Applications and Challenges of Blockchain Technology in Data Security and Data Governance

Zhili Deng[1,2], Dong Han[3,*], Jingwei Sun[4], Peng Sun[5]

[1]The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[2]Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[3]Houlang Technology (Hainan Special Economic Zone) Co., Ltd., Chengmai County, Hainan Province, China
[4]Guangzhou Wisdom Digital Innovation Service Co., Ltd, Guangzhou, China
[5]Guangzhou Xinhua University, Guangzhou, China
[*](Dong Han):pksoon@163.com

Abstract: *Blockchain technology has the potential to enhance data security and data governance. Its decentralized nature ensures data security and integrity, and prevents data leaks and abuse. In terms of data governance, blockchain technology can optimize data standardization, quality management, and classification, providing transparency and credibility for data decision-making[1]. Blockchain technology has broad applications in mobile applications, healthcare, and the Internet of Things (IoT), enabling verifiable and fair distributed machine learning. However, blockchain technology still faces challenges such as performance bottlenecks, security issues, and privacy protection. Further technological innovation and research are needed to address these challenges and find better solutions.*

Keywords: *Blockchain technology, data security, data governance, mobile applications*

1. Introduction

With the advent of the blockchain 3.0 era, blockchain technology is widely applied in various fields. Because of its special chain data structure, distributed algorithm, smart contract, and decentralized characteristics[2], it can solve data security problems and improve the methods and theories of data governance. Especially in logistics finance, the Internet of Things (IoT), supply chain management, and many other areas[3].

2. The Development of Blockchain Technology in Data Security

2.1. Impacts and Applications in Data Security

Blockchain technology has increased innovation efforts in development of many industries and fields,such as financial logistics[4], and even brings a breakthrough. However, as time passes, more information data has been accumulated, blockchain technology still handle current problems, but "congestion" cannot be avoid[5], as a result, even some problems can be solved or avoid, cannot completely prevent problems from occurring during the transaction process. In this case, demand technology innovation to improve performance, efficiency and security is urgent. If it’s used on a large scale in financial logistics, it will also face the risk of system paralysis during the transaction process. One of the reasons of this problems is non-uniformity of technology and evaluation criteria, and then different blockchain platforms lack inclusiveness and interoperability.

In the application of IoT, as one of the most important components of 6G technology, compared with traditional technology, blockchain can effectively solve the data security and privacy issues. For the problem caused by excessive amounts of data, such as network congestion or throughput limitation. The current common solution is to use uniform or random sharding techniques, which can solve part of problems above, but because of performance differences and collaborative transactions between nodes, they also may cause problems such as insecure shards, single points overheating or some issues else, which are even not suitable for complex and dynamic environments[6].

Supply chain management is the most typical application of blockchain. The data transparency and decentralization of blockchain can achieve the entire lifecycle management of products, enhancing the transparency of the supply chain and enabling traceability through blockchain technology to establish a corresponding credit ecosystem. However, because of transactions and data quantity increasing, the verification speed of blockchain will become slower and the cost will increase.[7]

2.2. Technological Developments in Data Security

Researchers have conducted extensive research on query methods on blockchain platforms and proposed many relevant research solutions. As a trusted machine working in a trustless environment, blockchain has been proven to be effective in data security, decentralization, and distributed ledger technology[8]. Many studies have been conducted on ensuring that data on the chain can be effectively queried and verified. Generally, blockchain query methods rely on using distributed databases, data indexing structures, and encryption algorithms to achieve efficient, verifiable, and secure queries[9]. Blockchain query technologies can be divided into three categories: efficient query technology, verifiable query technology, and privacy-preserving query technology.

Efficient query technology aims to provide high-performance data querying on the blockchain. The underlying data storage system exhibits high write performance when processing frequent query functions, but lacks read performance. Based on the assumption that both full nodes and clients are honest nodes, the existing query methods provide more efficient query services and richer query functions. There are three main strategies to improve query efficiency, including utilizing smart contracts, making underlying data accessible to other databases, or altering the internal design of the blockchain to include appropriate indexing structures. Adding an index structure typically retains the properties of a Merkle tree, even when introducing characteristics of other data structures such as B-trees, B+ trees, and Trie. While this structure allows for real-time querying and ensures query structures unchanged, it often sacrifices storage space and query functionality richness. The issue of unreliable query results is addressed through two strategies: utilizing smart contracts and external databases[10].

Verifiable query technology aims to provide query outputs that can be verified for security purposes. Authenticated data structures (ADS) based on cryptographic techniques or other techniques based on cryptographic, such as multi-party signatures, database fingerprints, and trusted execution environments, are the current main trends. While verifiable queries guarantee the reliability of query results, they incur slightly higher storage and processing costs.

Privacy-preserving query technology aims to provide query results without violating relevant security requirements and privacy. Assuming that full nodes and clients are semi-honest, with increasing data volume and application expansion, identity leaks, data leaks, query statements, and query results can be captured by attackers during the query process. Mainstream methods for protecting data privacy typically include storing encrypted data in cloud databases, indexing files to the blockchain, utilizing searchable encryption, private information retrieval, and differential privacy methods. These methods prevent the leakage of encrypted data while ensuring data integrity.

2.3. Challenges and Future Developments in Data Security

The existing data query strategy for blockchain faces many challenges. Currently, the most efficient method to handle blockchain queries is to combine distributed databases with blockchain queries. However, this approach cannot guarantee real-time data comparability between databases. Users may encounter issues when querying the most recent data from other subchain nodes, as the synchronization module still needs to enter the next synchronization cycle. To address this issue and achieve efficient querying, the mainstream solution now is to design new data structures. Furthermore, blockchain's data structure should not only provide efficient access but also maintain data consistency. When conducting data queries outside the blockchain, it is necessary to include an additional verification step to confirm the accuracy of blockchain results. Therefore, the second issue is to construct data structures to provide fast access to stored data while ensuring data integrity and reliability through effective verification methods. With the increase in data volume, storage scalability must also be considered. Currently, there are three trends for addressing data storage scalability: improving data structures, using sharding techniques, and cross-chain exchanges. Both sharding and cross-chain exchanges require many connection operations to interact, and cross-chain exchanges involve consistency, data structure configuration, and trust model selection issues. Therefore, ensuring high efficiency and security is a challenge that requires further research by scholars.

Another challenge faced by the existing blockchain data query strategy is related to privacy protection.

There are many encryption algorithms available to provide solutions for privacy protection during the query process and data privacy protection itself. However, due to the complexity of cryptographic primitives, significant processing is often required to implement effective indexing. This necessitates modifications to existing solutions to meet the practical goals. When generating index structures or storing files with searchable encryption, the average runtime is proportional to the number of files. This indicates that computational costs increase with the number of files, making it an emerging research direction to improve data availability under weaker threat assumptions.

3. Development of blockchain technology in data governance

3.1. Impact and Applications in Data Governance

Data is gradually becoming the cornerstone of production and life in the new era. In the era of big data, security risks such as privacy data leakage, data abuse, and data falsification leading to data distortion have always existed. The lack of a sound data security management system and the lack of autonomous capabilities in network systems highlight the issue of data security. Data security issues have become the main bottleneck restricting the development and utilization of data and value mining. The most fundamental way to solve data security problems is to conduct targeted security governance on data.

With the advancement of data governance practices in various industries and organizations, some changes and trends are gradually emerging. Data governance is becoming more intelligent, data governance and artificial intelligence have become two main research objects in recent year. On the one hand, through data governance, enterprises can improve data quality and enhance data compliance, thus providing high-quality compliant data for the application of artificial intelligence. On the other hand, artificial intelligence can play the optimization effects on data governance[6]. Artificial intelligence helps achieve the perfect integration of conceptual models and computer models, thereby optimizing data model management; it can collect unstructured data, extract key information, and help integrate metadata; it can help enterprises identify master data, define and maintain data matching rules, assisting in enterprise master data management[11]; it can define transformation rules, extract data quality assessment dimensions, and evaluate the effect of data cleaning and data quality through supervised learning and deep learning, maximizing dynamic improvement of data quality[12]; it can promote data classification and further enhance the data security protection system, promoting data security[13].

3.2. Technological Developments in Data Governance

Currently, common data governance technologies include data standardization, data quality management, and data classification and grading.

Data standardization operates a large number of data with different sources, formats, and content, so that it can be integrated and utilized, address data silo issues and promote data sharing. While many standardization-work can be automated, there is still a large number of manual work required because of data value preservation and standardization process inefficiencies.

Data quality management technology primarily focuses on verifying user data compliance by combining privacy computing technology and achieving data quality monitoring by implementing comprehensive security monitoring of the whole cycle of data storage scheduling. To verify the integrity of participating users and ensure the compliance of their data, the zero-knowledge proof techniques of privacy computing are commonly used. However, the zero-knowledge proof techniques of privacy computing are mostly considered as an auxiliary technology[14].

Data classification and grading mainly emphasizes implementing different management and protection measures for different data, facilitating a clear understanding of various data resources, and providing differentiated storage, circulation, and application pathways. This promotes a balance between data protection and utilization and serves as a crucial aspect of data governance. However, due to the lack of unified classification and grading standards, the demarcation lines for data levels are currently unclear [15].

3.3. The problems and challenges in data governance

In the era of big data, there have always been security risks such as privacy data leakage, data abuse, and data forgery leading to data distortion. Because of the lack of a sound data security management system and the lack of autonomy in network systems, data security problems have become a major bottleneck restricting the development and utilization of data as well as value mining[16]. The fundamental way to solve data security problems is to conduct targeted security governance on the data [17].

In the process of data governance, the main problems currently include[18]:

- The existence of data island problem makes it difficult to achieve governance. A large investment in data governance in the short term does not yield visible results, because the value assessment of data governance is difficult to quantify, the data is held by many different subjects, and the information between each subject is not circulated which prevent the true utilization of data. Due to data heterogeneity, it is impossible to simply integrate the data, and how to realize data sharing is also a very important problem.
- Ensuring data quality is difficult. The majority of strategic decisions currently heavily rely on data, so the importance of data quality is self-evident,. Low-quality data can lead to erroneous decisions, and decision-making errors due to low-quality or incomplete data can result in catastrophic situations.
- Outdated data management methods are prevalent. Many organizations still use relatively outdated methods to manage data, such as paper records, CD storage, or even manual records. Unclear role divisions, outdated management methods, and incomplete data systems make data sharing difficult, and the value of data cannot be fully tapped and demonstrated.

4. Conclusion

Currently, there are studies leveraging blockchain to enhance the transparency of data acquisition and sharing in areas such as mobile applications, healthcare, and the Internet of Things (IoT)[19]. The framework for data acquisition and sharing implemented based on blockchain technology increases the transparency of data flow through layered processing[20-23].

Blockchain can effectively address data governance issues, improve data utilization efficiency, share information, and provide privacy and security guarantees. In response to the problem of multiple data sources in a system, blockchain consensus nodes can streamline efficient business frameworks, obtain different restrictions on data access permissions, follow fixed processes for data acquisition, and set fixed channels for data sources, avoiding the inefficiencies caused by duplicate or multi-channel queries [24]. In the work of data collection and transmission, distributed work methods can reduce data duplication and loss, and data queries and processing can be traced back to their sources, the whole network nodes can be notified, and the data processing behavior can be proved by time stamps and transaction hashing.

Regarding data privacy and security, through consortium blockchain, the permissions of each node are clearly defined, the possibility of malicious behavior by nodes under the consensus mechanism is reduced， and the risk of sensitive data exposure is reduced, so organization structure is improved through the blockchain's authentication and traceability functions [26]. Based on blockchain, distributed data governance firstly sets up rights and responsibilities within the consortium blockchain before building the platform, and then continuously improves the network in practical work, implements organizational management concepts in actual execution. When data governance is based on blockchain network consensus of the whole network of blockchain, the changes in business iterations and technical characteristics will be reflected from various business departments to user clients, thereby avoiding unsustainable data governance caused by iterations or the problem of the quality of later data governance work. It will ultimately enhance management's confidence in data governance and ensure steady progress in digital transformation.

References

[1] Gao Hang, Yu Xue. Research on the application of blockchain technology in cross-domain data sharing and privacy protection [J]. China Management Informatization, 2023,26 (22): 185-187.

[2] Meng Xiaofeng, Liu Lixin. Blockchain and data Governance [J]. China Science Foundation of China, 2020,34 (01): 12-17.

[3] Data governace: status, technologies, applications and development trends. 2023 8th international conference on data science in cyberspace(DSC), P143-150

[4] Zhou Hongsen. Trial analysis of China's logistics finance innovation under the blockchain technology [J]. Transportation and Storage in China, 2023, (05): 65-66.

[5] Shen Dong, Lv Yi, etc. Application and analysis of big data analysis technology in communication network operation and maintenance [J]. Wireless Internet Technology, 2023,20 (12): 165-168.

[6] CAI Ziyue, Tan Beihai, Yu Rong, etc. Dynamic fragmentation of blockchain for 6G Internet of Things device collaboration [J]. Computer Engineering, 2024,50 (01): 50-59.

[7] Yazdinejad A, M.Parizi P, Dehghantanha A, et al. An energy-efficient SDN controller architecture for IoT networks with blockchain-based security [J]. IEEE Transactions on Services Computing, 2020,13(4):625-638.

[8] Ramesh S, Rizwan P, Arunprasath R, et al. Survival Study on Blockchain Based 6G-Enabled Mobile Edge Computation for IoT Automation [J]. IEEE, 2020, 202(8): 143453-143463.

[9] Pan Heng, Qian Haiyang, Yao Zhongyuan, etc. Review of typical blockchain storage and query techniques [J]. Journal of Zhengzhou University (Science edition), 2022,54 (06): 34-50.

[10] Wang Qiange, He Pu, Nie Tiezheng, et al. Overview of the data storage and query technologies of the blockchain system [J]. Computer Science, 2018,45 (12): 7.

[11] Li Yufei. Application of artificial intelligence in data governance [J]. ICT and Policy, 2019, (05): 23-27.

[12] ZHANG Rui, XUE Rui, LIU Ling. Security and privacy on blockchain [J]. ACM Compluting Surveys, 2019(3): 1-34.

[13] Gao Lei, Zhao Zhangjie, Song Jinsong, etc. Data security governance technology and practice in big data application [J]. Information Security Research, 2022,8 (04): 326-332.

[14] Zhang Zhengquan, Hu Sen, Mo Xiaokang. Review of zero-knowledge proof studies [J]. Digital Communications World, 2023 (06): 79-81.

[15] Zhang Zhipeng. The logic and regulation of blockchain enabling carbon data security governance [J]. Intelligence Magazine, 2023,42 (05): 86-93.

[16] Zhang Min, Wei Wei, Tan Tianyi, etc. Data classification and its development path [J]. Network security and Data Governance, 2022,41 (07): 18-22 + 29.

[17] Jennifer Z.S. What you need to know about Facebook & Cambridge Analytica[Z]. https://www.cbsnews.com/news/what-you-need-to-know-about-facebook-cambridge-analytica/. [2018-04-04]/[2024-03-15].

[18] Zyskind G, Nathan O. Decentralizing privacy: using blockchain to protect personal data//Proc of IEEE Security and Privacy Workshops. Washington: IEEE, 2015:180-184.

[19] Hou Peng, Li Zhixin, Zhang Fei, et al. Intelligent technology and practice of financial data security governance [J]. Journal of Network and Information Security, 2023,9 (03): 174-187.

[20] Surjandari, I., et al., Designing a Permissioned Blockchain Network for the Halal Industry using Hyperledger Fabric with multiple channels and the raft consensus mechanism. Journal of Big Data, 2021. 8(1): p. 10.

[21] XIONG Z H, ZHANG Y, LUONG N C，et al. The best of both worlds: a general architecture for data management in blockchain-enabled Internet of Things[J]. IEEE Network, 2020,34(1):166-173.

[22] Hye-Young P, Xiwei X, et al. Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance. IEEE Access, 2019,7:186091-186107

[23] Koliousis A, Watcharapichat P, Weidlich M, et al. Crossbow: scaling deep learning with small batch sizes on multi-GPU servers[J]. 201912(11): 1399-1412.

[24] Zhao Jian, Qiang Wenqian, An Tianbo, etc. Blockchain consensus mechanism based on random numbers [J]. Journal of Jilin University (Information Science Edition), 2023,41 (02): 292-298. DOI:10.19292/j.cnki.jdxxp. 2023.02.006.

[25] Zhu Jian, Yang Zhenna, Pang Long, etc. Industrial Internet Data Traceability Scheme based on blockchain [J]. Journal of Xi'an University of Posts and Telecommunications, 2022,27 (02): 102-110.