A summary of the research on the foundation and application of blockchain technology

Rongli Gai¹, Xiaoyan Du², Shuya Ma³, Na Chen⁴, Shouchuan Gao⁵

Faculty of Information Engineering, Dalian University, Liaoning 116622, China
¹gairli@sict.ac.cn
²1249858867@qq.com
³641656776@qq.com
⁴13546172387@163.com
⁵15966399712@163.com

Abstract. Blockchain is an emerging distributed database technology. It has the characteristics of decentralization, non-tampering, traceability and final consistency. Blockchain can solve data management problems in untrusted environments. Based on the research and analysis of the blockchain system, this article expounds the application of the blockchain system in the distributed database environment. First, this article introduces the concept of the blockchain system and the classification of the blockchain system from multiple aspects. Then it introduces the data storage technology and data encryption technology adopted by the blockchain system in detail. Finally, the application prospects of the blockchain system in today's society are introduced.

1. Introduction
Nowadays, with the increasing popularity of the Internet, a large number of applications need to permanently record what happened, behavior status and so on in a distributed environment for future queries, that is, distributed accounting. If the traditional centralized accounting method is adopted, the transaction data may be tampered with by the accounting party at random or if the attack is encountered, the data is difficult to recover. In order to solve these problems, the bitcoin system using distributed accounting was first proposed in 2008. Bitcoin is a virtual cryptocurrency based on decentralization, adopting peer-to-peer network and consensus initiative, open source code, and using blockchain as the underlying technology. Puneet Kumar Kaushal and others proposed⁶: Bitcoin supersedes fiat currency in multiple dimensions because it can be transferred internationally without any limits, transactions have either no fees or a very low fee, currently it does not need any personal information (useful for anonymity), is transparent as every user has a copy of public ledger, and secure as the underlying cryptographic algorithm provides security. Subsequently, the blockchain as the underlying technology used in the Bitcoin system gradually attracted the attention of the academic community. Blockchain is a “peer-to-peer” decentralized ledger technology, it provides a method to record and distribute information about transaction publicly on a peer-to-peer system of computers through the crypto protocol⁷.

Compared with centralized database management system a blockchain is a decentralized database which is distributed via replication among its contributing peers and data are not directly updated, but a
collection of change records represented by transactions are appended as aggregated entities called blocks\cite{3}. All participating nodes can directly store data, and the durability of the transaction is achieved by the growing data chain maintained by the participating nodes and the decentralized consensus mechanism, which ensures the credibility of the data based on verification. In addition, blockchain technology is also equipped with the strengths of sustainability, compatibility, data sharing, and interconnectivity\cite{4}. In the past decade, the blockchain technology has evolved and became viable for various applications beyond the domain of finance\cite{5}. Nowadays, blockchain technology is very popular. On October 24, 2019, the Political Bureau of the Central Committee of the Communist Party of China conducted a collective study on the current status and trends of blockchain technology development. General Secretary Xi Jinping emphasized when he hosted the study, "We must use blockchain as an important breakthrough in independent innovation of core technology. Main direction, increase investment, focus on conquering a batch of key core technologies, and accelerate the development of blockchain technology and industrial innovation." In the future, blockchain technology will be combined with other emerging technologies to apply to various applications. For example: the combination of blockchain and medical treatment, the combination of blockchain and education, the combination of blockchain and energy and so on.

The recognized basic architecture model of the blockchain system is mainly divided into 6 layers\cite{5}. Figure 1 shows the six-layer structure. The first layer from top to bottom is the application layer. The application layer encapsulates various application scenarios and cases, similar to various websites and apps used in daily life, such as social entertainment, e-commerce shopping, and news reading. The second layer is the contract layer, which contains scripts, algorithms, and smart contracts. Popularly speaking, it is a custom electronic contract. The reason why it is called a smart contract is because this contract can automatically trigger execution when the constraints are met. Manual intervention is required, and the contract can be automatically terminated when the conditions are not met. In theory, it can trigger the execution of all the terms agreed in advance. This is also one of the core technologies that the blockchain can liberate the credit system. The third layer is the incentive layer. The incentive layer is responsible for the incentive issuance system and distribution system. It can be simply understood as a salary and reward system. The fourth layer is the consensus layer. The consensus layer is responsible for the consensus algorithm and consensus mechanism. At present, the most common and mature ones are three consensus mechanisms: POW, POS, and DPOS. The fifth layer is the network layer, which is a peer-to-peer (P2P) networking mechanism, data propagation and data verification mechanism. Because of the characteristics of P2P, the blockchain has a mechanism for automatic networking and is often called a distributed autonomous system. The last layer is the data layer. This layer is the bottom layer of the blockchain model. It encapsulates the chain structure of the data block, as well as the asymmetric public key private key encryption technology and time stamping technology.
2. Distributed data management of blockchain system

Blockchain is a distributed database management system that solves digital currency transfer, exchange and payment functions. The characteristics of the blockchain are mainly the decentralization of the network structure, non-tampering, and open and transparent data. The blockchain is mainly for accounting applications in untrusted environments, and all participants can store complete shared ledgers. It can be seen that there are similarities and differences between the management methods of the blockchain system and the traditional centralized data management and distributed database systems. Blockchain uses distributed ledger to provide a shared, immutable, and transparent history of all the actions taken by network participants[6].

The distributed accounting method of the blockchain system makes it the same as the distributed database in the way of data storage management, that is, stores the structured data collection, these data logically belong to the same system, and are physically distributed in different computer networks On the ground. Blockchain systems also have many of the characteristics of distributed databases, such as: distribution, transparency, autonomy, and scalability. From traditional database perspective, blockchain can effectively be utilized to ensure order of transactions among number of participating distributed nodes[7].

In addition, the original purpose of the blockchain design is to solve the problem of the credibility of data under untrusted exchange. Here, the untrusted environment means that the node responsible for data storage may tamper with the data arbitrarily but other participating nodes cannot recognize this process, which will cause the problem of mutual distrust among participating nodes. However, in the traditional distributed database management system, the system is built in a trusted environment, and the nodes completely trust each other. Therefore, there are significant differences in data management between the blockchain system and the traditional distributed database, which are mainly reflected in five aspects. As shown in Table 1.
In summary, compared with the traditional distributed database, the blockchain system provides better distribution, transparency and credibility in the way of accounting, and provides a tamper-proof verification mechanism and smart contract mechanism in function, so it is more suitable for anonymous use in untrusted environments.

3. Blockchain classification

The potential of blockchain technology has received attention in the area of FinTech—the combination of finance and technology[8]. There are many ways to classify the blockchain. The most common ones are classified according to the deployment method of the blockchain system and classified according to the role of the node of the blockchain system.

3.1 Classification of blockchain system deployment methods

Current blockchain systems are categorized roughly into three types: public blockchain, private blockchain and consortium blockchain. The public chain is open to everyone. Any Internet user can join and read data at any time, and can send transactions and participate in the process of block consensus. Differently, only a group of pre-selected nodes would participate in the consensus process of a consortium blockchain[9]. The alliance chain is a blockchain system that is only open to specific organizational groups. The private chain refers to a blockchain in which the block accounting authority is controlled by an organization or organization, and its read permission is not open to the outside world or restricted to some extent. The distinction between the three of them is shown in Table 2. The network structure of the public chain is completely decentralized, the network structure of the alliance chain is partially decentralized, and there are multiple trusted centers in the private chain. In the public chain, since any participating node can read data arbitrarily, the data transparency of the public chain is very high. But this also causes the data processing speed to be very low. The characteristic of the alliance chain is that its consensus process is controlled by pre-selected nodes. The data of the alliance chain may only be read by participants. The structure is partially decentralized. Its fault tolerance and transaction efficiency are relatively moderate. The private chain usually adopts a partial decentralized structure with a trusted center. Since it does not require a complicated consensus mechanism, its accounting efficiency is much higher than that of the public chain and alliance chain systems.

| Network structure | Public chain | Alliance chain | Private chain |
|-------------------|--------------|----------------|--------------|
| Node size         | No control   | Controllable   | Limited      |
| Joining mechanism | Can participate at any time | Specific groups or limited third parties | Nodes within the organization |
| Bookkeeper        | Any participating node | Pre-selected nodes | Nodes within the organization |
| Data reading      | People have read | Restricted read | Restricted read |
| Incentives        | There are token incentives | No token incentive | No token incentive |
3.2 Classification of node roles in blockchain systems

In the blockchain, according to the different functions in the system, the nodes correspond to different roles, and the roles set by different blockchain systems are also different. The classification according to node functions can be roughly divided into six types: accounting function, data storage function, transaction submission function, routing function, wallet function and member management. In the accounting function, each node obtains the accounting right by calculating the hash value in the new block, and then adds the new block to the original blockchain. The data storage function is to store block data, block chain metadata and state balance in the system and verify its validity. The transaction submission function is that users submit transaction records to the blockchain system. The routing function refers to nodes discovering neighbor nodes in the P2P network and communicating and synchronizing blocks. The wallet function can generate and save the private key, public key and specification geology corresponding to the public chain account. Finally, member management is used to provide member registration and manage member identity certificates.

In addition, the blockchain can also be classified according to the role of the node. The role of the node can be divided into internal and external roles. The internal role mainly exists in the blockchain network and is responsible for maintaining the operation of the blockchain system. It includes all nodes, Accounting node, simple payment verification node, sorting service node. The full node is characterized by storing a complete copy of the blockchain data, which can provide query services for transaction records. The accounting node includes accounting functions, data storage functions, and so on. The main task of the simple payment verification node is to verify the transaction, but does not store block data. The task of the sorting service node is to sort the received transactions and package them to generate blocks.

The external role of the blockchain system exists outside the blockchain network. Specifically, it includes a pure miner node, which is responsible for computing tasks that solve PoW problems, and otherwise does not have any function of the blockchain system. The second is the trading client, which can be an application used by users, applied outside the blockchain.

4. Data storage technology of blockchain system

The blockchain itself is a series of connected data blocks generated by a cryptographic algorithm. Each block contains two parts: a block header and a block body[10]. The blockchain itself is a series of connected data blocks generated by a cryptographic algorithm. Each block contains two parts: a block header and a block body. The block header is used to form the connection relationship between blocks, mainly including the hash value, time stamp, random number of the previous block and the root hash of the transaction. The block body is mainly used to store transaction technology and transaction details.

4.1 The main data structure of the blockchain system

At present, a large number of blockchain systems for the purpose of issuing digital currencies are designed based on Bitcoin's blockchain structure, that is, a Merkle tree structure is used to generate the root hash value of the transaction[11]. The Merkle trusted tree is created to solve the authentication problem in multiple one-time signatures. The Merkle trusted tree structure has the advantages of a large number of signatures at a time and has a significant advantage in authentication. The structure of the Merkle tree is shown in Figure 2. Each block will have a Merkle tree, which starts from a leaf node (bottom of the tree), and a leaf node is a transaction hash (bitcoin uses double SHA256 hash). The number of leaf nodes must be double, but not every block contains double transactions. Because, if the number of transactions in a block is singular, then the last leaf node (that is, the last transaction of the Merkle tree, not the last transaction of the block) is copied into a double number. From bottom to top, pairwise, connect two node hashes, and use the combined hash as the new hash. The new hash
becomes the new tree node. Repeat this process until there is only one node, the root of the tree. The root hash will then be used as the only indicator of the entire block transaction, save it to the block header, and use it for proof of work.

![Merkle tree structure](image)

**Figure 2** The structure of the Merkle tree

At present, most blockchain systems store block headers in the form of data files. Block data and metadata are stored using a data storage system based on the impairment model. The blockchain system mainly uses part-time databases such as LevelDB, and is used to improve the storage write efficiency and query access efficiency of transactions through the LMS-tree structure\(^{[12]}\). Here we introduce the organization of transaction data, index data and other meta-information of the main blockchain system.

The transaction data of the blockchain is Bitcoin. Bitcoin is the first system of currency which is completely decentralized and beyond the control of any monetary power\(^{[13]}\). Bitcoin can divide the data into four parts and store them in the blockchain system. Bitcoin's status data and block metadata are stored using LevelDB, which greatly improves the efficiency of access. The index data of the blockchain system is Ethereum. Ethereum, the well-known blockchain platform, does not have any limit for block size, unlike Bitcoin\(^{[14]}\). In the Ethereum blockchain system, the final data storage form is based on Key-Value key-value pairs, and uses LevelDB as the underlying database to store data. Ethereum uses LevelDB as the underlying database, it is a very efficient database implemented by google. Both key and value are byte arrays of arbitrary length. The database provides basic interface operations including Put(), Delete(), Get(), Batch(), and supports the atomicity of batch operations. Finally, let me introduce the concept of the Hyperledger. Hyperledger is a distributed ledger platform for running Chaincode. Hyperledger Fabric also have been developed in order to record the data for various business purposes, such as, logistics, energy trading, currency exchange and so on. Hyperledger Fabric is a permissioned and private blockchain. The biggest difference between the super ledger system and Bitcoin and Ethereum is that it supports multiple chains, each of which corresponds to a set of ledgers. The data contained in each ledger includes block data, block index, status data and historical data. Therefore, compared to the Bitcoin system and Ethereum, the super ledger has more obvious characteristics of distributed databases.

### 4.2 Storage optimization technology of blockchain system

With the development of the Internet, blockchain has now become the infrastructure of more and more decentralized models. However, problems are withholding the realization of the potentials\(^{[15]}\). In order
to improve data access efficiency and storage efficiency, the blockchain system uses three storage optimization methods: use database system to manage block data, use efficient index structure to improve data access efficiency and use distributed storage strategy to reduce node storage load.

Efficient database system can effectively improve the efficiency of data storage. Bitcoin and Ethereum and other LevelDB systems store data. In order to provide larger-scale data storage capabilities, the BigchainDB system directly increases the characteristics of the blockchain based on the distributed database system. In the database field, it directly implements the data storage system for the blockchain system, such as the ForkBase storage system, which involves the index structure of the university and the combination of the super ledger.

The second storage optimization method is the efficient index structure of the blockchain system. Using an efficient index structure can greatly improve data access efficiency. At present, the most used Bloom Filter index structure in the blockchain, it can quickly determine whether an element exists in a set, and at the same time determine the size of the storage space, so as to filter the block.

The last method of storage optimization is to adopt a distributed storage strategy. We know that the blockchain uses a distributed storage strategy to improve the efficiency of storage space. In order to realize the integration of multiple distributed blockchain storage nodes into a unit, literature proposes a distributed blockchain storage strategy based on a publicity unit, which at the same time ensures that this unit contains at least one complete area Blockchain copy.

4.3 Confidentiality mechanism of blockchain data storage
As we mentioned earlier, the blockchain system uses a completely shared ledger. Although this increases the transparency of the transaction, it also brings the risk of privacy leakage. However, the blockchain system must not only prevent personal information from being leaked due to release, but also provide privacy protection for the content of the transaction.

Sidra Malik et al. proposed a trust chain as a three-tier trust management framework that uses a consortium blockchain to track the interactions between supply chain participants and dynamically allocate trust and trust based on these interactions. Reputation score. This solves the trust problem associated with the data itself.

HYUNIL KIM and others proposed a DML model that allows blockchain privacy protection to systematically solve privacy, security, and performance issues. This model can handle any differential private learning algorithm that requires the definition of non-deterministic functions.

In addition, the most commonly used privacy protection method in blockchain systems is the encryption protection mechanism. The encryption protection mechanism is based on the principle of cryptography, using symmetric encryption algorithm and asymmetric encryption algorithm to realize that only relevant parties can view related data. This effectively protects the privacy of the data.

5. The application and prospect of blockchain
Blockchain is a new emerging technology that can bring a variety of benefits to enterprise mobile applications. In addition to that, it can add the element of confidentiality into the application and develop the advanced peer-to-peer payment platforms with the help of the distributed nature of the Blockchain. At present, many fields have combined their own business with blockchain, which not only solves the drawbacks of their own business, but also improves the drawbacks and deficiencies of the blockchain system itself.

HYE-YOUNG PAIK et al. Blockchain can be used in data management analysis, which improves people’s understanding of blockchain technology as a data storage system and promotes the application of it to large software systems. The existing distributed storage mechanism of the blockchain has achieved decentralization, and at the same time solved the problems of security and throughput, but there are still deficiencies in programmability, scalability, and confirmability, which is also the challenges faced by blockchain systems in data analysis and storage.

Blockchain enables shared access to information which is broadcasted across a network based on the trust of its participants. Industries aiming at services for the Internet of Vehicles (IoV) consider
blockchain as a leading technology for handling managerial as well as transmission aspects of vehicles [27]. Research groups in Mobility Open Blockchain Initiative (MOBI) have highlighted the potential use of this technology for IoV[28]. Vishal Sharma proposed a model[29], which can control the number of transactions through distributed cluster optimization, thereby handling the energy requirements of the Internet of Vehicles with blockchain functions.

In addition, blockchain technology also provides more secure and convenient bookkeeping services in the pan-financial field. In the field of credit management, the blockchain has built a consortium chain architecture and provides data that can be shared. In addition, it has important applications in distributed social networks and charity fields.

6. Conclusion
As an emerging distributed database technology, blockchain is playing an increasingly important role in life. This article introduces the role of blockchain in distributed database environment in detail. And classify the blockchain from multiple angles, and introduce the data storage technology of the blockchain in detail. In addition, it also looks forward to the future development prospects of blockchain. In the future, blockchain will play an indispensable role in multiple fields such as healthcare, finance, and education.

Acknowledgments
This work was financially supported by the Provincial Natural Fund Guidance Program: Research on Complex Curved Surface Trajectory Planning Algorithm (2019ZD0309)

References
[1] P. K. Kaushal, A. Bagga and R. Sobti, "Evolution of bitcoin and security risk in bitcoin wallets," 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur, 2017, pp. 172-177, doi: 10.1109/COMPTELIX.2017.8003959.
[2] Q. K. Nguyen and Q. V. Dang, "Blockchain Technology for the Advancement of the Future," 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, 2018, pp. 483-486, doi: 10.1109/GTSD.2018.8595577.
[3] M. H. Ziegler, M. Groβmann and U. R. Krieger, "Integration of Fog Computing and Blockchain Technology Using the Plasma Framework," 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Seoul, Korea (South), 2019, pp. 120-123, doi: 10.1109/BLOC.2019.8751308.
[4] S. Yu, K. Lv, Z. Shao, Y. Guo, J. Zou and B. Zhang, "A High Performance Blockchain Platform for Intelligent Devices," 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, 2018, pp. 260-261, doi: 10.1109/HOTICN.2018.8606017.
[5] Yuan Yong, Wang Fei-Yue. Blockchain: The State of the Art and Future Trends. ACTA AUTOMATICA SINICA, 2016, 42(4):481-494 (in Chinese)
[6] F. Baig and F. Wang, "Blockchain Enabled Distributed Data Management - A Vision," 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), Macao, Macao, 2019, pp. 28-30, doi: 10.1109/ICDEW.2019.00-39.
[7] Tien Tuan Anh Dinh, Ji Wang, Gang Chen, Rui Liu, Beng Chin Ooi,and Kian-Lee Tan. Blockbench: A framework for analyzing private blockchains. In Proceedings of ACM SIGMOD. ACM, 2017.
[8] H. Okada, S. Yamasaki and V. Bracamonte, "Proposed classification of blockchains based on authority and incentive dimensions," 2017 19th International Conference on Advanced Communication Technology (ICACT), Bongpyeong, 2017, pp. 593-597, doi: 10.23919/ICACT.2017.7890159.
[9] Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, 2017, pp. 557-564, doi:
10.1109/BigDataCongress.2017.85.

[10] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. White Paper, 2008

[11] Merkle R.C. Protocols for public key cryptosystems. //Proceedings of the 1980 Symposium on Security and Privacy, Oakland, USA, 1980: 122-133

[12] O’Neil P , Cheng E, Gawlick D, O’Neil E. The Log-Structured Merge-Tree (LSM Tree). Acta Informatica, 1996, 35(4): 351–385

[13] P. K. Kaushal, A. Bagga and R. Sobti, "Evolution of bitcoin and security risk in bitcoin wallets," 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur, 2017, pp. 172-177, doi: 10.1109/COMPTELIX.2017.8003959.

[14] S. Rouhani and R. Deters, "Performance analysis of ethereum transactions in private blockchain," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 70-74, doi: 10.1109/ICSESS.2017.8342866.

[15] S. J. Moon, I. H. Park, B. S. Lee and J. Ju Wook, "A Hyperledger-based P2P Energy Trading Scheme using Cloud Computing with Low Capability Devices," 2019 IEEE International Conference on Smart Cloud (SmartCloud), Tokyo, Japan, 2019, pp. 190-192, doi: 10.1109/SmartCloud.2019.00039.

[16] BigchainDB GmbH. BigchainDB 2.0 The Blockchain Database. White Paper, 2018

[17] Wang S , Dinh T.A , Lin Q, et al. ForkBase: An Efficient Storage Engine for Blockchain and Forkable Applications. //Proceedings of 44th International Conference on Very Large Data Bases, Rio de Janeiro, Brazil, 2018:1137-1150

[18] Xu Zi-Hua, Han Si-Y uan, Chen Lei. CUB, a Consensus Unit-based Storage Scheme for Blockchain System. //Proceedings of IEEE 34th International Conference on Data Engineering, Paris, France, 2018: 173-184

[19] Zhu Lie-Huang, Gao Feng, Shen Meng, et al. Survey on Privacy Preserving Techniques for Blockchain Technology. Journal of Computer Research and Development. 2017, 54(10):2170-2186(in Chinese)

[20] S. Malik, V. Dedeoglu, S. S. Kanhere and R. Jurdak, "TrustChain: Trust Management in Blockchain and IoT Supported Supply Chains," 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 2019, pp. 184-193, doi: 10.1109/Blockchain.2019.00032.

[21] H. Kim, S. Kim, J. Y. Hwang and C. Seo, "Efficient Privacy-Preserving Machine Learning for Blockchain Network," in IEEE Access, vol. 7, pp. 136481-136495, 2019, doi: 10.1109/ACCESS.2019.2940052.

[22] G. S. Praba Devi and J. C. Miracil Joyce Pamila, "Accident Alert System Application Using a Privacy-Preserving Blockchain-Based Incentive Mechanism," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 390-394, doi: 10.1109/ICACCS.2019.8728507.

[23] I.-C. Lin and T.-C. Liao. “A survey of blockchain security issues and challenges,” International Journal of Network Security, vol.19, no.5, pp. 653-659, Sept. 2017.

[24] H. Paik, X. Xu, H. M. N. D. Bandara, S. U. Lee and S. K. Lo, "Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance," in IEEE Access, vol. 7, pp. 186091-186107, 2019, doi: 10.1109/ACCESS.2019.2961404.

[25] Cao Y uan, Zhang Cong, Ding Zhao-Y un, Jiang Xin-Wen. Blockchain Technology on DAG Principle and Practice. Beijing: China Machine Press,2018(in Chinese)

[26] J. Gu et al., “Consortium blockchain-based malware detection in mobile devices,” IEEE Access, vol. 6, pp. 12118–12128, 2018.

[27] H. Liu et al., “Blockchain-enabled security in electric vehiclescloud and edge computing.” IEEE Netw., vol. 32, no. 3, pp. 78–83,May 2018.

[28] E. Uhlemann, “Time for autonomous vehicles to connect [connectedvehicles],” IEEE Veh. Technol. Mag., vol. 13, no. 3, pp. 10–13, Sep. 2018.

[29] V. Sharma, "An Energy-Efficient Transaction Model for the Blockchain-Enabled Internet of
Vehicles (IoV)," in IEEE Communications Letters, vol. 23, no. 2, pp. 246-249, Feb. 2019, doi: 10.1109/LCOMM.2018.2883629.