

A Scalable Double-Chain Storage Module for Blockchain

Hui Han^{1,2}, Wunan Wan^{1,2,*}, Jinquan Zhang^{1,2}, Zhi Qin^{1,2}, Xiaofang Qiu^{1,2}, Shibin Zhang^{1,2} and Jinyue Xia³

¹School of Cybersecurity, Chengdu University of Information Technology, Chengdu, 610225, China

²Advanced Cryptography and System Security Key Laboratory of Sichuan Province, Chengdu, 610225, China

³International Business Machines Corporation (IBM), New York, 10041NY212, USA

*Corresponding Author: Wunan Wan. Email: nan_www@cuit.edu.cn

Received: 14 February 2022; Accepted: 26 April 2022

Abstract: With the growing maturity of blockchain technology, its peer-to-peer model and fully duplicated data storage pattern enable blockchain to act as a distributed ledger in untrustworthy environments. Blockchain storage has also become a research hotspot in industry, finance, and academia due to its security, and its unique data storage management model is gradually becoming a key technology to play its value in various fields' applications. However, with the increasing amount of data written into the blockchain, the blockchain system faces many problems in its actual implementation of the application, such as high storage space occupation, low data flexibility and availability, low retrieval efficiency, poor scalability, etc. To improve the above problems, this paper combines off-chain storage technology and deduplication technology to optimize the blockchain storage model. Firstly, this paper adopts the double-chain model to reduce the data storage of the major chain system, which stores a small amount of primary data and supervises the vice chain through an Application Programming Interface (API). The vice chain stores a large number of copies of data as well as non-transactional data. Our model divides the vice chain storage system into two layers, including a storage layer and a processing layer. In the processing layer, deduplication technology is applied to reduce the redundancy of vice chain data. Our double-chain storage model with high scalability enhances data flexibility, is more suitable as a distributed storage system, and performs well in data retrieval.

Keywords: Blockchain; storage model; off-chain storage; de-duplication

1 Introduction

Blockchain is a concept proposed by Satoshi Nakamoto in the article [1], which describes the conceptual framework of peer-to-peer network technology and cryptography, timestamp technology, and blockchain technology. Blockchain is the integration of multiple existing and mature technologies, and it is defined as an application model for distributed data storage databases, peer-to-peer transmission, consensus mechanisms, encryption algorithms, and other technologies. At present, typical blockchain



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

systems include Bitcoin [1], Ethereum [2], Hyperledger Fabric [3], etc. All of them are able to operate in an untrusted environment and implement the abstraction of the ledger.

When acting as a distributed ledger [4], the blockchain system has to store a large number of copies of data to ensure the integrity and security of the data. The storage model naturally leads to a lot of data redundancy. Currently, the total data size in the Bitcoin network alone had reached 399.6 gigabytes. With the dramatic increase of data in the blockchain network, the flexibility and scalability of the blockchain system are reduced. The excessive volume of data makes it much more difficult for the blockchain to manage the data such as querying and updating.

The data storage module is mainly responsible for the persistence, coding, and decoding of blockchain data and the tasks of storing and accessing data in the system. Reference [5] describes the storage requirements in blockchain technology. At present, the block information of blockchain system mostly uses Key-Value database such as level database (LevelDB) [6] for data storage. LevelDB is an unstructured data storage system based on the Key-Value model, which improves the efficiency of writing and querying transaction data through Log-Structured-Merge tree (LSM-tree) structure. However, because accessing blockchain data requires traversing the entire system, the execution efficiency of the blockchain's storage structure is low. In order to optimize storage performance, many studies have improved the storage model of blockchain, they have respectively used data slicing techniques [7], data structure improvements [8,9], and modified encoding methods [10,11] to improve the storage performance of blockchains. These technologies are adapted within the blockchain. But the data on the blockchain is appended only, so it does not appear to solve the storage problem substantially in the long run. This is where the emergence of the idea of off-chain storage has led to many new optimization solutions. The need for new types of off-chain storage is introduced in [12]. The classical off-chain storage model EtherQL [13], an efficient query layer for Ethereum, uses a traditional database as the medium for off-chain storage. Traditional databases, however, tend to operate in a completely trusted environment as a prerequisite and there is data isomerism between traditional databases and blockchain networks, which makes the security and integrity of data can be compromised. Thus, a new model for off-chain storage becomes particularly important.

To sum up, the main challenges of blockchain as a model of non-reducible, append-only distributed storage are as follows:

1. The scalability of blockchain is reduced due to the sharp increase of data volume,
2. The structure of the off-chain storage system is different makes it difficult to maintain off-chain data safely and completely,
3. Inflexible data storage and inefficient query ability.

Based on the problems and challenges of using blockchain as a distributed storage system proposed above, this paper makes the following work:

1. The way of using vice chain storage is adopted to increase the throughput of the blockchain storage system,
2. Deduplication technology is brought into the vice chain to reduce the storage burden of the vice chain and improve its scalability.

The organization of this article is as follows: Section 2 summarizes the relevant work and technology of this paper. Section 3 describes the architecture of the double-chain storage model proposed in this article in detail. Section 4 analyzes the efficiency theory of the storage model and Section 5 concludes the work.

2 Related Work and Technology

2.1 Related Work

As an emerging technology, blockchain has strong scalability and can be well applied in many fields, it is used as a security database [14] in the fields of medicine, finance, and the Internet of Things and its decentralized nature makes it possible to use it to fight the Coronavirus Disease 2019 (COVID-19) [15], which is a big challenge facing all over the world. Blockchain can be thought of as a data structure that can run in an untrusted environment, cannot delete data, and only supports data appending. Blockchain is a distributed ledger technology built on a variety of technologies. Each block contains multiple transaction data and the full node will back up data backups of the entire system, which makes it obviously problematic. Bitcoin [1], as the originator of blockchain system, is a ledger model based on transactions. Its data structure is a list of data associated with hash Pointers. It adopts the structure of the Merkle tree [16] to generate the root hash value of transactions. Merkle Tree replaces traditional Pointers with hash Pointers, and the whole Merkle Tree can be detected by mastering the root hash so that the history of the block is protected accordingly. Ethereum 2.0 [2] requires participants to have relatively stable identities. It introduces a state tree, which maps from account address to account state. At this point, the state of the accounts in Ethernet constitutes a Merkle Tree to maintain consistency of state among all nodes. Since Merkle Tree does not provide good search and update methods, it is too expensive to insert new data when a new transaction is created if sorting would result. In order to solve the above problems, the Ethereum system introduced Merkle Patricia Tree (MPT) [17], it is both tamper-proof and, like Merkle Tree, no part of the entire tree changes as long as the root hash remains unchanged, so the state of each account can be guaranteed integrity, and the key-value pairs are preserved in the state tree. Hyperledger [3] has simple storage performance based on key-value pair mode, which has a certain improvement in efficiency compared with Ethereum. In article [18], the performance of blockchain as a data processing platform is evaluated. Blockchain technology can improve security during data transmission [19], however, the literature [20] has described that the performance of blockchain as a distributed ledger needs to be improved in terms of scalability, security, availability, and transaction retrieval speed.

In order to make better use of blockchain technology, research on optimizing the storage structure of blockchain systems has been continuing. Elastico [7] is the first public chain consensus protocol based on slicing, which is based on the idea of slicing transactions to reduce the redundancy of duplicate data. However, even after transactions are sliced, each node still needs to store the complete ledger. Similarly, the idea of data slicing is also applied in the storage scheme in literature [21]. The data is divided into cold and hot zones by using the least recently used (LRU) algorithm. The literature [22] proposes a scalable storage capacity model of blockchain called Elastic Chain, which reduces data redundancy on the premise of effectively ensuring data security and enables blockchain to support the storage of larger amounts of data to a certain extent. But due to the increase of blockchain data and users, the demand for querying transaction data in the blockchain is also rapidly increasing. Scalable storage and efficient retrieval (SE)-chain [8] proposed a new storage model based on Elastic Chain. It made structural improvements to the Merkle Tree structure used by traditional blockchain, introduced a balanced binary tree, and improved a balanced scaling algorithm to reduce the number of data duplicates. In this way, the scalability and data retrieval efficiency of blockchain can be increased. In paper [10], Erasure Coding was adopted to increase scalability and make its queries more efficient. Although the above schemes can improve the efficiency of blockchain storage to a certain extent and reduce data redundancy, all data is still stored on the blockchain. Therefore, it is still hard to resist the poor performance of blockchain storage caused by the increasing amount of data. At this time, off-chain storage is essential to a blockchain storage system. The paper [12] also introduces the

necessity of off-chain data storage through calculation. For example, the idea of an off-chain database proposed in the etherQL system [13]. Using an external database as a medium to store more data can effectively reduce the amount of data on the chain, and by listening to the interface to achieve data query operations. In addition to this system, there are many examples of off-chain storage, most of them, however, use traditional databases, cloud servers, and other storage means currently used on the Internet as the carrier for off-chain storage. That poses some problems, for example, there is isomerism between the data of traditional databases and the blockchain system. In the aspect of maintaining the consistency of data, there will be a lag in updating time. When the system crashes, it is difficult to recover. An asynchronous environment can lead to not being able to confirm whether the server is dead or delayed for too long.

2.2 Blockchain Distributed Storage

Blockchain technology is essentially a distributed database maintained by multiple participants. The current accepted basic model architecture of the blockchain system is shown in Fig. 1, which mainly includes the application layer, contract layer, incentive layer, consensus layer, network layer, and data layer. The data storage structure and data organization mode of blockchain systems are different from other distributed databases. Blockchain organizes transactional records into blocks. Each block contains a block body and a block header, and the chain structure by means of the hash value of the previous block recorded in the block header. This structure can ensure the tamper-proof, traceback, and verifiability of data, which makes it a great application demand in the untrusted environment. Distributed storage management is mainly focused on the contract layer, network layer, and data layer of the blockchain structure. The contract layer is responsible for processing the data in the blockchain system. The blockchain is mainly oriented to the untrusted storage environment, so the data are stored in a decentralized, fully replicated, distributed manner. In the network layer, the distributed algorithm and encryption signature are realized, and the nodes are linked through a peer-to-peer network structure, which has distribution, transparency, and scalability. The data layer is used to store the data information on the whole blockchain, which can be both transactional and non-transactional. The data storage structure in the blockchain system is mainly maintained in the data layer. Too much information stored in the chain will occupy too much block storage space, so to improve the efficiency of distributed storage of blockchain, we should mainly consider optimizing the structure of the data layer.

2.3 Off-Chain Storage Technology

At present, the off-chain data of models that use off-chain storage mostly refers to the non-transactional data like images, videos, files, etc. These data sets require too much storage space and storing them directly in the blockchain would place a great burden on the blockchain system. The basic idea of off-chain storage is to store data with high storage space occupancy and high demand for privacy protection off the chain, and only part of the data content is placed on the blockchain to prove that the transaction occurs, which is the main function of the main chain. As a result, systems like etherQL [13] store those data in the form of off-chain storage. In this form, some structured or unstructured data is encrypted by using hash functions, and then cloud storage and traditional database are used for off-chain storage. The structure of these models is shown in Fig. 2. These methods can reduce data redundancy on the chain and greatly increase the throughput of the blockchain itself. Therefore, the design idea of our scalable double-chain storage model is mainly derived from this model. However, due to the isomerism between cloud server or traditional databases and blockchain systems, there will be some uncontrollable problems, such as the raised difficulty

of blockchain data management, low retrieval efficiency, and query no response cannot determine whether the server is down, etc. In this paper, we introduce a double-chain model to improve storage efficiency and avoid the problems caused by isomerism, which can both reduce the redundancy of data on the chain as well as enhance availability and scalability. The security and integrity of data can also be improved to a great extent.

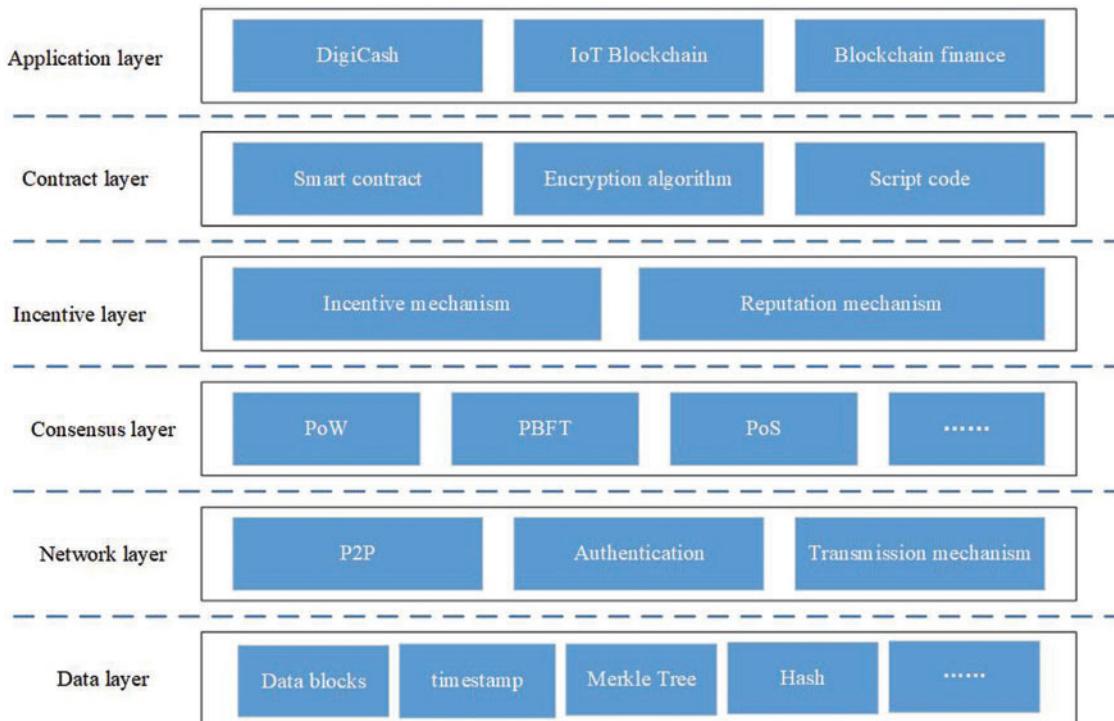


Figure 1: Blockchain layered structure

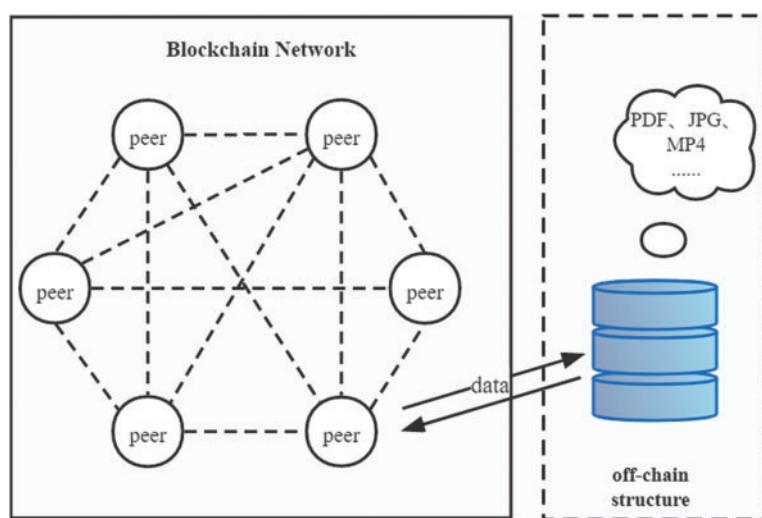


Figure 2: Off-chain storage model

2.4 Data De-Duplication

De-duplication [23], also known as capacity-optimized protection technology in the industry, is an efficient data streamlining technique that is widely used in various storage systems. The principle of the de-duplication system to detect duplicate data is to use a function $h(x)$, if two data blocks $a \neq b$, then $h(a) \neq h(b)$, and vice versa if $a = b$, then $h(a) = h(b)$. The research process of de-duplication mainly includes granularity division, fingerprint query, and time of reducing data redundancy. The real value of de-duplication lies in processing data that has already been stored. De-duplication technology is used to delete duplicate data to reduce data redundancy. The application of this technology in the storage system of blockchain can control the rapid growth of data to a certain extent, improve storage space and reduce the overall storage cost of blockchain. In this paper, de-duplication technology is mainly applied to the vice chain for fingerprint management. Since the storage model of blockchain generates hash fingerprints for each transaction or other related data, this makes it possible to detect duplicate data by data block fingerprint collision and delete a certain amount of duplicate data, thus reducing the number of copies to achieve storage space optimization.

3 Scalable Double-chain Storage Model (DCS)

In order to reduce the amount of data storage on the major chain to improve its scalability, our designed storage model is based on the idea of off-chain storage technology and added a blockchain called vice chain to store a large number of data copies and other non-transactional data. In this way, the storage burden on the major chain is reduced and the storage efficiency of the major chain is optimized.

In our double chain storage model, the major chain is the traditional peer-to-peer network, and the communication between nodes remains the same as in the traditional blockchain, but a huge number of data stored on the major chain is reduced and only a small number of key data is kept, and then an API is added to perform query and insert operations. The main job of the major chain is to supervise the side chain without storing all copies.

The vice chain is divided into two layers, namely the storage layer and the processing layer, respectively. The processing layer first performs the operation of generating fingerprints from transaction data and non-transaction data, then uses de-duplication technology to reduce the great many of data redundancy according to the fingerprint collision, and imports the deleted data into the storage layer and use the MPT structure to perform storage operations on the data. The storage model framework based on the double chain designed by us is shown in Fig. 3.

Due to the fact that blockchain can operate in untrustworthy environments, many industries hope to use blockchain as a storage model. The full replication model of blockchain determines that scalability will be greatly affected when large amounts of data are written into the blockchain system. The main point of this scheme is to use two chains to store different data separately, so as to reduce the pressure on the major chain storage and make it more flexible. Among them, the data stored in the blocks of the major chain include previous block hash, nonce, MPT root, timestamp, and other key data information. In addition, the major chain provides an API query interface and supervises the vice chain, and retrieves detailed data when necessary, and some non-transactional information is stored off the major chain. Considering the traditional off-chain storage will make the blockchain system less secure due to the asynchronous nature of data update, the storage model we designed is to increase the scalability of the blockchain itself on the basis of security by adding a vice chain into storage, and the vice chain storage still uses the blockchain system to store a large number of transaction copies and other non-transaction data such as images, audio, etc.

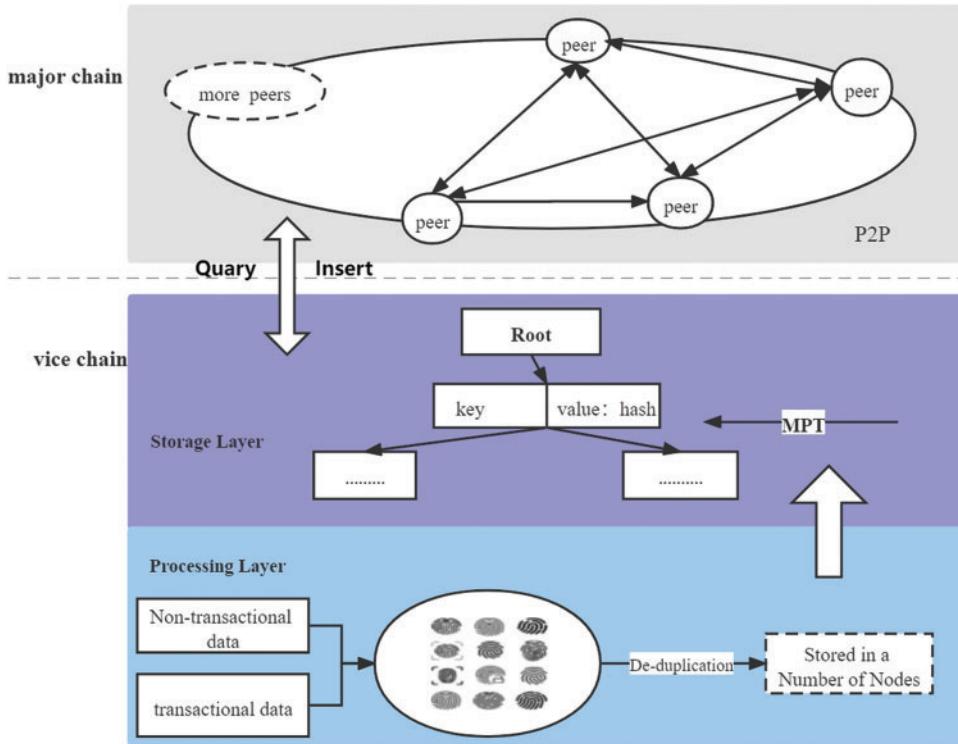


Figure 3: Double-chain storage architecture

3.1 Major Chain Storage Structure

According to the above data allocation, our model greatly reduces the storage pressure of blocks on the major chain. Although the type and quantity of data storage are reduced, in order to ensure security, the major chain will listen to and manage the data of the vice chain by setting a blockchain database API interface with the vice chain, establishing index and link between the major and vice chains through the API interface, and ensure its security through the major private key pair. Once the data is updated, key data will be updated in the API and inserted into the major chain, and then related fields in the major chain will be modified. The structure and content setup of the major chain is shown in Fig. 4, which is similar to the structure of a traditional blockchain, but does not require the storage of a complete tree structure or other non-transactional data. The stored fields are included in the figure, the root node of the transaction tree and the hash of the three-level transaction tree, the purpose of that is to avoid the risk of hash collision when there is a malicious node in the vice chain, the data is still secure, as well as version information, timestamps, and other regular information.

3.2 Vice Chain Storage Structure

The vice chain needs to store the complete copy of transaction data and some other hash of non-transaction type data. In order to reduce the burden of the vice chain, the storage structure of the vice chain has been improved to some extent. We divided it into a storage layer and a processing layer, the storage layer references the Merkle Patricia Tree used in Ethereum [2] to store all transaction data, MPT is the product of combining Merkle tree and tire tree, which can effectively reduce the storage space and increase the retrieval efficiency at the same time, while the de-duplication technique

is introduced in the processing layer to reduce the amount of data redundancy, so as to improve the storage performance of the vice chain.

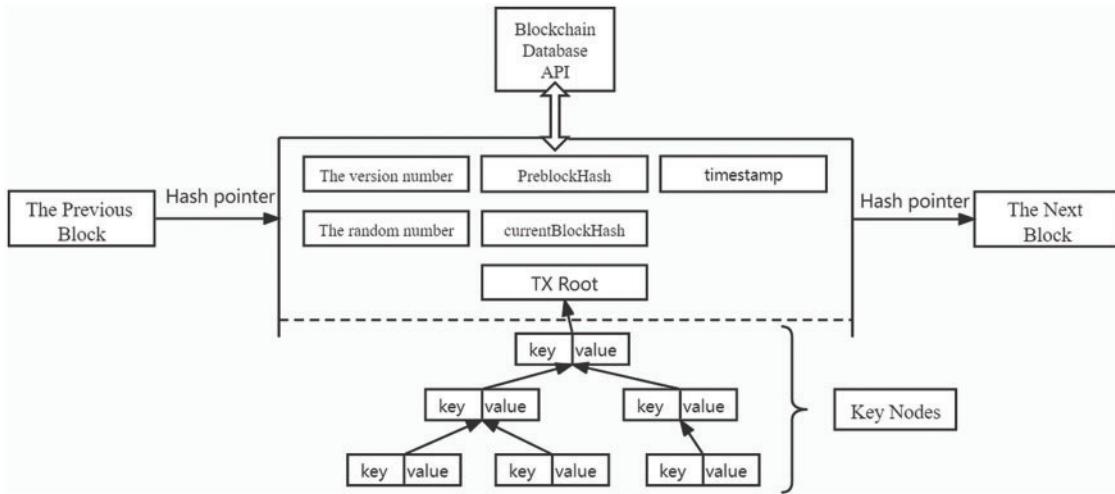


Figure 4: Major chain structure

3.2.1 The Storage Layer

The storage layer uses the data structure called Merkle Patricia Tree (MPT), which is used in Ethereum, to manage the data for accounts and generate hashes for transaction collections. MPT can store key-value pairs of arbitrary length by combining Merkle Tree and Tire Tree. The key is used to store keys belonging to the scope of the node, value is used to store the contents of the node and can quickly calculate the hash fingerprint of the maintained data. The advantage of tire trees is the ability to improve indexing efficiency and reduce excessive disk accesses. The vice chain is a way to store details of transactions that don't have to be made public on the major chain. Therefore, the vice chain structure can also be used for more secure storage of part of the data that needs strong confidentiality. In addition, there are a large amount of non-transactional data in the vice chain, which may be structured or unstructured. For these data, after a series of processing through the processing layer, a hash index table is established for the major chain to conveniently retrieve it. The storage layer in the vice chain has the same structure as the MPT is shown in Fig. 5.

3.2.2 The Processing Layer

Our storage model also introduces a processing layer to process data in the vice chain system. When a new transaction is generated, the tree structure needs to be updated and the full node needs to keep complete copies of the data, thus leading to a large amount of data redundancy. Therefore, On the premise of ensuring security, de-duplication technology was introduced to the processing layer. This idea is derived from the current popular distributed file management system InterPlanetary File System (IPFS) [24], which greatly reduces the data storage pressure through this technology. In this paper, some improvements are made to this approach to make it more applicable to blockchain systems. As the data in the vice chain includes both transactional and non-transactional data. For non-transactional data, duplicate data is removed by directly comparing the hash fingerprints generated on the relevant data as a way to reduce data redundancy; For transactional data, they have higher security requirements, so we should find out the number of identical transactions n according to

the fingerprint characteristics of transactional data blocks, and then refer to the duplication ratio adjustment algorithm to calculate the minimum duplication number c out of block i according to formula (1), in which N represents the total number of nodes in the vice chain and n represents the total number of the same transactions currently. Next, delete the same data whose number is greater than the minimum number c of repetitions.

$$C = 2^{-\left\lceil \frac{n-i}{100} \right\rceil} \times N \quad (1)$$

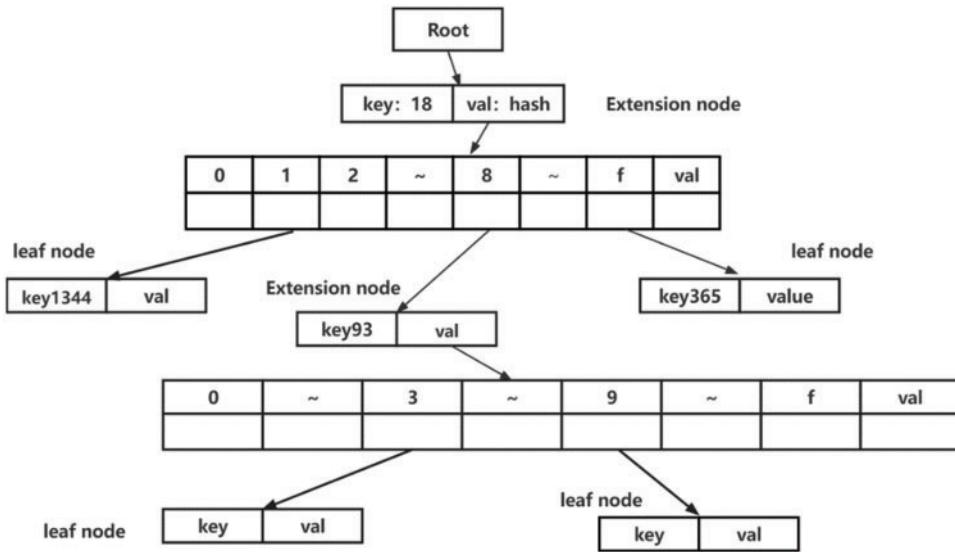


Figure 5: MPT structure

4 DCS Model Efficiency Analysis

4.1 Retrieval Efficiency Analysis

This model uses an off-chain storage model to reduce the storage pressure on the major chain, and the storage data structure used in the model is the Merkle Patricia Tree. Compared to traditional blockchain storage systems using Merkle Tree, MPT combines the Patricia Tire tree structure, which is a more space-efficient Tire, and for each node of the tree, if a node has only one child node, merge it with its parent node. Fig. 6 respectively shows the retrieval speed of Merkle Tree and Merkle Patricia Tree as the built-in structure of blockchain, and the response result of query request initiated by the major chain when Merkle Patricia Tree is applied to the vice chain of this model. According to the comparison, the retrieval efficiency of Merkle Patricia Tree is much better than Merkle Tree. While compared with MPT is used in the internal structure, the retrieval efficiency is reduced when it is used on the vice chain. That's because the deletion of some copies and the structure of the major and vice chain, the communication speed between nodes is reduced and the retrieval speed becomes low is also a normal phenomenon. Nevertheless, compared with the performance of traditional Merkle Tree, the retrieval efficiency of our model is also improved.

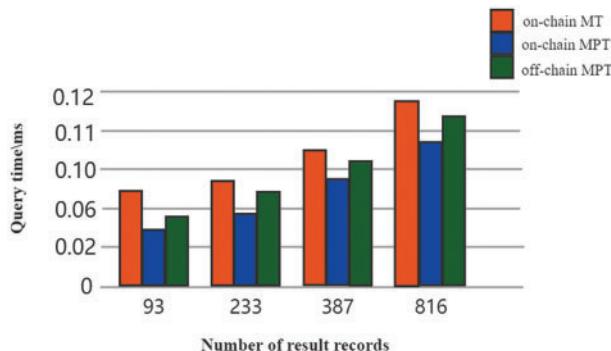


Figure 6: Retrieval average time

4.2 Storage Space Analysis

In our storage model based on double chain, only the root hash of the transaction and some important nodes and copies are stored on the major chain. Compared with the fully replicated blockchain storage system, the amount of storage per block on the major chain has been greatly reduced. When our MPT stored per block with a full binary tree, each block contains at least $2^n - 1$ nodes, while our major chain only needs to store three layers' nodes and the root hash, under the same conditions, only needs to store 16 nodes. So, in this case, the block on the major chain saves at least 2^{n-4} nodes' space. When it extended to the whole blockchain system, the storage performance of our major chain has been greatly improved. The reduced amount of stored data can make the major chain more flexible, and the same storage space can store more data.

The major chain needs to supervise the data of the vice chain, and a large number of copies and some non-transactional data are kept by the vice chain. More storage pressure is shifted to the vice chain storage structure. And we remove the duplicated non-transactional data in each block and the transactional data with more than the minimum number of copies according to the de-duplication technology, which can reduce the storage space waste of the vice chain and make it more flexible on the basis of ensuring security.

4.3 Scalability Analysis

When traditional blockchain is used as a distributed storage model, other industries often want to be in combination with it because of its security, when a great number of data is written to the blockchain system, too much data causes the storage performance of the blockchain to decline and cannot better adapt to the demand of big data on the Internet. The API is set on the major chain of our model, which can be associated with other unstructured systems while managing the vice chain. The double-chain storage structure also ensures the security of our model. Meanwhile, our double chain model has 1.64 times the storage space of traditional blockchain systems, with high throughput, can adapt well to big data environments, and perform well in terms of scalability.

5 Conclusion

As a distributed storage model, blockchain has brought new opportunities to the IoT and financial industries, while the dramatic increase in the volume of data in blockchain makes it much less flexible and makes it meet bottleneck period quickly in application. Therefore, in this paper, the traditional

storage model is improved by combining off-chain storage technology and de-duplication technology based on the problems exposed by blockchain in terms of scalability. Through theoretical analysis, our improved storage model has some degree of improvement in scalability and query efficiency.

In the future, the research focus of the blockchain data storage module lies in 1) Combine blockchain distributed storage with other emerging technologies to solve more cutting-edge problems. 2) Optimize the function and performance of the underlying data storage system, and design efficient storage and query strategies for unstructured systems and users; 3) As the use of blockchain systems becomes more and more widespread, cross-chain transactions also occur between different blockchain systems, so how to make storage and query for data in heterogeneous different blockchain systems.

Funding Statement: This work is supported by the Key Research and Development Project of Sichuan Province (No.2021YFSY0012, No. 2020YFG0307, No.2021YFG0332), the Key Research and Development Project of Chengdu (No. 2019-YF05-02028-GX), the Innovation Team of Quantum Security Communication of Sichuan Province (No.17TD0009), the Academic and Technical Leaders Training Funding Support Projects of Sichuan Province (No.2016120080102643).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>.
- [2] G. Wood, “Ethereum: A Secure decentralised generalised transaction ledger,” 2014. [Online]. Available: <https://gavwood.com/paper.pdf>.
- [3] An introduction to hyperledger, 2018. [Online]. Available: https://www.hyperledger.org/wp-content/uploads/2018/08/HL_Whitepaper_IntroductiontoHyperledger.pdf.
- [4] Q. Lin, P. Chang, G. Chen, K. Tan, Z. Wang *et al.*, “To-wards a non-2pc transaction management in distributed database systems,” in *Proc. of ACM Int. Conf. on Management of Data (SIGMOD)*, San Francisco, CA, USA, pp. 1659–1674, 2016.
- [5] B. Koteska, E. Karafiloski and A. Mishev, “Blockchain implementation quality challenges: A literature,” *Workshop of Software Quality, Analysis, Monitoring, Improvement and Applications*, vol. 11, pp. 11–13, 2017.
- [6] C. Mumford, D. Apostolou and V. Costan, “LevelDB” 2020. [Online] Available: <https://github.com/google/leveldb>.
- [7] J. Wang, C. Han, X. Yu, Y. Ren and R. S. Sherratt, “Distributed secure storage scheme based on sharding blockchain,” *Computers, Materials & Continua*, vol. 70, no. 3, pp. 4485–4502, 2022.
- [8] D. Y. Jia, J. C. Xin, Z. Q. Wang, H. Lei and G. R. Wang, “SE-Chain: A scalable storage and efficient retrieval model for blockchain,” *Journal of Computer Science and Technology*, vol. 36, no. 3, pp. 693–706, 2021.
- [9] S. H. Souk, S. H. Ahmed and D. Kim, “Hierarchical and based hash naming with compact tire name management scheme for vehicular content centric networks,” *Computer Communications*, vol. 71, no. 0, pp. 73–83, 2015.
- [10] X. Qi, Z. Zhang, C. Jin and A. Zhou, “BFT-Store: Storage partition for permissioned blockchain via erasure coding,” in *2020 IEEE 36th Int. Conf. on Data Engineering (ICDE)*, Dallas, TX, USA, pp. 1926–1929, 2020.
- [11] H. Wu, A. Ashikhmin, X. Wang, C. Li, S. Yang *et al.*, “Distributed error correction coding scheme for low storage blockchain systems,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7054–7071, 2020.
- [12] M. Ault, “Why new off-chain storage is required for blockchains,” 2018. [Online]. Available: <https://www.ibm.com/downloads/cas/RXOVXAPM>.

- [13] Y. Li, K. Zheng, Y. Yan, Q. Liu and X. Zhou, "EtherQL: A query layer for blockchain system," *Database Systems for Advanced Applications*, vol. 10178, pp. 556–567, 2017.
- [14] P. Wang and W. Susilo, "Data security storage model of the internet of things based on blockchain," *Computer Systems Science and Engineering*, vol. 36, no. 1, pp. 213–224, 2021.
- [15] S. H. Alsamhi, B. Lee, M. Guizani, N. Kumar, Y. Qiao *et al.*, "Blockchain for decentralized multi-drone to combat COVID-19 and future pandemics: Framework and proposed solutions," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. , pp. 9, 2021.
- [16] M. Yu, S. Sahraei, S. Li, S. Avestimehr, S. Kannan *et al.*, "Coded merkle tree: Solving data availability attacks in blockchains," *Cryptography and Security*, vol. 12059, pp. 114–134, 2020.
- [17] C. Chinchilla, "Merkle Patricia Tree," 2014. [Online]. Available: <https://github.com/ethereum/wiki/wiki/Patricia-Tree>.
- [18] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi *et al.*, "Untangling blockchain: A data processing view of blockchain systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1366–1385, 2018.
- [19] S. Hsiao and W. Sung, "Utilizing blockchain technology to improve WSN security for sensor data transmission," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 1899–1918, 2021.
- [20] H. T. Vo, A. Kundu and M. Mukesh, "Research directions in blockchain data management and analytics," in *EDBT*, USA, pp. 445–448, 2018.
- [21] J. Zhang, S. Zhong, J. Wang, X. Yu and O. Alfarraj, "A storage optimization scheme for blockchain transaction database," *Computer Systems Science and Engineering*, vol. 36, no. 3, pp. 521–535, 2021.
- [22] D. Jia, J. Xin, Z. Wang, W. Guo and G. Wang, "Elastic chain: Support very large blockchain by reducing data redundancy," in *Proc. APWeb WAIM. Macau, China*, vol. 10988, pp. 440–454, 2018.
- [23] K. Srinivasan, T. Bisson, G. Goodson and K. Voruganti, "iDedup: Latency-aware, inline data deduplication for primary storage," *Fast*, vol. 12, pp. 1–14, 2012.
- [24] J. Benet, "IPFS-Content addressed, versioned, P2P file system," 2014. [Online]. Available: <http://arxiv.org/pdf/1407.3561v1.pdf>.