2021 International Conference on Intelligent Computing, Automation and Applications (ICAA) | 978-1-6654-3730-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAA53760.2021.00108

Overview of the Basic Principles of Blockchain

Buquan Liu*
Research Institute of Innovation, Kylin Software Company Limited, Tianjin 300450
*Email: liubuquan@kylinos.cn

Abstract—**In the Internet, the authenticity of data between computer nodes is difficult to be guaranteed. However, as the underlying implementation of bitcoin, blockchain adopts decentralized and distributed ledger to allow false transactions and data tampering, and ultimately provides a reliable and trustworthy solution for bitcoin transactions on the unreliable and untrusted Internet. This paper introduces the basic principles of blockchain from the aspects of the relationship between blockchain and bitcoin, two mathematical algorithms involved in blockchain, the logical structure of blockchain, the types of blockchain, and the application of blockchain. This paper explains the core concepts of blockchain, such as decentralization, trustlessness, mining, proof of work, points out several problems that need to be paid attention to in the practical development of blockchain application, and puts forward the approach of integrating blockchain and traditional database to solve the practical application.**

I. Introduction

The price of bitcoin has continued to rise since it appeared in 2009. On March 10, 2021, each bitcoin has reached US $56000 [1]. As the underlying technology of bitcoin, blockchain [2] has been highly concerned by government departments and enterprises, and has been widely studied and applied in financial and non-financial fields. In essence, blockchain is a data chain made up of blocks, which is distributed to multiple computer nodes. These computer nodes are independent and trustless with each other. There is not an authoritative central node among them. Therefore, blockchain is considered as a decentralized and trustless system. Decentralization is the method while trustlessness is the purpose. The latter needs to solve through the former. Many people emphasize the former, but the latter is the fundamental. Trustlessness refers to removing the trust of the third party, and the transaction is conducted directly by both sides without the trust third party. However, the blockchain itself does not require mutual trust between the two sides, and the transaction data is allowed to make fraud and deception, which may lead to multiple blockchains at the bottom of bitcoin. However, through consensus mechanism, reward, fraud punishment and other measures, the blockchain finally establishes a reliable and trustworthy solution for bitcoin. However, in some small applications, consensus mechanism and reward can not fundamentally solve the problem of trustlessness or data fraud. Many blockchain applications currently developed are built on the basis of trust to some extent, which is quite different from bitcoin's blockchain. Even so, some of the systems still solve problems that are difficult to handle by other technologies, and play a good role in reality [3-4]. This paper focuses on the basic principle of bitcoin blockchain, and provides some reasonable suggestions for the development and application of blockchain technology.

II. Blockchain and Bitcoin

A. History of Blockchain

The relationship between blockchain and bitcoin is that bitcoin is the upper application while blockchain is the bottom software. Blockchain has experienced three stages, namely 1.0, 2.0 and 3.0 respectively.

Blockchain 1.0 is an era of virtual currency represented by bitcoin, which realizes the decentralization of currency and means of payment, and a large number of counterfeit currencies emerge. Blockchain 1.0 supports virtual currency transactions between two people, which cannot be popularized to other industries.

Blockchain 2.0 adds the concept of smart contract. A transaction allows multiple people to sign and send it to others. Smart contract provides a wider range of application scenarios in the financial field, such as commodity trades under multi-party contract. The representative of blockchain 2.0 is Ethereum [5], which supports smart contract and provides programming interface. Users can develop their own virtual currency. Blockchain 2.0 is usually used in the financial area.

Blockchain 3.0 is applied to other areas rather than the financial industry. Blockchain 3.0 is known as a new generation of technological innovation after Internet, which is enough to promote greater industrial reform. Because it no longer depends on a third party or an institution to obtain trust and establish credit, the technology can improve the work efficiency of the whole system. The representative of blockchain 3.0 is hyperledger fabric [6], where there is not a service charge in transaction.

978-1-6654-3730-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICAA53760.2021.00108

B. *Value and Acquisition of Bitcoin*

Bitcoin is highly speculative and lacks third-party supervision since its birth. For example, RMB, US dollar and euro all rely on national credit and gold reserves as circulating currencies, but bitcoin's credit is some mathematical algorithms. Because it does not have a clear value, it is considered illegal activities in China.

Bitcoin can be obtained in at least three ways. One is to buy from some trading websites; the other is to transfer through bitcoin software; the third is to mine through bitcoin software. Mining is one of the core concepts of blockchain, which will be gradually described later. This paper does not discuss illegal transactions, but only reveals the basic principle of blockchain.

III. Two Mathematical Algorithms

Blockchain uses two basic mathematical algorithms: asymmetric encryption and hash. Through both algorithms, the data is encrypted and decrypted so that it is difficult to be cracked by a third party.

A. *Asymmetric Encryption*

Symmetric encryption has only one key, which is used for encryption and decryption. For example, "123456" is used as the password to compress a folder, and also for decompressing.

Asymmetric encryption has two different keys: public key and private key. If the public key is used to encrypt the data, only its private key can be used to decrypt. Moreover, if the private key is used to encrypt the data, only its public key can be used to decrypt. Usually, the public key is sent to other people, and the private key is kept by senders; other people can encrypt through the public key, but only use the private key to decrypt, so the security is guaranteed.

B. *Hash*

Modular operation in mathematics is a very simple hash algorithm. For example, the remainder of 1,2,3,4,5,6,710001 divided by 4 is 1,2,3,0,1,2,3,1 respectively. In other words, it can be written as: hash(1) = 1, hash(2) = 2, hash(3) = 3, hash(4) = 0, hash(5) = 1, hash(6) = 2, hash(7) = 3, hash(10001) = 1. It can be referred that different numbers may have the same hash value.

Scientists have created many more complex hash algorithms, such as MD5, SHA1, SHA256. The following table lists the values of each hash algorithm corresponding to the string "Hello World!".

These hash algorithms have the following characteristics.

1) Given a string key, it is easy to calculate the hash value;

2) Given a hash value, it is difficult to deduce the corresponding key.

Theoretically, there may be multiple keys corresponding to the same value, but in fact, the probability is very little. It is approximately thought that key and value are one-to-one correspondence, and value is called data fingerprint. Data fingerprint means that it is hard to fake. A value can only find a matching key.

TABLE 1. HASH VALUES FOR STRING "HELLO WORLD!".

Hash algorithm	Hash value	Bits
MD5	ED076287532E86365E841E92BFC50D8C	128
SHA1	2EF7BDE608CE5404E97D5F042F95F89F1C232871	160
SHA256	7F83B1657FF1FC53B92DC18148A1D65DFC2D4 B1FA3D677284ADDD200126D9069	256

IV. Logical Structure of Blockchain

A. *Link List*

Blockchain is composed of block and chain. Blocks are used to hold data. For bitcoin, one block is generated every 10 minutes to store all transactions within the period. Transaction stores the information of currency ownership from one person to another. Each transaction requires a certain service charge, which is called gasoline in the blockchain.

Generally, data link can be divided into forward link, reverse link and bidirectional link.

1) Forward link. Support searching backward from the head of the link list.

2) Reverse link. It supports the backward search from the tail of the link list.

3) Bidirectional link. It supports two-way search and is an efficient data structure.

Blockchain adopts reverse chain, which is easy to trace from back to front. Compared with several structures of link lists, the following conclusions are drawn.

Proposition 1. High performance is not a consideration of blockchain.

In addition, the first block in the blockchain is called genesis block.

B. Decentralization

As shown in Figure 1, the above is a centralized architecture. The database in node 5 can be shared by the other four nodes, which is easy and efficient to read and modify.

The below is a decentralized P2P architecture, all nodes are peer-to-peer, and each node keeps a complete blockchain. This architecture is essentially a backup mechanism of multiple copy with better reliability.

Proposition 2. Blockchain should not be too large.

For bitcoin, anyone in the world can participate in mining. If the blockchain is too large, ordinary computers may not be able to save the whole blockchain, and many people will not be able to participate. For more than 10 years, the current blockchain is only 300g, which can be saved by ordinary computers.

Proposition 3. Don not update blockchain too fast.

Again, this is the blockchain of bitcoin. A block needs to include almost all transactions over a period of time. Because transactions are transmitted between all nodes of the Internet, if the update is too fast, some transactions may not be packed into the blockchain. In addition, because all nodes can mine, there will be multiple blockchains on the Internet.

C. Mining

Blockchain is composed of multiple blocks. The so-called mining refers to the process of generating a new block.

As shown in Figure 2, a block is composed of block header and block body [7-8].

The block contains all transactions within 10 minutes. Four transactions are shown in the figure. Hash1, hash2, hash3 and hash4 are obtained by computing the hash value of each transaction. Then the hash value every two is computed until hash1234 is gotten, which is Merkel root in the block header. If a transaction is added, deleted or modified in the block, Merkel root will change.

Block header mainly includes six elements: hash(parent), version, Merkle root, timestamp, difficulty target and nonce. The so-called mining is to determine these six values.

The previous block also contains these six elements, whose hash value is recorded as hash(parent). Therefore, if a certain block is forged, it must modify the hash(parent) of all blocks after it. This is certainly a big task.

In order to generate a block every 10 minutes, the difficulty target is set. With the increasing performance of computer, the value is changed dynamically. As shown in the following formula, the system will adjust the difficulty every 2016 blocks according to the mining time of the previous cycle. For a specific block, the value is fixed.

DifficultyTarget = 0x00000000FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF / Difficulty

It can be found that the difficulty of block 674000 is 21448277761059.70. The hash value of the six elements is '00000000000000000007f8d1652e3bf35c9d2672be4834fef6ebc6f1c407bd8c', shown in block 674001 as hash(parent). There are 19 zeros in front of this value. In order to meet this condition, a lot of calculations must be carried out to find out these elements.

Among the six elements, hash(parent), version, Merkle root and difficulty target are all fixed. The timestamp represents the generation time of the block, and nonce is variable. In the distributed system, it is not required that the clock of each computer is exactly the same, and the change of timestamp is of little significance to others. Thus, in order to simplify the mining difficulty, the time stamp can also be determined in advance. In this way, we only need to find the random number 'nonce' in the six elements.

Nonce is usually tried one by one from 0 to see if the hash value of six elements meets the difficulty. If a random number meets the difficulty, it means that the mining is successful. After the new block is generated, it is broadcast to other adjacent nodes.

The nonce of block 674000 is 4275554107, which shows that the workload of mining is very large. In blockchain, such workload is called proof of work (POW).

Those engaged in mining are called miners. Since the first successful miner will be rewarded with bitcoin and receive all the transaction fees, the miners are trying to improve the efficiency of mining.

Proposition 4. High performance is the pursuit factor of miners.

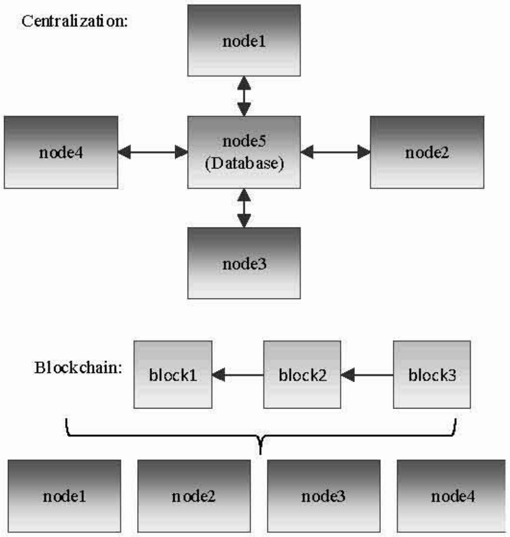

Figure 1. Centralization vs Blockchain.

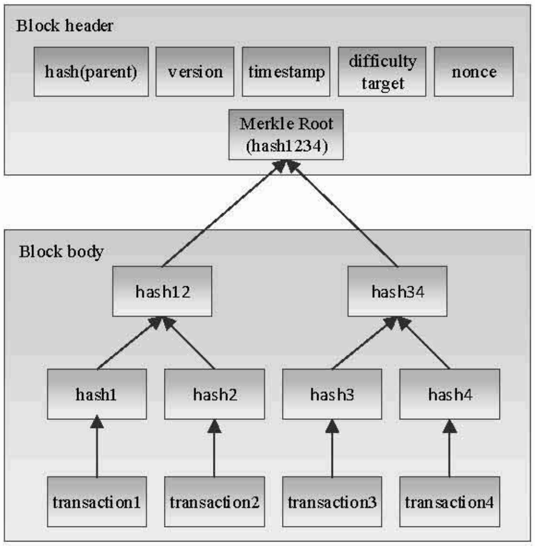

Figure 2. Block Structure.

D. *Consensus Mechanism*

If only one legal block is produced in the same period, and there is no other legal block competing with it, the block will be received. If there is a competitive block, more than 51% of the nodes confirm that the most difficult and longest chain will eventually be received, and other legal blocks that are not received will become isolated blocks and be eliminated. The so-called consensus mechanism means that it has been confirmed by most (more than 51%) nodes.

E. *Trustlessness*

In the Internet, there is no absolute trust between nodes. In fact, if most of the nodes are trustless, the blockchain will have no significance. In addition to decentralization, blockchain at least adopts these measures to prevent counterfeiting.

1) When a node receives a block from another node, it will check and confirm the blockchain through the hash value. If the blockchain is fake or not the longest, it will be discarded.

2) There will be a reward for successful mining and service charges for all transactions in the block. Since Nakamoto created the first block and won 50 bitcoins, the reward will be halved for every 210000 blocks (about four years). When there is no reward, you can still get the service charges.

3) Counterfeiting costs a lot. For the blockchain composed of block 1, block 2,..., block n. If block 1 is faked, it is necessary to recalculate the nonce of the block. As a result, the hash(parent) in block 2 changes so that it is also necessary to recalculate the nonce of block 2. Finally, it is necessary to calculate the nonce of all subsequent blocks. In fact, if the counterfeiter's computing power exceeds 51% of the mining nodes in the whole network, the counterfeiting is successful.

Proposition 5. Blockchain fraud costs a lot.

V. Blockchain Type

Blockchain is divided into three types: public blockchain, private blockchain and consortium blockchain.

A. *Public Blockchain*

Bitcoin's blockchain is a kind of public blockchain, which has the following characteristics.

1) Public, open and decentralized.

2) The data is open all over the network, which is not suitable for banks, governments, securities.

3) Data cannot be tampered with.

4) The efficiency of transaction processing speed is low: all nodes participate in the whole network, and each node participates in the decision-making for the authenticity.

5) The issue of e-money is not supported by governments.

Private blockchain and consortium blockchain are the results of improving the openness and efficiency of public blockchain.

B. Private Blockchain

Private blockchain is not open to the outside. Only authorized nodes can participate in it. It is a partially decentralized blockchain. The main users are financial institutions, large enterprises, government departments, etc. For example, the digital currency issued by the central bank can only be recorded by the central bank, and individuals cannot participate. In addition, Alibaba, Baidu, Jingdong and other large companies will also build their own private blockchains, focusing on the data security.

C. Consortium Blockchain

Consortium blockchain, namely industry blockchain, is a blockchain constructed between companies and organizations. Usually, multiple nodes are selected to participate in the block consensus decision-making, while other nodes only participate in the transaction but not in the decision-making. This is a polycentric blockchain. User groups include banks, securities, insurance, group enterprises, etc.

VI. Blockchain Application

A. Development Considerations

With the development of blockchain, more and more blockchain applications begin to appear. In particular, due to the government's encouragement and support, we need to avoid the overheating situation of rushing to adopt the technology without considering the reality. Before deciding to use blockchain to develop applications, we need to consider the following aspects comprehensively.

1) Traditional methods. If an application can be implemented with traditional methods, it does not need to use blockchain technology.

2) Decentralization. This only means multiple copies of data, not the absolute reason to use blockchain. MySQL, Oracle, MongoDB and Ceph also support backups and have high performance.

3) Trustlessness. This means the need to prevent data fraud, and is a reason to adopt blockchain.

4) Traceability. Blockchain can trace the fraud data, and databases generally do not have the characteristics of traceability. This is also a reason to use blockchain.

5) Performance. If the data is updated frequently, using blockchain means a large amount of data and requires mass storage. This is not suitable for some unqualified applications. In addition, it is very inconvenient and inefficient to retrieve data in the blockchain.

B. Development Method of Integrating Traditional Database

After deciding to use blockchain technology to develop an application, we have to ask a question: whether to save all data to blockchain?

Blockchain is not as good as traditional database in performance and data retrieval. In addition, it will be particularly large if all data are saved to blockchain. A feasible method is to use 'blockchain + Database' to save data in the application. The important information that needs to be traced to prevent fraud can be put into the blockchain, and the other information can be put into the database. In blockchain applications, data is usually stored in multiple places. Therefore, the database should have the characteristics of high performance and multi **backup support. Even Ceph, a distributed storage system with high performance, high availability and high scalability, can also be considered.**

VII. Conclusion

Blockchain has the characteristics of decentralization, trustlessness and no third party. It is essentially a distributed data storage system. A blockchain is saved to multiple computer nodes. Through consensus mechanism, reward, fraud punishment and other measures, the blockchain finally establishes a reliable and trustworthy solution for bitcoin. However, blockchain also has some shortcomings, such as poor performance, difficult retrieval, and so on. Then, public blockchain, private blockchain, consortium blockchain for different applications have emerged. To some extent, private blockchain and consortium blockchain are no longer decentralized or trustless. During the development of application, we should consider a variety of factors to determine whether it is necessary to use blockchain technology. In some blockchain applications, we can use the way of 'blockchain + Database' to save data, that is, the key information to be traced is saved to the blockchain, and the other information is saved to the database or distributed storage system.

References

[1] Bitcoin home. [EB/OL]. [2021-03-10] https://history.btc126.com/today/btc/03-10/.

[2] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. [EB/OL]. [2021-03-10] https://bitcoin.org/en/bitcoin-paper.

[3] Wang P, Wei B and Wang C 2020 Application of blockchain technology in government data sharing. Application, doi:10.11959/j.issn.2096-0271.2020037, pp. 105-114.

[4] Zhang Y and Li N 2020 Data sharing technology of logistics application system based on Blockchain. Electronic Design Engineering, 28(14), pp. 72-76.

[5] Ethereum. [EB/OL]. [2021-03-10] https://ethereum.org/en.

[6] Hyperledger. [EB/OL]. [2021-03-10] https://www.hyperledger.org/use/fabric.

[7] Tencent. [EB/OL]. [2021-03-10] https://cloud.tencent.com/developer/news/167233.

[8] Li H. [EB/OL]. [2021-03-10] http://www.woshipm.com/blockchain/1022259.html.