 2014, 151: 1-32.
[15] Wirdum V A . Ethereum Classic Community Navigates a Distinct Path to the Future[OL]. 2017.02.15. Stallings W. Cryptography and network security: principles and practice[M]. Pearson Education India, 2003.
[16] Liu AD, Du XH, Wang N, Li SZ. Research Progress of Blockchain Technology and its Application in Information Security. Ruan Jian Xue Bao/Journal of Software,2018,6,14:1-24.
[17] Merkle R C. Protocols for public key cryptosystems[C]//Security and Privacy, 1980 IEEE Symposium on. IEEE, 1980: 122-122.
[18] Fan J, Yi LT, Shu JW. Research on the technologies of Byzantine system. Ruan Jian Xue Bao/Journal of Software,2013,24(6):1346-1360 (in Chinese).
[19] Joseph Poon and Thaddeus Dryja. The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments.