Research on Distributed Data Sharing System based on Internet of Things and Blockchain

Enchang SUN

School of Information and Communications Engineering, Faculty of Information, Beijing University of Technology, Beijing 100124, China
E-mail: ecsun@bjut.edu.cn

Kang MENG

School of Information and Communications Engineering, Faculty of Information, Beijing University of Technology, Beijing 100124, China
E-mail: 1605378608@qq.com

Ruizhe YANG

School of Information and Communications Engineering, Faculty of Information, Beijing University of Technology, Beijing 100124, China
E-mail: yangruizhe@bjut.edu.cn

Yanhua ZHANG

School of Information and Communications Engineering, Faculty of Information, Beijing University of Technology, Beijing 100124, China
E-mail: zyh2863049377@163.com

Meng LI

School of Information and Communications Engineering, Faculty of Information, Beijing University of Technology, Beijing 100124, China
E-mail: limeng720@bjut.edu.cn

Abstract Aiming at the problems of the traditional centralized data sharing platform, such as poor data privacy protection ability, insufficient scalability of the system and poor interaction ability, this paper proposes a distributed data sharing system architecture based on the Internet of Things and blockchain technology. In this system, the distributed consensus mechanism of blockchain and the distributed storage technology are employed to manage the access and storage of Internet of Things data in a secure manner. Up to the physical topology of the network, a hierarchical blockchain network architecture is proposed for local network data storage and global network data sharing, which reduces networking complexity and improves the scalability of the system. In addition, smart contract and

Received August 15, 2020, accepted February 24, 2021

Supported by the National Natural Science Foundation of China (61901011), the Foundation for University Key Teachers from the Ministry of Education of China (201806545031), the Foundation of Beijing Municipal Commission of Education (KM202010005017), the Project of Enterprise and Business Cooperation (KLA-GIO2019100006124)

distributed machine learning are adopted to design automatic processing functions for different types of data (public or private) and supervise the data sharing process, improving both the security and interactive ability of the system.

Keywords blockchain; smart contract; Internet of Things; data sharing

1 Introduction

In the information age, data as an important strategic resource, is experiencing explosive growth. All walks of life have realized that both the innovative application of enterprises and scientific theoretical research need to be driven by a large number of reliable data as the cornerstone, which is accompanied by a sharp increase in the demand for data sharing. From the perspective of enterprises, data sharing can make full use of the value of data to provide users with personalized services and improve the quality of enterprise services. From the perspective of scientific research, high credibility of data can provide a good guarantee for the rapid development of research fields such as big data analysis and artificial intelligence, etc. However, most traditional data sharing platforms build data centers and manage shared data in a centralized manner, which results in weak data interaction between data owners and risk of data abuse and privacy leakage in data sharing. Furthermore, it affects the enthusiasm of data owners to share and hinders the sharing of data resources and value mining^[1–3].

On the other hand, blockchain technology integrates distributed storage, point-to-point communication, distributed consensus mechanism, encryption algorithm and other technologies to establish a set of trust mechanism through the joint maintenance of ledger by multiple nodes, so as to make ledger data public, transparent, traceable, and not tampered which highly meeting the demand of data sharing^[4]. As the underlying technology of Bitcoin, blockchain solves the problem of Double Spending of electronic currency, thus attracting public attention^[5]. Inspired by Bitcoin, Ethereum allows participants to write smart contracts in the platform by building a mature Turing complete programming language, making the blockchain step out of the limitations of electronic currency application and become a basic technology of distributed architecture^[6].

Based on the distributed trust mechanism of blockchain, researchers have proposed many corresponding data storage and sharing systems and platforms for different applications. In the field of education, Xia, *et al.* built a sharing platform of learning record data among educational institutions based on the blockchain of Hyperledger Fabric, broke the data sharing barriers among various institutions, and solved the problems such as the inconvenient for employers to obtain the learning record data and the lack of learning record data in educational institutions^[7]. Based on the consortium blockchain, Wu, *et al.* designed the smart grid data storage system, and used the smart contract to set the sharing authority to solve the secure storage and sharing of network monitoring data in the smart grid^[8]. Gao, *et al.* proposed a medical data sharing scheme based on blockchain, which is used to realize the sharing of medical research data among research and development enterprise in the medical field, promoting the cooperative research and development of pharmaceutical products among enterprises, and thus reduce the cost of drug research and development among enterprises^[9]. However, these studies are limited to specific applications, and do not take into full consideration to construct the system from the

aspects of data access, storage, sharing and analysis, etc. In addition, they do not take into account the scalability of the blockchain and the matching of the system size and other issues.

In this paper, we propose a distributed data sharing system based on Internet of Things and blockchain. Through the functional architecture of data access, data processing and data application, the hierarchical multi chain structure and the corresponding smart contract covering the automatic access, storage, access and processing of data in the Internet of Things with different permissions and levels are realized. Besides both data privacy and data value mining are taken into account. Finally, a secure and reliable, extensible and automated large-scale Internet of Things data sharing system is formed.

The rest of this article is organized as follows. In Section 2, we introduce the related technologies. In Section 3, the construction of our blockchain data sharing system is described including system functional architecture, system network architecture and the consensus protocol based on our system. In Section 4, the realization of sharing for Internet of Things data through smart contract based on the construction proposed in Section 3 is introduced in detail. Finally, in Section 5, the whole paper is summarized and the conclusion is given.

2 Related Technologies

2.1 Blockchain

Blockchain is a distributed ledger combining distributed storage, P2P, consensus algorithm, encryption algorithm and other technologies. It is characterized by distribution, security and reliability. The blockchain storage is based on data blocks and connected in chronological order. Data blocks are generated by the consensus of all distributed nodes and guaranteed by certain economic incentives that all nodes actively participate in the activities of the blockchain. All nodes in the distributed system are equal, without any special centralized node. Moreover, each node will verify and broadcast data blocks, so as to ensure that malicious nodes will not affect the operation of the entire blockchain system. Blockchain can be divided into public blockchain, private blockchain and consortium blockchain. The public blockchain is called permission-less blockchain. Any organization or individual can participate in consensus and have access to read and write data. The private blockchain is suitable for the internal system of the unit or organization, and the read and write permission of its data is controlled by the organization, which cannot completely solve the trust problem. The consortium blockchain is also called a permissioned blockchain, whose consensus is to be joined by members of the blockchain, whose data reading and writing permissions are set according to the consensus proposal, and the participation of nodes requires the consent of other nodes.

2.2 Blockchain Consensus Algorithm

As a distributed system, blockchain faces the same problem as other distributed systems: How to reach consensus efficiently. This is achieved through the blockchain consensus algorithm, which includes the Proof of Work (PoW), Proof of Stake (PoS), Delegated Proof of Stake (DPoS) and the Byzantine Fault Tolerant (BFT).

The PoW consensus algorithm is the consensus algorithm used in Bitcoin, as well as the consensus algorithm used in many current digital currency systems. Nodes in the blockchain

compete with each other through computing resources and the winner of the competition can get the right to keep accounts and gain system rewards by generating data blocks. However, the competition results in a waste of resources and a relatively long period of consensus, which is not suitable for business applications.

In PoS consensus algorithm, the nodes that occupy a large number of interests in the system are responsible for generating data blocks and ensuring the operation of the blockchain. Only those who have rights and interests in the system can compete for the right to keep accounts which greatly reduce the waste of resources and time for reaching consensus comparing to PoW consensus proposal.

DPOS consensus algorithm, the essence is the voting mechanism. Token holders in the system vote for consensus nodes for verification and billing. As an efficient and flexible consistency algorithm. The DPOS uses stakeholder voting power to resolve consensus issues in a fair and democratic manner.

BFT consensus algorithm is a consistency algorithm based on message passing. Practical Byzantine Fault Tolerant (PBFT) algorithm improves the efficiency and reduces the complexity of original BFT algorithm^[10]. It can be used to build high-availability systems to tolerate Byzantine Fault and have excellent performance in asynchronous environments. This paper adopts this consensus algorithm to improve the usability and performance of the system.

2.3 Blockchain Smart Contract

Smart contract is a special program that can be automatically executed on the blockchain^[11]. Its feature is that the program code and data are all stored on the chain, so it has strong tamper-proof and high degree of decentralization. The smart contract is created and invoked in the form of transactions. The contract program is executed on all nodes in the distributed network, so the failure of any node will not affect the operation of the contract program. And Ethereum and Hyperledger Fabric are currently the mainstream blockchain that supports smart contracts.

2.4 IPFS File System

IPFS (InterPlanetary File System)^[12] is a globally connected distributed file system that combines the advantages of distributed hash tables, fast switching, version control systems, and self-certified file systems. It is content addressable, non-tamper with, and decentralized. When storing a file, IPFS will calculate the file fingerprint according to the file content. When obtaining a file, IPFS will retrieve the file from the storage node according to the file fingerprint and return it to the user after correction. IPFS is divided into private cluster and public cluster. Public cluster refers to the distributed network composed of IPFS nodes of the whole network, and anyone can join the network as a host. Private clusters are limited to use within a group or organization, and nodes with the same swarm-key can participate in the network.

3 Construction of Blockchain Data Sharing System

The data sharing system of the Internet of Things proposed in this paper builds the system functional architecture and network architecture based on the scalability of the system, the security of the data sharing and the value exploration.

3.1 System Functional Architecture

As shown in Figure 1, the functional architecture of the system proposed in this paper is generally divided into three layers. The first layer is the data access layer that gets data and uploads data through data transmission protocol. The second layer is the data service layer that connects the data access layer, receives the uploaded data and carries out data storage, data sharing and data analysis. The last one is the data application layer that is implemented through API interfaces provided by the network service layer. The whole system completes the data collection, data uploading, data storage, data sharing, data analysis and the final data application process based on the functional three layer, forming a complete set of blockchain data sharing architecture. In this three layer architecture, blockchain technology is introduced to constitute the data service layer, so that shared data storage and shared record storage cannot be tampered, which meets the security requirements of data sharing system.

Figure 1 System functional architecture

3.1.1 Data Access Layer

The data access layer mainly realizes the collection of data, which is uploaded through the corresponding data transmission protocols (HTTP protocol, TCP/IP protocol, NB-IOT, etc). Note that the data of different business processes will show different characteristics. Even in the same scene, different data collection methods and equipment will lead to different data formats, sizes and other aspects, so the data show strong heterogeneity. In the Internet of Things, the

data collection methods may be QR Code, RFID, sensors, multimedia, etc, therefore the data access layer needs to design the corresponding data interface according to the data types and formats to complete reliable and agile data access, so that the reliability of the data source of each link is guaranteed.

3.1.2 Data Service Layer

Data service layer is a service platform combined with blockchain, distributed storage via IPFS, artificial intelligence, servers (register server, knowledge server, web server, etc) and interfaces and therefore is mainly responsible for data storage, data sharing, data analysis and interactions with the data access layer and application layer. Specifically, blockchain allows data sharing to run completely in a distributed system without a third party. It's decentralized feature can avoid the problem that the system cannot operate due to the single point failure of the system. It's data tamper-proof feature can ensure the authenticity of shared data and high transparency, which makes the system have the auditability of data sharing records and facilitate the supervision of the data sharing process.

Blockchain, combined with IPFS technology, provides a hybrid, reliable, and distributed storage for raw data, metadata, and shared record data. The raw data is encrypted using symmetric-key algorithm to generate ciphertext and uploaded to the IPFS distributed storage system, while the metadata information and the shared record data produced in the process of data sharing are uploaded to blockchain. It realizes the entire process of data storage and sharing, guaranteeing shared data security and auditing.

Blockchain, combined with artificial intelligence technology provides data analysis functions for the data application layer according to the upper application functions, and returns data analysis results according to the input. According to the distributed characteristics of the system, the distributed machine learning framework based on artificial intelligence algorithm is adopted where the agent in each node server of the blockchain runs local data analysis, and the analysis results are uploaded to the cloud, so as to avoid the consumption of data aggregation and the risk of data leakage during data analysis.

3.1.3 Data Application Layer

The application layer can realize functions through the interfaces provided by the service layer (RestAPI, Webservice and other data interfaces), such as low-level device management, user management, data sharing and query, and data sharing process supervision and other applications, which satisfy the requirements in agriculture, industry, medical treatment, civil aviation and other fields.

3.2 System Network Architecture

As shown in Figure 2, system network proposed in this paper is a two-layer network, consists of local network and global network to be responsible for the storage of local data and sharing of global data respectively. Compared with the single-layer network architecture^[13–15], each organization manages its own local network in the two-layer network architecture and global network will play the role of bridge to connect each local network, reducing complexity of

networking. When a new member wants to join the blockchain network, it only needs to authenticate the identity of the bridge node in the local blockchain network, which greatly improves the scalability of the network.

As is known to all, Internet of Things data is characterized by high concurrency and large volume. Due to the huge number of data records, real-time data writing becomes a bottleneck and query analysis is extremely slow, making it a new technical challenge. The traditional single-layer network does not fully take advantage of the characteristics of the Internet of Things data, and its performance improvement is extremely limited. Therefore, it can only rely on investing more computing resources and storage resources for processing, and the operation and maintenance cost of the system rises sharply. So, this paper proposes a two-layer network architecture to solve the problems effectively. It vertically divides network into global network and local network for different network responsibilities. Global network is used for data sharing, query and deployment of global consortium blockchain to check data compliance and share data. Local networks are used to collect and store Internet of Things data and deploy local consortium blockchain. Data with large storage capacity and sensitive information is stored in the local blockchain and connected to global network through the local consortium blockchain smart contract. This way of deployment based on the specific business scenarios of data sharing, is not only beneficial to reduce network load, global data privacy protection, but also meet the requirements of the parties to the sharing system.

Figure 2 System network architecture

3.2.1 Local Network

The members in the local network form the local consortium blockchain, which is mainly composed of each business node, data processing node and data sharing node. Among them, the same type of business nodes join in the same business subchain, receiving and uploading the

data to the local blockchain and storage system according to the corresponding data format, so the physical isolation of business data storage is achieved. The data processing node is responsible for data analysis and value mining of local blockchain data information as well as the storage of local model parameters, while the data sharing node is used for interactive sharing with other local blockchain networks. By setting up local network and combining distributed storage and data analysis, autonomous storage and analysis of shared data can be realized, which can reduce the pressure of single-layer blockchain network data processing and the risk of shared data reveal.

3.2.2 Global Network

The global network builds a global consortium blockchain among the members of each organization, which is composed of data sharing nodes in each local consortium blockchain and intelligent processing node, responsible for the process scheduling and supervisory record functions of data sharing between each local network in the whole system. Specifically, each local network sends data sharing request through its data sharing node and accesses the data according to the sharing permission set by the data owner. The intelligent processing node in the global network receives the models and parameters of local data analysis on the data processing nodes in each local consortium blockchain. Then this node operates global data analysis and value mining. For network expansion in the global consortium blockchain, when a new member joins the blockchain network, it only needs to authenticate the identity of the data sharing nodes (bridge node) of each local blockchain network over the global blockchain, which greatly improves the scalability of the network.

3.2.3 Blockchain Network Consensus Protocol

In this network, both global and local consortium blockchain adopt the PBFT consensus algorithm. The algorithm flow is shown in Figure 3.

Request: First, the client initiates a Request to the primary node in the blockchain network, which is responsible for generating new blocks to package transactions in the network.

Pre-prepare: The primary node receives the transaction request sent by the client, sorts the transaction collection in the network, packs it and generates a new block, then signs the newly generated block and sends it to other nodes in the network.

Prepare: After receiving the data block, digital signature and other information from the primary node, the node will execute the transaction in order, and the execution result will be signed and broadcast to the whole network.

Commit: The node receives the validation results of other nodes and summarizes them. When a node receives $2f$ (f is the tolerable Number of Byzantine nodes) messages from other nodes and meets the validation criteria, the Commit message will be broadcast to the whole network.

Reply: The primary node collects the Commit message from the nodes. When it receives at least $2f + 1$ Commit messages, the consensus completes and the new block is linked to the blockchain.

Although PBFT consensus algorithm is used in both two layer, but we have made some changes in it's implementation. In the local consortium blockchain, the primary node and the data sharing node is the same physical entity. Here, we select nodes with relatively high configuration (computing power, storage resources, etc) as the host node to prevent the occurrence of limited local network throughput due to the poor performance of the primary node. In the global chain, for higher reliability, the primary node is elected by all nodes in the global network to package and broadcast the transactions in the network.

Figure 3 PBFT consensus algorithm

4 Data Sharing Based on Smart Contract

4.1 Smart Contract Design

In the system functional architecture, the three-layer network architecture proposed adopts the storage mode of blockchain plus IPFS to realize the safe storage of shared data, which meets the security requirements of the system. In the system network architecture, two-layer network architecture design is adopted to reduce the complexity of networking and verification of new members, which meets the scalability requirements of the system. And in this section, the intelligent and automatic data sharing function of the system is realized by using the flexibility and programmability of smart contract based on the system architecture proposed above. The proposed smart contract adapted to the system network architecture is shown in Figure 4, which is composed of global consortium blockchain contract and local consortium blockchain contract. The local consortium blockchain contract is deployed in the local consortium blockchain node to realize the local raw data storage management function. It consists of data member management smart contract (DMMC) and local data storage smart contract (LDSC). Global consortium blockchain contracts are deployed in global consortium blockchain nodes to implement data sharing logic, consisting shared member management smart contract (SMMC) and shared data management smart contract (SDMC).

Figure 4 Smart contract structure

DMMC is used to record the data user identity (DU-ID), the corresponding public key (PubKey), and the mapping of LDSC in the Local chain. The data provider can be a real user or a smart device in the Internet of Things. During system initialization, information about existing members is stored in the contract.

LDSC is used for storage protection, update, query and other functions of local data, including data storage smart contract (DSSC) and data update smart contract (DUSC).

- DSSC is used to store metadata information for local raw data, including data identity, data owner, data fingerprint (hash value), privacy level (public or private), IPFS address, creation time, last modification time, and so on. The data identity is the unique identity of the data related to the data entity. The data owner is the DU-ID in the DMMC, and the data fingerprint is the hash value of the raw data.
- DUSC is used to update data, and only the data owner has permission to do so, so the contract first verifies that the user is the data owner and then modifies the data fingerprint, IPFS address, and last modification time in the metadata information.

SMMC is used to record the shared user identity (SU-ID), the corresponding public key (PubKey), and the shared data management smart contract (SDMC) mapping for each organization in the global blockchain data sharing business.

SDMC realizes the storage management function for data information in global chain, including data digest storage smart contract (DDSC), shared permission control smart contract (SPCC), and shared records monitoring smart contract (SRMC).

- DDSC is used to record data summary information for sharing in the local consortium blockchain for viewing by other organizations data sharing nodes, including data identity,

data owner, creation time, and data description information. The data fingerprint is the hash value of the raw data.

- SPCC is used to control the access rights of the data and store the access permission list information of the data.
- SRMC is used to keep authorized access records of each organization member, including data identity, sharing user identity and sharing time, etc. With the supervisory function in data sharing logic, it can effectively avoid the occurrence of data abuse.

4.2 Implementation of Data Processing Function Based on Smart Contract

4.2.1 User Registration

In the whole system, user identity is the basis of data protection and sharing on blockchain. The user registration can be divided into the local network data-user identity registration and the global network sharing-user identity registration. The registration process is shown in Figure 5, where a new user (data user or shared user) initiates a registration requests, and then enrolled by the certificate authority (CA) issues the identity certificate and generate the public-private key pair for the user. Public key is stored in the DMMC or SMMC. When the storage is complete, the user's unique identity ID (DU-ID or SU-ID) in the contract is returned to the user with the private key for local storage.

Figure 5 User registration process

4.2.2 Data Access and Storage

Correspond to different layers of the network proposed in this paper, data access is divided into local data access and global shared data access.

In the process of data sharing in the Internet of Things, the smart contract on the local consortium blockchain and the global consortium blockchain is used to cooperate with the IPFS private cluster to collect and store the data of the Internet of Things on the blockchain, so as to prevent it's content from being tampered and destroyed. The whole Internet of Things data is saved in a private IPFS cluster under the blockchain in the form of DocJSON, and the file fingerprint is generated and stored in the contract on the blockchain, as shown in Figure 6, which is the JSON data saved on the chain. The Figure 6(a) is the local data store in local network, including data header about the metedata of row data and dead body related to IPFS. The Figure 6(b) is data synopses stored in global consortium blockchain for sharing.

```
{
  "File_Header" : [
    "Data_ID" : "aj07411101000",
    "Version" : "1",
    "Timestamp" : "20200314",
    "Last modifide" : "20200415",
    "Owner" : "Tom",
    "Privacy level" : "public"
  ],
  "File_Body" : [
    "Hash" : "e150a1ec81e8e93e1eae2c3a77e66ec6dbd6a3b460f89c1d08aecf422ee401a0",
    "IPFS" : "QmQU2gS4gZ7Tp1TECjDUxdQFd9bBBEWxDxPPfhLfYHVuei"
  ]
}
```

(a)

```
{
  "Data_ID" : "aj07411101000",
  "Timestamp" : "20200314",
  "Description" : "关于行李运输安检环节上链数据"
  "Owner" : "Tom",
}
```

(b)

Figure 6 (a) Local data in JSON. (b) Shared data synopses in JSON

The local data access process designed in this paper is shown in Figure 7. The data user initiates a data up-link request to the local consortium blockchain and uploads the raw data and the locally saved digital identity information of the user. The local consortium blockchain calls the DMMC to query the user's public key information according to the user's digital identity. If the user's public key information does not exist, the up-link request fails, otherwise, the information of user's public key is returned. After querying the public key information, the public key is used to encrypt the raw data to generate ciphertext uploaded to the IPFS system, and then obtain the IPFS stored hash value address. Finally DSSC is used to initiate a transaction to store the metadata information and return the stored results after building metadata information containing the raw data, data identity, data owner, data fingerprint (hash value), privacy level (public and private), IPFS address, creation time, and last modified time.

Figure 7 Local data upload process

Figure 8 Data synopses upload process

The shared data access process designed in this paper is shown in Figure 8. The shared user accesses the local consortium blockchain through the data sharing node, obtains the metadata information uploaded locally and determines the privacy level of the data. Then data summary information will be constructed for the data with the data privacy level of public, including

data identity, data owner, creation time and data description, and then it accesses DDSC to initiate a transaction to store the data summary information. After that, the shared users $U_1, U_2, \dots, U_n$, are confirmed to the shared permission list $[(U_1, \text{SU-ID1}), (U_2, \text{SU-ID2}), \dots, (U_n, \text{SU-IDn})]$, and finally the shared user access SPCC to initiate a transaction to store the shared permission list in the blockchain.

4.2.3 Data Sharing and Acquisition

We have designed and implemented the sharing and acquisition functions of the data in blockchain, as shown in Figure 9. First of all, the shared user A initiates a sharing request to the data with the specified Data-ID in the global consortium blockchain and uploads its digital identity information. The global consortium blockchain accessing SPCC determines whether shared user A is in the data share list. If it exists, the Data-ID is sent to shared user B, who accesses the local consortium blockchain to obtain information about the Data-ID data. The local consortium blockchain obtains the metadata information of raw data from DSSC, particularly its IPFS address, then gets the public key of the data user from DMMC and returns it to the shared user A. The shared user A obtains the ciphertext of the raw data through the IPFS address value, and uses the public key of the data user to decrypt the ciphertext to obtain the raw data. After successful sharing, the shared user A accesses the SDMC to store the shared record of this article data identity, the shared object identity, shared time.

Figure 9 Data sharing and acquisition process

4.2.4 Data Privacy and Analysis

For private data on the local blockchain, our system employs distributed machine learning to provide private data analysis function within the data application layer. The collaborative training over machine learning can be done by configuring agents on local network data process-

ing nodes and the global network intelligence processing nodes. First, the local training model parameters are uploaded to the global network after the data processing node learns the local strategy for the data, and then the intelligent processing node trains and updates the global model. In this way, the learning ability can be improved with a wider range of samples, so as to achieve the improvement of optimized performance.

The whole system process is shown in Figure 10.

Figure 10 Analysis process based on distributed machine learning data

The first stage: The intelligent processing node in the global network generates the initial model and a pair of public and private keys according to the data analysis requirements, and stores the initial model and public key in the global consortium blockchain.

The second stage: Each local data processing node obtains the initial model and public key in the global consortium blockchain through the data data sharing node, stores them in the local consortium blockchain, and completes the initialization on the agent.

The third stage: The data processing nodes train the initial model based on the local stored data, encrypt the model after the local training with the public key and transfer it to the local blockchain for storage.

The fourth stage: The data sharing node gets the training model in the local consortium blockchain and transfers it to the global consortium blockchain.

The fifth stage: The intelligent processing node gets the model parameters transmitted by each local consortium blockchain. After decrypting with the private key, it calculates the weighted average value of all model parameters according to the size of the training data of the participants, updates the initial model parameters and transmits them to the global blockchain.

The sixth stage: The data processing node in each local consortium blockchain gets a new training model, updates the local training model, and performs the next iteration until the model converges to get the final training model.

5 Conclusion

This paper proposes a distributed data sharing system based on the Internet of Things and the blockchain technology. With the blockchain and IPFS technology, it provides a mixed, reliable and distributed storage mode for the raw data, metadata and shared record data to obtain the overall trusted interactive ability of the system. The local storage and global sharing of data in the network are realized by using the hierarchical network architecture to improve network scalability. On this basis, the consortium blockchain is combined with smart contract, distributed machine learning technologies to complete automatic processing functions for different types of data (public or private) and supervision for data sharing process, improving the privacy protection ability of system data sharing, and obtaining data sharing value.

References

- [1] Elhoseny M, Abdelaziz A, Ahmed S. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. *Future Generation Computer Systems*, 2018, 78: 641–658.
- [2] Yao X. Design and implement of share data center in digital asian-pacific region. 2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Macau, 2016: 182–186.
- [3] Zhao Q. Resource sharing strategy of regional scientific and technological innovation based on big data analysis platform. 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Changsha, 2018: 134–138.
- [4] Swan M. *Blockchain: Blueprint for a new economy*. USA: O'Reilly Media Media Inc, 2015: 35.
- [5] Nakamoto. Bitcoin: A peer-to-peer electronic cash system. <http://bitcoin.org/bitcoin.pdf>, 2008.
- [6] Vitalik B. A next-generation smart contract and decentralized application platform. <http://bitcoin.org/bitcoin.pdf>, 2015.
- [7] Xia H F, Xu Q. Learning record sharing account book construction with hyperledger fabric technology. *Chinese Modern Electronics Technique*, 2020, 43(2): 80–83.
- [8] Wu Z Q, Liang Y H, et al. Secure data storage and sharing system based on consortium blockchain in smart grid. *Chinese Journal of Computer Applications*, 2017, 37(10): 2742–2747.
- [9] Gao M J, Wang H Q. Blockchain-based searchable medical data sharing scheme, *Chin. Journal of Nanjing University of Posts and Telecommunications: Natural Science Edition*, 2019, 39(6): 94–103.
- [10] Castro M, Liskov B. Practical byzantine fault tolerance. *Proceedings of the Third Symposium on Operating Systems Design and Implementation (OSDI)*, (USENIX, New Orleans), 1999.
- [11] Mohanta B K, Panda S S, Jena D. An overview of smart contract and use cases in blockchain technology. 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, 2018: 1–4.
- [12] Kumar R, Tripathi R. Implementation of distributed file storage and access framework using IPFS and blockchain. 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019: 246–251.
- [13] Huang Z, Su X, Zhang Y, et al. A decentralized solution for IoT data trusted exchange based-on blockchain. 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2017: 1180–1184.
- [14] Liang W, Tang M, Long J, et al. A secure fabric blockchain-based data transmission technique for industrial Internet-of-Things. *IEEE Transactions on Industrial Informatics*, 2019, 15(6): 3582–3592.
- [15] Liu B, Yu X L, Chen S, et al. Blockchain based data integrity service framework for IoT data. 2017 IEEE International Conference on Web Services (ICWS), Honolulu, HI, 2017: 468–475.