Research on Data Integrity Verification Technology Based on Blockchain

Xiaoshu Wang

Dalian Vocational and Technical College (Dalian Open University), Dalian City, Liaoning Province, 116037
20061215411@dlvtc.edu.cn

Abstract. With the rapid development of Internet of Things technology, it has changed People's Daily production and life to a great extent, but the large number of applications of Internet of Things devices affect the security level of data to a certain extent. In order to ensure data integrity in the Internet of Things environment, it is necessary to conduct in-depth research on blockchain-based data integrity verification technology, analyze the authenticity structure based on data sources, and carry out data integrity verification exploration according to data source identification structure. The experimental system for data integrity verification is proposed. In this way, the effectiveness of blockchain-based data integrity verification technology can be guaranteed, and the integrity of data can be guaranteed under the condition of decentralization, while the scattered storage of data and data privacy can be protected.

1. Introduction

At present, the application of computer technology is becoming more and more common, especially in the era of big data and cloud, computer has become an important basic equipment for processing data and information. However, in the process of using computer to carry out data processing and application, the confidentiality, integrity and availability of data may be damaged due to various factors. The integrity of the data is damaged, it cannot be restored to the original data, resulting in permanent loss of data. This will seriously affect the application value of the data, and may even lead to the loss of enterprises or individuals. In order to ensure the integrity and security of data in the information age, it is necessary to conduct in-depth discussion on the data integrity verification technology based on blockchain and design a data integrity verification scheme based on blockchain. Thus, the data source integrity verification technology is widely used.

2. Research status of data integrity verification technology based on blockchain

Blockchain in the application process can provide inherent data storage space, to ensure that the existing transaction data will not be modified or deleted. Moreover, in the application of blockchain technology for transaction verification, the authenticity and credibility of information data can be analyzed without any third-party authentication. Historical transactions are imtamperable and can ensure the authenticity and effectiveness of the stored data. In the application process of blockchain technology, the application of encryption and digital signature identity authentication and authenticity technology is more common, which can highlight the advantages of block chain read and write access control, and further improve the security of block chain technology in the application process.

At present, with the continuous development of information technology, the application of blockchain technology has gradually matured and improved. The characteristics of non-tampering and...
Decentralization of its data are an important basis for exerting the advantages of blockchain technology. Effective security prevention and management of data collection, transmission and stored process in the Internet of Things environment is an important measure to improve the level of data security. If effective security measures are not taken in the process of data processing and application, data may be lost, leaked or expired. All these problems will have a certain impact on data integrity and harm the application value of data in the Internet of Things. Therefore, it is necessary to effectively apply and utilize the blockchain technology. The unique decentralized and anonymous trust management mechanism of blockchain technology can verify the integrity of data without the need to transfer the complete data to the network server, which is conducive to solving the data integrity verification problem with privacy protection requirements [1].

Figure 1 Block structure diagram

Data integrity verification mechanism is the research focus in the current data management process. Data integrity verification mechanism can ensure the safe storage of local data and the security of remote data at the same time. At present, the research on data integrity verification mechanism is continuously developing, and the more common integrity verification mechanisms include the following: First, some researchers put forward the data ownership proof, which is mainly to complete the integrity verification process of data in the file block. Although this mechanism can effectively identify the integrity of remote data, it cannot effectively identify the availability of remote data. Second, some researchers put forward the theory of data recovery. The recoverable proof mechanism based on sentinel can not only accurately judge whether the remote data is damaged, but also recover the damaged data to a certain degree. Third, open validation model validation scheme. This verification mechanism requires the use of third-party verification institutions to verify the integrity of data, and can also use data structures to complete dynamic data verification. Fourth, the application of random source technology, this technology can not only verify the integrity of the data, but also protect the privacy of the data, can hide the data information that needs to be verified in the cloud server, to prevent third-party verification agencies from obtaining the data content. Fifth, secure cloud storage service architecture DIAAS. This architecture requires a dedicated data integrity service to perform integrity verification. Sixth, the data integrity verification
mechanism based on block chain technology. In the continuous development of blockchain technology, some researchers envision a verification mechanism based on the hashing method. This mechanism is similar to the theory of third-party verification of the integrity of data records. In addition, there are other blockchain data integrity verification methods based on audit tracking method, which pay more attention to the privacy of data and emphasize the non-repudiation of privacy audit logs. At present, some researchers have proposed a data integrity verification framework based on blockchain. In the construction process of the framework, data verification work should be carried out based on the data integrity verification protocol of the own system. In the process of research on this verification scheme, it is necessary to conduct in-depth exploration on the validity of the data integrity verification scheme based on the Internet development environment [2].

3. Data integrity verification mechanism
In the process of further research on blockchain-based data integrity verification mechanisms, it is necessary to understand that the overall goal of proving data integrity is consistent, to ensure that the data is recorded as expected, to ensure that it is the same as the original record in subsequent retrievals. Therefore, the main purpose of data integrity is to prevent abnormal changes in information. If the system can provide an integrity system, it needs to ensure that the data is sent to the receiver from an authenticated data source and that the data has not been tampered with. This ensures that the data is non-repudiation. The integrity verification protocol of data storage is not mainly guaranteed by replication technology, because in the process of data storage, the data integrity may be affected by accidental errors or intentionally damaged data by the data storage party. In the process of designing the data integrity verification protocol, it is necessary to ensure the security of local data storage, remote data storage and cloud data storage. Only in this way can the integrity and authenticity of the data be truly guaranteed. At present, in the research process of data integrity verification mechanism, the P2P storage environment data integrity protocol has been studied deeply. The protocol mainly includes the following contents: during the operation of the protocol, the owner of the data will complete the data replication and store the copy in the data store. The data store and the data owner negotiate the storage time of the data, which is checked periodically by the data validation. In the verification protocol operation, the interactive form of knowledge proof should be used to enable the data storekeeper to prove the stored data protocol to the data verifier, which mainly includes query and response message. The data verifier needs to periodically send query messages to the data store. The data store can calculate the corresponding response message based on the data and the label and send the message to the data verifier. The data verifier makes accurate judgment on the integrity of the data based on the metadata. Metadata is an important part of the validation message, which is generally calculated by the data owner and sent to the data verifier [3].

![Figure 2 Data owner and data validator interaction model](image_url)

In the process of data integrity verification in the cloud information era, two protocols of data ownership proof and recoverability proof are needed. Ensure the reliability of data integrity verification results. The data ownership proof is to use the query response protocol, the user can check the application
of the data, to ensure that the server can have the original data. In order to achieve this goal, a message authentication code symmetric encryption, and the application of other methods are more common, in the file before send to distrust cloud server, user metadata can be stored in the file, the user will be sent to the corresponding file cloud servers, can be sent via metadata and data comparison, the integrity of the data can be validated. The restorability verification method with CAPTCHA can also be used for remote validation of language server data without even needing to recover from the cloud server to the local disk. Proof of data ownership and proof of recoverability can be applied when verifying semi-trusted and untrusted server data [4].

4. Design and implementation of data integrity verification based on blockchain

4.1. Overall system design

In the design process of data integrity verification scheme based on blockchain, the design work should be carried out according to the main characteristics of blockchain. Blockchain has the characteristics of tamper-proof, decentralization and redundant storage. The participating computing nodes are peer to peer, and there is no central node. As a result, can block chain data network as a peer-to-peer network, all nodes can save a copy of the block chain, each node area through the block header information to ensure the consistency of the copy, each block by the hash value and the connection between the previous block can effectively prevent data information is malicious tampering. If you want to modify a block, you must recalculate the hash values of all previous blocks in front of the block, which will greatly increase the complexity of calculation, so as to prevent information tampering. This is the main advantage of the application of blockchain technology in secure storage systems. In addition, the use of blockchain distributed storage system can also prevent a single point of failure. Once the central node is down or fault, the distributed storage can use blockchain to complete the recovery of key data metadata. In particular, the tamper-proof feature of blockchain can ensure the integrity and validity of the data stored on the blockchain to the greatest extent. This is also a necessary condition for the construction of a secure storage system. The specific design framework is shown in Figure 3, which mainly includes Internet of Things devices, blockchain, smart contract, database, etc. First, the database is a traditional distributed storage system, whose main function is to complete the data storage of Internet of Things devices and the corresponding metadata storage. In order to provide reliable support for data integrity verification. Second, blockchain. The main function of the blockchain is to ensure the global state of the blockchain in the data storage area, and all server nodes can query to the blockchain. In this way, the state of the local blockchain is synchronized to ensure that the state of the blockchain is unified. In addition, the metadata stored by the blockchain is read-only. If the data stored by the user changes, new metadata can be generated, and the newly generated metadata can be stored in the blockchain through the smart contract, which can prevent malicious tampering of information data to the greatest extent. Third, smart contracts. Smart contracts allow IoT devices to communicate with the blockchain network. After the smart contract between IoT devices and the production chain, they interact with the database in the overall design architecture network. The corresponding function in the smart contract can complete the registration and deletion of the corresponding Internet of Things devices in the Ethernet address, and can complete various functions such as data storage and data integrity verification [5].
4.2. System development environment design
In the process of designing the data integrity verification scheme based on blockchain, the system development environment must be reasonably selected. In this experiment, the hardware environment is Intel(R)Core(TM) i7-8550U CPU (1.80GHz), and the RAM is 8GB. The experimental running system is Ubuntu16.04LTS 64-bit operating system. The language of smart contracts is solidity. The blockchain environment uses the Solidity programming language, the development framework is Truffle, and the visual private chain client Ganache.

4.3. System function design and implementation
In the design process of integrity verification experimental system based on blockchain, its function design mainly includes two functions: data storage function and data integrity verification function. In the design of the data storage function, the main role of the blockchain node device is to collect various metadata additional information such as the hash value of the data, the time and digital signature of the data, and the owner of the data. So as to construct a new blockchain. When designing the data verification function, the main role of the blockchain node is to complete the request to receive the verification, update the blockchain copy, retrieve the relevant metadata of the given authenticity identification, and return the retrieval result. To ensure the smooth implementation of the data integrity verification scheme. There are two types of entities in the design process of functional modules. In the design of entity components, it mainly includes IoT nodes and blockchain nodes. Function modules include authenticity identification function, data storage function, blockchain consensus function and data integrity verification function. In the specific design process, it is necessary to ensure that IoT nodes correspond to IoT devices in the IoT structure, whose main role is to collect various data in the IoT environment. In the process of functional module design, it is necessary to ensure that the blockchain node can correspond to the peer node in the network system.

The specific functional design includes the following contents: First, the authenticity of the identification function module design. In this module design, it is necessary to ensure that the blockchain nodes can provide the data information collected under the Internet of Things environment. Before the data is stored on the blockchain, this feature needs to be utilized to preprocess the data to build a data structure that identifies the authenticity of the data source. In the application of the data integrity
verification experimental system based on the complete block chain in the Internet of Things environment, the authenticity of the data source and the storage space of the block chain need to be fully considered. In the design process of this functional module, the main content is to decide which data to put on the blockchain and which data to put off the blockchain. Second, the design of data storage function module. In this module design, the data structure that needs to be stored should be provided to the blockchain node, the corresponding script should be used to perform the storage operation, and the validity of the storage operation should be verified, and finally the data information should be stored in the blockchain. This function is similar to the transaction function in the network blockchain. Verification of transactions is a unique mechanism in the blockchain network, and all blockchain nodes must verify transactions accurately. Third, block chain consensus function module design. In this functional design process, the block construction and the block forming function of the chain in the block chain should be completed. All data stores and records must be verified by the consensus function module before they can be assembled into blocks. After validation of the block, the block can be linked to the latest block of the blockchain, which is the core content of the blockchain structure and is similar to the consensus protocol in the network blockchain. Fourth, the design of data integrity verification function module. In the design process of this module, relevant data need to be obtained from the blockchain and compared with the metadata. Thus, data integrity verification results are obtained. Blockchain nodes should provide authenticity identification of acquired data, and use corresponding scripts to perform data acquisition operations to verify the consistency between acquired data and data provided by IoT nodes. This function is similar to the transaction function of block chain technology in the use process [6].

5. Conclusion
In a word, in the design process of blockchain-based data integrity verification system, the corresponding functions must be comprehensively analyzed to ensure the reliability and effectiveness of the data integrity verification system. In addition, after the completion of the design of the data integrity verification system, the operation effect of the whole system needs to be analyzed, and the results of functional test and benchmark test can be statistically compared. The final test results are obtained by executing the corresponding test procedures for different test cases. This ensures the correctness of the data storage and validation to the maximum extent possible, and helps to ensure the integrity of the data. In the context of the continuous development of information technology, it is necessary to further explore and study the data integrity verification system based on block chain. In the
course of future research, it is necessary to improve the efficiency of smart contract storage in the blockchain. Due to the relatively small scale of experimental verification, the specific experimental process can be optimized and improved in the future research. In the design, you can build an application that supports more sophisticated data integrity validation using rich data storage formats. In addition, the evaluation model can be used for in-depth analysis of the security of related programs to further improve the reliability and effectiveness of the data integrity verification system.

Reference
[1] Zhang Hanchen. Research on Data Service Model and Key Mechanism of Internet of Things Based on Blockchain [D]. Beijing University of Technology.
[2] Wang Zhanbing. Research on the Integrity Verification Mechanism of Cloud Database Based on Blockchain [D]. Wuhan University, 2018.
[3] Liu Wenzhuo. Cloud data integrity protection scheme based on blockchain [J]. China New Communications, 2020, v22 (09):130-131.
[4] Guo Ruixue. Research on Network Data Privacy Protection Method Based on Blockchain Technology [J]. Modern Information Science and Technology, 2020, v.4(13):172-174.
[5] Zhang Jiashun, Chen Xiubo. Blockchain-Based Data Integrity Verification Scheme for Cloud Storage [J]. 2021.
[6] Wei Yan. Research on Data Integrity Verification Technology Based on Blockchain [D]. Nanjing University of Posts and Telecommunications.