PROCEEDINGS OF SCIENCE

A Data Tamper Protection Method based on Block Chain

Fuken Zhou[1]
Department of Computer Science, Guangdong Neusoft Institute
Foshan, 528000, China
Guangdong Key Laboratory of Big Data Analysis and Processing
Guangzhou, 510006, China
*Email:*zhoufuken@nuit.edu.cn

Zepeng Yu
Department of Computer Science, Guangdong Neusoft Institute
Foshan, 528000, China
*Email:*yuzepeng@nuit.edu.cn

Yongjie Xie
Department of Computer Science, Guangdong Neusoft Institute
Foshan, 528000, China
*Email:*Xieyongjie@nuit.edu.cn

Tian'en Chen
Department of Computer Science, Guangdong Neusoft Institute
Foshan, 528000, China
*Email:*ChenTen@nuit.edu.cn

Yongfa Liao
Department of Computer Science, Guangdong Neusoft Institute
Foshan, 528000, China
*Email:*Liaoyongfa@nuit.edu.cn

Data centric storage with network as the carrier is vulnerable to external attacks, resulting in data loss and tampering. We propose a method of tamper proofing for block chaining data without centralization and centralization. According to the technology principle of bitcoin block chain, the network data information is protected by block chain encryption technology, and the security of high information tamper resistant is realized. A block chain data storage algorithm based on hash table is hereby proposed, which can get enough address space to save all the data, and make block chain form a time sequence to achieve the safe transmission of data. The research results show that the data protection method has very high tamper resistance ability because of the high security of encrypted digital block chaining technology. The block chain implements the value transfer without relying on centralized organization.

ISCC2017
16-17 December 2017
Guangzhou, China

[1] Acknowledgements：This work by the Foshan Science and Technology Innovation Project (2016AG100792).

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(ISCC 2017)052

1.Introduction

With the rapid development of computer network technology, computer data storage will inevitably encounter some safety problems, such as natural disasters, network hardware and software, human error, etc., these security issues are likely to bring immeasurable losses to enterprises. How to protect data security has become the primary problem to be solved. The above problems have their effective solutions, such as offsite backup, firewall, data encryption, rights management and personnel training etc..

At present, the common data protection technology mainly has the following several kinds:

Redundant Array of Inexpensive Disks (RAID)

RAID is made up of a hard disk controller to control multiple types, capacity, interface and even brand consistent mutual connection, dedicated hard disk or ordinary hard disk to synchronize multiple hard disk read and write, reduce errors, increase the efficiency and reliability of the technology in a rapid, accurate and safe way of reading and writing disk data so as to increase the speed of data reading and security.

Peer-to-peer Remote Copy (PPRC)

Memo ry level data replication technology, where the data of the local disk and the remote mirroring disk are synchronized at real time and simultaneously in the online state. When the production disk fails, it can be switched to the mirror disk immediately. The remote disk mirroring has two modes, synchronous mode and asynchronous mode. The former can ensure good data consistency, but affects performance. The latter is the opposite, and the asynchronous mode is usually used.

Snapshot

Grab all the data on a disk at a certain moment, that is, the camera presses the shutter to leave the image. The source of backup, used to solve some logic fault, such as system crashes, mis-operation etc.. Snapshots can be done instantaneously, leaving only the shadow of the original volume.

The block chain cannot be tampered with features is supported by distributed storage and consensus mechanism. This feature can be used to solve the problem of data tamper proofing. At present, this article, only for two cases, prevents data tampering: (1) the intruder skips the monitoring system to manipulate the data directly in the database; (2) illegal users enter the system by using the system to modify the data.

2.Technology of Block Chain

The block chain structure shown in Fig. 1:

we use the four layers: data layer, network layer, consensus layer and application layer of the block chain to construct a data protection technology system based on te block chain by using private chain. The data layer mainly includes data block, Merkle Tree, timestamp, hash function; network layer includes distributed algorithm, encrypted signature and peer-to-peer network; the consensus layer includes consensus algorithm, a few obey the majority mechanism.

PoS(ISCC 2017)052

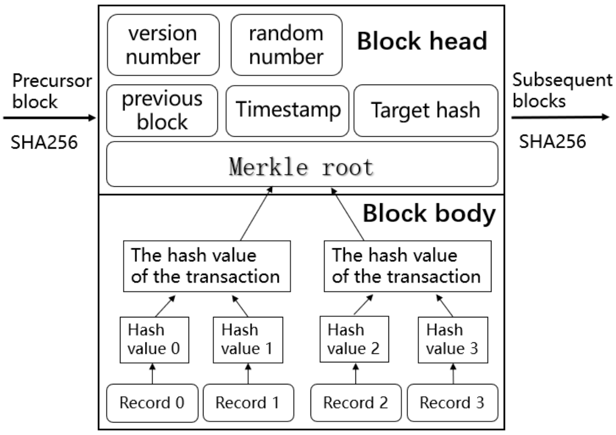


Figure 1: Data Structure of Block Chain

2.1 Data Layer

The data layer of the block chain is a chain structure in which data blocks are encapsulated, associated asymmetric data encryption techniques and timestamps, etc., which belong to the lowest data structure of the entire block chain technology. Most of the technology can be used for so long enough to prove its its safety and reliability, and the block chain puts these technologies together in a smart manner.

2.2 Network Layer

The network layer of the block chain includes P2P networking mechanism, data propagation mechanism and data authentication mechanism. P2P network is different from the network structure composed of C/S mode. In the peer-to-peer network of P2P, the status of each node is equal, so it is also called peer-to-peer network, and the client can communicate directly with the client. P2P network is divided into two kinds of structured and unstructured, structured P2P network by using a consistent hash table to build each node's routing table, unstructured P2P network routing between nodes in the way of broadcast, each Nodes to read and send data to their neighbor nodes for data transfer and broadcast in the network.

2.3 Consensus Layer

The block chain is a decentralized distributed database, so the records need to be distributed to different regions of the network nodes to participate in the decision-making. The decision-making rule is the core of the consensus mechanism. The consensus mechanism determines which one of the most recent data blocks is to ensure that the information exchange can be done without error. The consensus mechanisms need to weigh the relationship between safety and efficiency. The more complex security measures, the processing speed will be slower. In order to improve the processing speed, efforts must be made to simplify the complexity of security measures. At present, the consensus mechanism of block chain technology is: Pow, PoS, DPoS, Pool Mechanism and PBFT Mechanism, etc..

3. Algorithm Design

3.1 Build Block Chain

With reference to the Bitcoin block chain technology principle, this paper designs the block chain as follows:

(1) Verification: the so-called block chain verification is a pseudo-verification method, that is, a public key can generate a number of private keys, which can be achieved by assigning private keys. So the system also uses this approach to establish a composite validation model. The model can share a key to multiple users, who have the right to access the block chain, but others who do not have a key assigned are not authorized to access it. Supposing that in a data transmission, use A and B respectively to represent both sides of the transmission.

(2) Assuming that the block chain storage space in the system be L, the 256 hash algorithm can be used to obtain enough address space to save all data, which is consistent with the design of the bitcoin block. As each time the data transmission, the block chain can be seen as a time series, each of which contains the last transmission address (about 160), so after the data hash, the first two bits are used to identify the data type, followed by all locations used to store data addresses. L [k] indicates the block of the current most recent transaction.

3.2 Data Structure Design

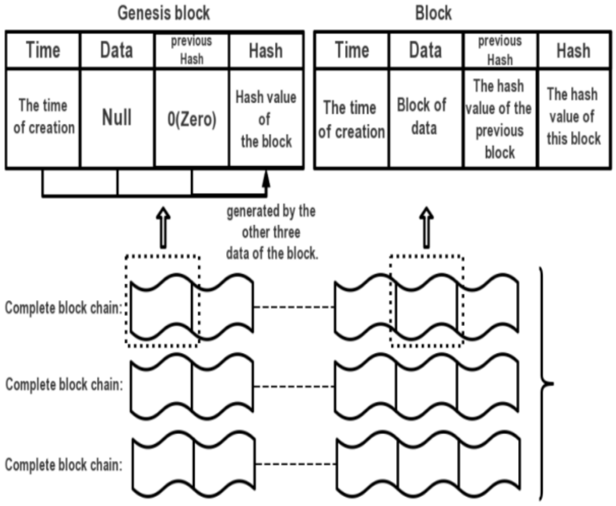


Figure 2: The Block Diagram of the Block Formation

As shown in the Fig., each record of the data after each change will generate a block, and the record is converted to Json structure stored in the block data; and the hash value of this block is generated by the other three data of the block. And then calculate the hash value of a Merck tree root hash value, each time before adding the block, firstly check the root of the hash value; after verification, continue to add data. Each block is responsible for a record, and a block chain is responsible for an entire database. The intermittent inspection of the root hash value can determine whether the data has been tampered with, once tampered with, then issued an alarm.

3.3 Block Chaining

In the structured network, this paper uses DHT's distributed hash algorithm to solve the problem of structured distributed storage.

DHT (Distributed Hash Table) is a way of distributed storage, DHT network allows each client to be responsible for a routing and storage of a small part of the data, not to use the server can achieve the entire network addressing and storage.

Based on the Chord version of the DHT implementation: the use of consistent hash as a hash algorithm, each node also needs to store n other nodes of information, the set of these information is called the finger table .The nodes in the consistency hash also have such a table, but in Chord, the nodes in the table are no longer directly adjacent nodes, and their spacing (the

PoS(ISCC 2017)052

interval of the ID) will be arranged in a 2i relation (i refers to the table in the array subscript). The routing relationship between nodes thus formed is actually the permutation relationship required by the binary search algorithm. The number of hops required for the query is reduced from O (N) to O (log (N)). Even in large-scale P2P networks (for example, N = 100,000,000), the number of hops is only O (8), and each node needs to store only 27 other nodes.

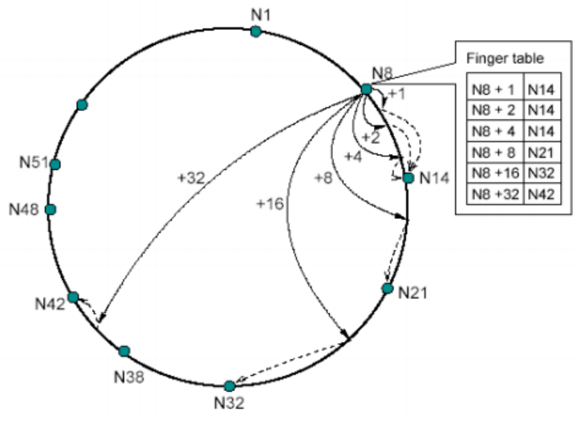


Figure 3: Chord's Model Maps

3.4 Distributed Consistency Algorithm

The core problem of distributed networks is how to reach a consensus efficiently, just as the existing social system. A high degree of centralization and a centralized decision-making society are more likely to reach consensus, like dictatorship and autocracy, but the social satisfaction is low; a society with a low degree of centralization and a decentralized decision-making is more difficult to reach a consensus, like a democratic vote, but the whole society is more satisfied.

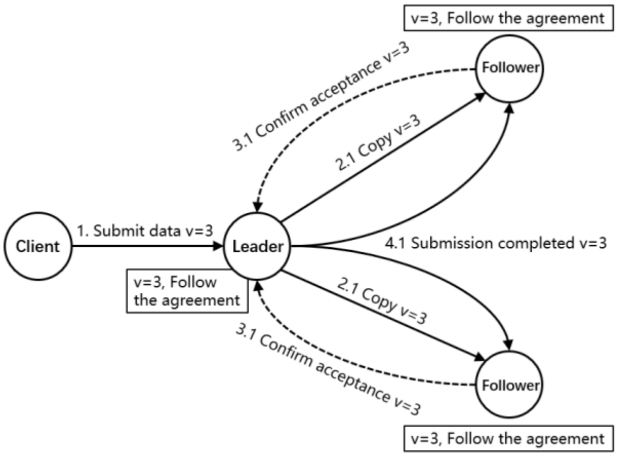
The Raft algorithm implementation process is similar to the election, and the candidate needs to convince the voter (the server) to vote for him and, once selected, follow the operation. In the Raft algorithm, any server can play one of the roles:

(1)Leader: process all client interaction, log copy, etc., generally only one Leader at a time.

(2) Follower: similar to voters, is completely passive.

(3) Candidate: similar to the proposer lawyer, can be chosen as a new leader.

The raft's stage is divided into two steps: one is the election; the second is to lead the followers to normal operation. Any one of the servers can be a candidate, the candidate to the other Follower election vote to vote, waiting for the vote to respond, as long as the majority of N / 2 +1 votes, candidates can become the Leader, who can then issue instructions to the follower for normal operation.


v=3, Follow the agreement
Follower
3.1 Confirm acceptance v=3
2.1 Copy v=3
Client
1. Submit data v=3
Leader
4.1 Submission completed v=3
v=3, Follow the agreement
2.1 Copy v=3
3.1 Confirm acceptance v=3
Follower
v=3, Follow the agreement

PoS(ISCC 2017)052

Figure 4: Raft Implementation Flow Chart

The leader's election process is time-limited and the specific time allocation is shown below:

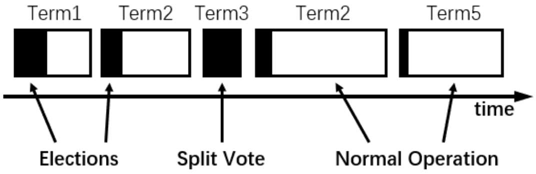


Figure 5: The Time Allocation Map of the Election Process

4 .Application Layer Implementation

This paper combines the characteristics of the block chain to construct a data protection technology system. It mainly includes hash authentication, distributed storage and consensus algorithm, etc. in order to prevent the data from being tampered with and the data recovery function. The data block is combined with a timestamp to generate a hash value as a data validation flag; by using the block chain for distributed storage of data, and in the peer-to-peer network broadcast, the whole network to achieve the block chain data synchronization and consensus verification to prevent data has illegal tampering effect and can further ensure the safety of the data.

The specific flow chart is shown below:

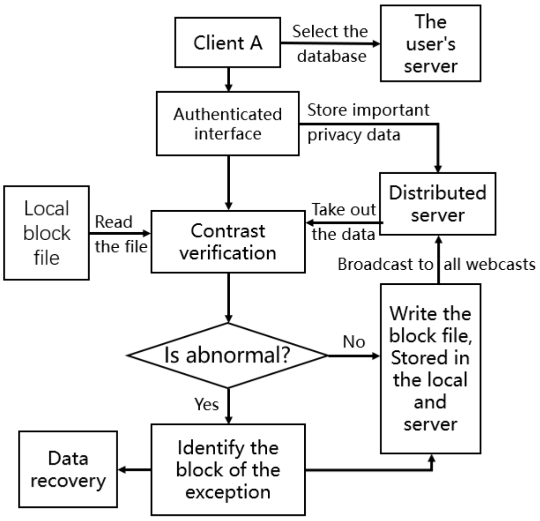


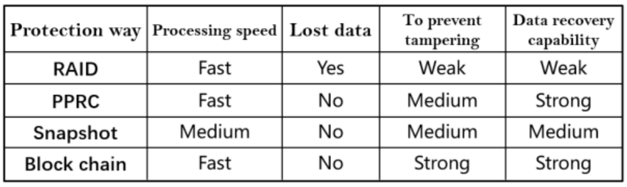
Figure 6: Anti-Tampering Design Flow Chart

Users can choose to protect the sensitive data tables or databases, and call the corresponding interface through the interface to connect the user’s database and our server database, the important data stored in our servers. In this sense, they can protect their data.

To prevent the tampering against the second case: illegal users to enter the system, the use of the system to modify the data:

This change does not seem to be perceived by the system and the user, but as long as the system interface is invoked, the block chain will generate a non-tamperable record after each operation, similar to an unchangeable operation Log, and broadcast the whole network. All the changes in the data will be stored in all peer-to-peer network. When the user once found the data is wrong, such user can view the change records, compare and find out the abnormal data, make changes and recovery through the system platform.

Through data analysis and test of the application layer, the comparison of the various protection techniques for data protection and the prevention of tampering is shown in the following Table 1:



Protection way	Processing speed	Lost data	To prevent tampering	Data recovery capability
RAID	Fast	Yes	Weak	Weak
PPRC	Fast	No	Medium	Strong
Snapshot	Medium	No	Medium	Medium
Block chain	Fast	No	Strong	Strong

Table 1: Technical Analysis Comparison Table

The use of hash encryption, the original content and user personal information can not be seen and cracked by other users, which also protect the privacy of user data, making even the data transfer data are stored in different clients, each user information still can not be Obtained by any client.

Finally, the data in the encryption process, in fact, is not to encrypt the whole multimedia data, but only through the encrypted data fingerprint to identify a data set, so the storage and verification process, not very time-consuming, and The data itself is stored in the database and extracted and modified as needed.

5.Conclusion

We introduce a new data security technology, namely the block chain technology. By using the centralized storage method, the traditional data center concept is replaced by the distributed concept, and the central identification of data security is transformed into the identification of distributed network nodes. This makes the data fingerprint difficult to tamper (unless 51% of the nodes are controlled), protect the security of data, and achieve the value transfer which does not rely on the centralized organization. In addition, each change of data is recorded on the block chain, so it can track any change trajectory of the data, as another part of the block chain technology worth study.

References

[1]Swan M. Blockchain: *Blueprint for a New Economy*[M]. O'Reilly Media, Inc. 2015.

[2]Crosby M, Pattanayak P, Verma S, et al. *BlockChain Technology: Beyond Bitcoin* [M]. Berkeley: Applied Innovation Review,2016.

[3]Wust K. *Security of Blockchain Technologies* [D]. Zurich: Department of Computer Science, 2016.

[4]Q Xia. *Trust-less Medical Data Sharing Among Cloud Service Providers Via Blockchain:IEEE Access*[C],2017

[5]P Jiang,*Blockchain-based private keyword search in decentralized storage: Future Generation Computer Systems*[J], 2017

[6]PK Sharma. *A Distributed Blockchains-Based Secure SDN Architecture for IoT Networks*:IEEE Communications Magazine, 2017

PoS(ISCC 2017)052