Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/14/20230755

Introduction and analysis of bitcoin-based blockchain technology principle

Wangshu Zhu

Wuhan University, Wuhan, 430072, China

2020301041170@whu.edu.cn

Abstract. Blockchain technology is a technology that inherently solves trust issues. It has the characteristics of decentralization, distributed storage, tamper resistance, security, and transparency. It ensures reliable communication between nodes that do not trust each other through consensus mechanisms, smart contracts, and other means. Blockchain stores each transaction data on each transaction node to make the data public and transparent, and generates the data into a blockchain.. As a relatively new distributed database system, blockchain has expanded to many other fields since the initial digital currency, but its development is seriously constrained by problems such as large storage overhead and low query efficiency. In order to find a suitable optimization method, select the representative blockchain system Bitcoin, analyze its data structure, data storage and data query processing mechanism, discuss the problems existing in the two functions of storage and query, summarize the existing relevant optimization methods, and look forward to the main research problems of the blockchain system represented by Bitcoin in the future.

Keywords: blockchain, bitcoin, data structure.

1. Introduction

As one of the currently widely concerned and rapidly developing technologies, blockchain was first proposed in 2008 by Satoshi Nakamoto to solve the trust problem of traditional currencies relying on third parties [1]. Due to its advantages such as decentralization, non-tampering, traceability, high reliability, and high availability, it has been used as a safe and credible new distributed database system and has been applied to digital finance, agriculture, etc [2], medical treatment, smart city and many other fields [3-6].

However, in practical applications, the current blockchain system still has problems such as high storage overhead and low query efficiency that seriously affect the combined application of blockchain and other fields. Experts and scholars in related fields have made a lot of research and attempts for this. This article is based on previous research and practical applications, focusing on Bitcoin, a typical representative of the blockchain system. By analyzing its defects and optimization methods in storage and query, it hopes to provide reference for the future development of blockchain technology.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/14/20230755

2. Principle analysis of blockchain technology

2.1. Introduction to blockchain

Blockchain technology verifies and stores data through the block chain data structure, generates and updates data through the consensus algorithm of distributed nodes, ensures data transmission and access security through cryptography, and is composed of automated script codes. Smart contracts are used to program and manipulate data, and its essence is a new distributed infrastructure and computing paradigm [7]. At present, all blockchain applications can basically be divided into three types, namely public chain, alliance chain and private chain [8].

Bitcoin is a peer-to-peer electronic cash payment system, which belongs to the public chain in the blockchain, so there is no review mechanism and access requirements for users, mainly using the transaction model of Unspent Transaction Output (UTXO). As the first application of blockchain technology and the most successful decentralized cryptocurrency so far, Bitcoin has the longest and most powerful blockchain currently, and the most classic method of blockchain is still following the Bitcoin The storage structure and query technology adopted by the coin, and many subsequent innovative methods are also developed and evolved from this [9].

2.2. Bitcoin data structure

Blockchain adopts block chain structure as its core data structure, that is, a chain composed of blocks, each block is divided into two parts: block header and block body [10]. As shown in Figure 1, the block header of Bitcoin is responsible for the realization of most functions. It mainly consists of three sets of block metadata. The first set of metadata refers to the hash value of the parent block, which is used to compare the block with the previous one. blocks are connected. The second set of metadata is related to mining competition, that is, version number, timestamp, etc. The third set of metadata contains the Merkle tree root, a data structure that efficiently summarizes all transactions in a block. The block body only records and stores the transaction information within a certain historical time.

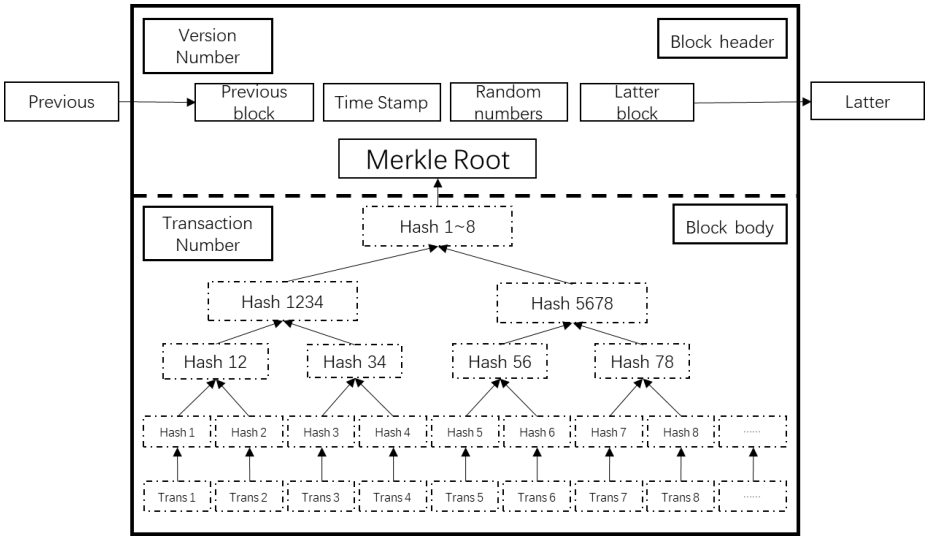


Figure 1. Block chain structure.

DOI: 10.54254/2755-2721/14/20230755

Table 1. Bitcoin block structure.

<table>
<tr><td rowspan="8">Block Structure</td><td>Block Header</td><td colspan="3">Version number
Previous block hash
Merkle tree root hash
Difficulty target
Time stamp
Random numbers</td></tr>
<tr><td rowspan="7">Block Body</td><td colspan="3">Transaction counter</td></tr>
<tr><td rowspan="7">Trade 1,2,…,n</td><td colspan="2">Num of inputs</td></tr>
<tr><td rowspan="2">Transaction inputs</td><td rowspan="2">The hash of the last transaction referenced
The output index of the last transaction
The input script</td></tr>
<tr><td colspan="2">Num of outputs</td></tr>
<tr><td rowspan="2">Transaction outputs</td><td rowspan="2">The number of bitcoins for the output transaction
The length of the output script
The output script</td></tr>
</table>

The detailed structure of a Bitcoin block is shown in Table 1. Each transaction is composed of input and output, and the transactions are connected to each other. The input of each transaction corresponds to the output of a previous transaction [11]. UTXO- based transactions contain reference information to the historical transactions of "unspent coins". The input content of the transaction mainly includes the hash of the last transaction, the output index of the last transaction and the input unlocking script. The output content of the transaction mainly includes the output transaction The number of bitcoins, the length of the output script, and the output lock script.

The Merkle tree is a binary tree composed of transaction hashes, originally proposed by Ralph Merkle [12], which is used to store the transactions of each block in Bitcoin. Each leaf node corresponds to a transaction hash, and the hash operation is performed layer by layer, and finally all data blocks are calculated into a Merkle root value, and the Merkle root can be recomputed to verify that the Merkle proof is valid [13], such as Figure 1 shows.

2.3. Bitcoin data storage methods

The storage system of Bitcoin mainly includes the file system and Level-DB [14,15]. As shown in Figure 2, the file system contains block data and "undo" data. Block data includes block body and block header information. "undo" data is the data for rollback correction when the chain fork occurs in the system. The purpose of the file system is to append data in the form of logs and perform retroactive operations on blocks. But in practical applications, searching through traversal is extremely inefficient, so the Level-DB system was introduced.

Bitcoin stores state data and index data in Level-DB in the form of key-value pairs, and is divided into state database and index database according to the format of the stored data. Among them, the state database is responsible for the UTXO of all current transactions, and the index database is responsible for the position of the transaction in the block, which is convenient for quickly locating related transactions. Therefore, the purpose of the Level-DB system is to improve access efficiency and assist system query and traceability.

DOI: 10.54254/2755-2721/14/20230755

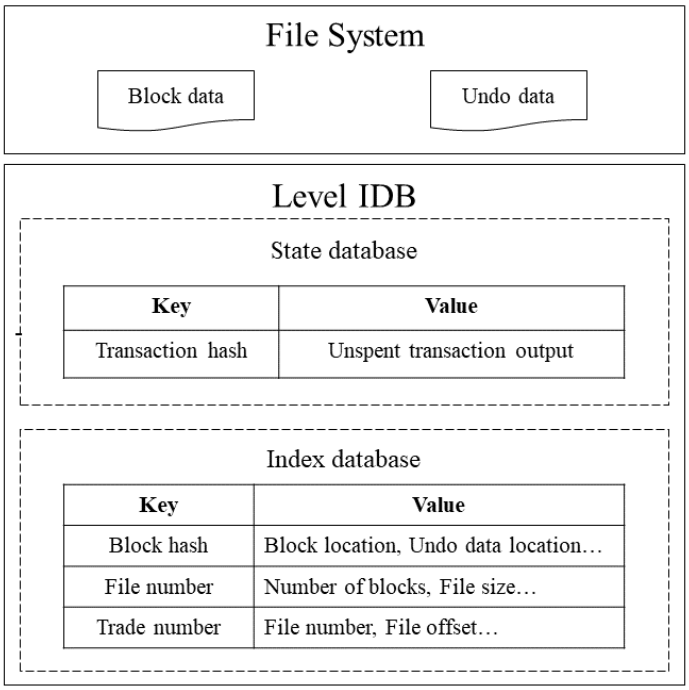


Figure 2. Bitcoin data storage system.

2.4. Bitcoin data query method of blockchain

Simplified Payment Verification (SPV) and index query in actual query processing [15].

The query mechanism of SPV is based on the Merkle tree structure mentioned above, which is generally applied to transaction payment verification. According to the structural characteristics of the Merkle tree itself, the SPV light node in the blockchain can verify whether a certain transaction exists in the blockchain only by virtue of the stored block header and the Merkle tree path, which greatly saves storage space. Lowering the query threshold for users. As shown in Figure 3, for example, to verify whether transaction 1 exists in the blockchain, it is only necessary to obtain the information of Hash34 and Hash5678 through the SPV mechanism.

DOI: 10.54254/2755-2721/14/20230755

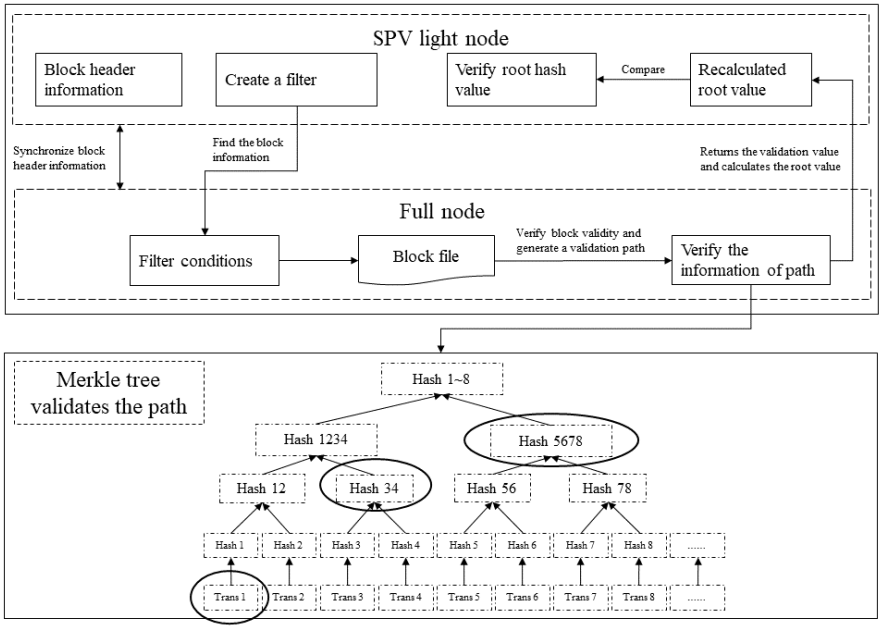


Figure 3. Process of Bitcoin SPV query.

The Bitcoin transaction query based on the index data in Level-DB is realized through the transaction number or transaction hash with the help of index information. The index information mainly includes the exchange file number, block position, and transaction offset in the block, etc. In the actual query, the system first obtains the relevant index information of the target transaction through the index database, and then locates the location of the target transaction in the file system according to the index information to obtain the target information. The specific query process is shown in Figure 4.

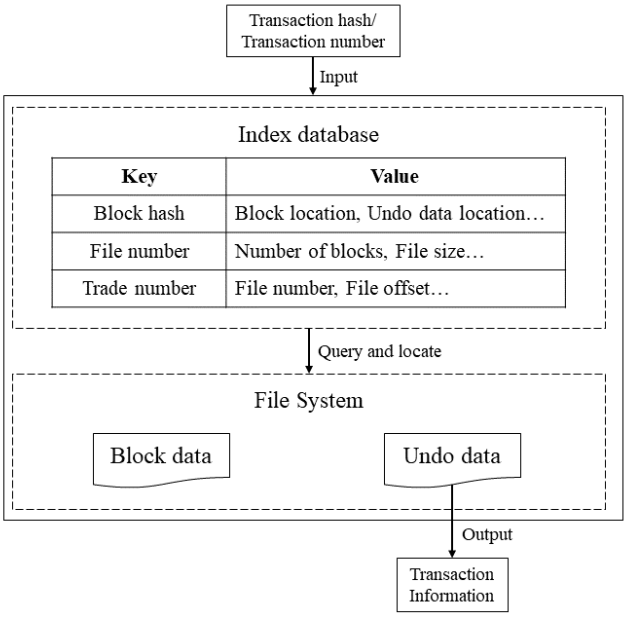


Figure 4. Process of Bitcoin index query.

DOI: 10.54254/2755-2721/14/20230755

3. Existing problems and optimization methods

3.1. Storage

Bitcoin adopts the multi-node high-redundancy mechanism of the blockchain, that is, full back up of node ledgers, status data, index data and other information, in order to realize the advantages of timely sharing, timely updating, and high consistency of the blockchain. But at the same time, it also led to the problem of node storage explosion, which made the resource overhead of node storage extremely large.

3.2. Query

Bitcoin adopts a block chain structure and Level-DB, an unstructured database oriented to content queries. Although it can be stored persistently and prevented from tampering, it also has shortcomings that cannot be ignored, such as low query efficiency and rare data types saved.

3.3. Optimization Method

There are still many research and practical problems in Bitcoin storage and query, and many experts and scholars in enterprises, open-source communities or academic fields have made optimization and improvement methods.

Clipping-based storage optimization method. Bitcoin-core mentioned a pruning strategy for the problem of storage explosion, that is, after the UTXO set of block data is constructed, the historical transaction information with low storage value is pruned and discarded, and the deletion operation is regarded as a Transactions are recorded in the blockchain to free up disk space and reduce storage pressure [16].

Sharding-based storage optimization method. The blockchain sharding technology can be traced back to the sharding protocol Elastico proposed by Luu, which separates the nodes in the network from the database based on the calculation results of PoW and assigns them to different areas (data servers) , each slice only needs to manage part of the data, thereby alleviating the pressure on data storage and processing [17,18].

Multi-chain based query optimization method. Cai built a double-chain model including account chain and transaction chain according to the difference in storage content [19]. By classifying and storing account information and post-transaction information, the system can process transactions concurrently, which greatly improves transaction processing and the speed of the query.

Query optimization method based on improved data structure. Zhang designed a credible query verification method by introducing the concept of Merkle mountain range (MMR) , and realized the dynamic addition of data to the blockchain to solve the problem of query verification information occupying too much storage resources in SPV light nodes. problem, improving the efficiency of query verification [20].

4. Technical outlook

4.1. Optimization of storage performance

The multi-node high redundancy mechanism for full backup of block data is an unavoidable problem for the blockchain. The idea of fragmentation and multi-chain is the current mainstream idea to solve this kind of problem, but there are still some problems in various coordination mechanisms of this type of method [17]. Only by improving the redundancy problem of full data backup can the possibility of storing large amounts of data in the blockchain be guaranteed, thereby maintaining the long-term operational efficiency of the blockchain network.

4.2. Selection of system structure

The typical blockchain represented by Bitcoin always has the problem of data explosion determined by the blockchain structure itself. Experts and scholars in related fields try to escape from the shackles of the block chain structure, and actively seek other possible new architectures to fundamentally solve the

DOI: 10.54254/2755-2721/14/20230755

storage and query problems of traditional block chains. For example, the DAG-type blockchain based on directed acyclic graph (DAG) technology stipulates that each transaction can be used as a block to connect with multiple previous transactions, and finally form a graph structure. In addition to the DAG-type blockchain, there are also new-type blockchains such as partition-type blockchains. However, due to problems such as late start, complex structure, and immature technology, these new system structures are still difficult to achieve in terms of decentralization and system security. Needs to be refined and validated.

5. Conclusion

Blockchain is still essentially a data ledger or database, so storage and query functions are still its core technology, and Bitcoin is by far the most mature and widespread application of the technology. This paper focuses on the analysis of the data structure, storage and query methods of Bitcoin, summarizes the existing problems and optimization methods, and looks forward to the future research direction of related technologies. As an emerging technology, blockchain has attracted the attention and attention of various countries and fields, and its combined application with other fields has developed rapidly, extending from the initial digital currency field to medical care, education, government affairs, and the Internet of Things. Although there are still many problems in the application of blockchain technology in actual scenarios and fields, it is the continuous emergence and overcoming of problems that promote the continuous development of technology.

References

[1] Nakamoto S . Bitcoin: A Peer-to-Peer Electronic Cash System[J]. consulted, 2008.
[2] Singh D, Monga S, Tanwar S, et al. Adoption of Blockchain Technology in Healthcare: Challenges, Solutions, and Comparisons[J]. Applied Sciences, 2023, 13(4): 2380.
[3] Tschorsch F , Scheuermann B . Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies[J]. IEEE Communications Surveys & Tutorials, 2016:2084-2123.
[4] Kamilaris A , Fonts A , Prenafeta-Boldu F X . The Rise of Blockchain Technology in Agriculture and Food Supply Chains[J]. Trends in Food Science & Technology, 2019, 91:640-652.
[5] Tsung-Ting, Kuo, Hugo, et al. Comparison of blockchain platforms: a systematic review and healthcare examples.[J]. Journal of the American Medical Informatics Association Jamia,
[6] Bhushan B , Khamparia A , Sagayam K M , et al. Blockchain for Smart Cities: A review of Architectures, Integration Trends and Future Research Directions[J]. Sustainable Cities a
[7] Chuen D , Deng R , Chuen D . Handbook of Digital Currency[J]. 2015.
[8] SHI J S,LI R.Overview of blockchain access control under the Internet of things [J].Journal of Software,2019,30 (6):1632-1648.
[9] Li J , Wolf T . A One-Way Proof-of-Work Protocol to Protect Controllers in Software-Defined Networks[C]// Symposium on Architectures for Networking & Communications Systems. IE
[10] LI M,SONG W P,HAO H,et al.IEEE Standard for Data Format for Blockchain Systems[J].Institute of Electrical and Electronics Engineers,2020,2(2418):1-32.
[11] CHEN H,WANG Y J.A Lightweight Scalable Protocol forPublic Blockchain[J].Journal of Computer Research and Deve-lopment,2020,57(7):1555-1567.
[12] Merkle R .MERKLE,RALPH ON NANOTECHNOLOGY AND BIOTECHNOLOGY[J]. Bio/technolgy, 1995, 13(5):440-440.
[13] ZHANG W B.Constructing blockchain world state Merkle Patricia Trie subtree:USA,10 929 374[P].2021-02-23.
[14] Kan J , Kim K S . MTFS: Merkle-Tree-Based File System[J]. IEEE, 2019.
[15] Gervais A , Karame G O , Gruber D , et al. On the Privacy Provisions of Bloom Filters in Lightweight Bitcoin Clients[J]. ACM, 2014.
[16] BLOCKSTREAM CORPORATION INC.Redefine trust[EB/OL].[2021-12-10].https://blockstream.com/.
[17] LUU L,NARAYANAN V,ZHENG C,et al.A secure sharding protocol for open

DOI: 10.54254/2755-2721/14/20230755

blockchains[C]//The 2016 ACM SIGSAC Conference on Computer and Communications Security.Vienna:ACM Press,2016:17-30.

[18] GARAY J A,KIAYIAS A,LENOARDOS N.The bitcoin backbone protocol:Analysis and applications[C]//Proceedings of the 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques.Sofia,Bulgaria,2015:281-310.Abbas H H, Kareem M I A, Gheni H M. A survey on security and policy aspects of blockchain technology[J]. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2023, 21(2): 302-313.

[19] Priya N. Analysis of B2B Blockchain Apps Using Hyperledgers and Their Implications in This Digital Era[M]//Bankruptcy and Reorganization in the Digital Business Era. IGI Global, 2023: 23-37.

[20] BAI C. State-of-the-art and future trends of blockchain based on DAG structure[C]//International Workshop on Structured Object-oriented Formal Language and Method,November 16,2018,Gold Coast:Springer,2019:183-196.