Blockchain based Secure Data Sharing Model

Shi Wang
College of Computer Science
Inner Mongolia University
Hohhot, China
781240654@qq.com

Jing Liu
College of Computer Science
Inner Mongolia University
Hohhot, China
liujing@imu.edu.cn

2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD) | 978-1-7281-6597-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/CSCWD49262.2021.9437751

***Abstract*—There are mainly three traditional data sharing methods. The first is the most direct data copy, the second is to share data based on a data sharing protocol, and the third is to share data through a data center. These methods have a common feature, that is, the data requester will get the data of the data owner. This may cause serious problems in data security, such as data leakage and data abuse. As a data center is a centralized organization, there are risks such as data loss and data tampering. In addition, various countries have also issued a series of policies on data security issues, such as the GDPR implemented by the European Union in 2018. The blockchain technology using a decentralized model can be used as a new attempt to solve the above problems. This paper studies a data sharing scheme based on blockchain, and proposes a model that combines the Ethereum blockchain and federated learning ideas, and uses off-chain storage methods to share data. In this model, users can upload data description information to the blockchain through smart contracts, and can retrieve the required data through keywords, and then send the data identification and data processing model to the data owner in the form of transactions. The data owner can use this model to process the data, and finally return the result to the data requester. Because the data owner is in full control of his data and does not expose the source data to the outside, the use of this model for data sharing can effectively avoid problems such as data leakage, data loss, and data abuse.**

Keywords—data sharing model, blockchain, ethereum, smart contract

I. INTRODUCTION

With the rapid development of the Internet today, technologies such as mobile communications, social networks, and sensors not only connect humans, the virtual world, and the real world with each other, but also produce massive amounts of data at all times. These data records are eventually collected by service providers and stored on their own devices. Enterprises can gain valuable knowledge by mining and analyzing these data. Therefore, both for industry and academia, data has become increasingly important and has become an important intangible asset [1]. Although enterprises produce and accumulate huge amounts of data in the process of production, operation and management, but because they have different data, their own data cognition is limited, so for each organization, the data is also scarce. Therefore, in various fields such as scientific research, financial risk control, precision marketing, insurance pricing, and health care, there is a strong demand for data sharing among various organizations.

Although there is a need for data sharing among different agents, behaviors of data sharing is not very active. This is mainly because the traditional methods of data sharing among enterprises cannot solve the data security risks faced by the data sharing process. In traditional data sharing behaviors (such as enterprise data buying or selling or signing a data confidentiality agreement between companies), data requesters directly obtain the data of the data owner and then process it, which may result in data leakage and Risk of data misuse. These data sharing behaviors are private activities among enterprises, and it is difficult for the supervisory authorities to supervise them. As a result, the supervisory authorities cannot stop the non-compliant data sharing in advance, and it is difficult to accurately determine the liability afterwards. As a result, the losses suffered by the enterprise cannot be comprehensive make up. And relying on a third-party data center for enterprise data sharing requires that the company fully trust the third-party data center. However, the reliability of third-party data centers is difficult to accurately assess, and centralized data centers may also experience downtime and firewalls being compromised. Once the data is lost, it will bring great loss and harm to the enterprise.

Some of the characteristics of blockchain technology can be used as new solutions to try to solve the above problems. Blockchain technology was originally the core supporting technology of Bitcoin, a digital currency. It is a chain data structure composed of a set of data blocks connected in series by hash pointers. With the widespread promotion of Bitcoin, the blockchain technology has also developed rapidly, and the Ethereum platform has since appeared. The biggest feature of Ethereum is that it supports Turing complete programming language, which can write logic complex code that runs on the chain. This brings a wealth of possibilities for the application of blockchain. Blockchain technology uses data encryption, time stamping, and distributed consensus to implement trusted, point-to-point transactions, coordination, and collaboration in untrusted distributed systems. At the same time, the blockchain built on the P2P network is very easy to expand, and the data stored on it is also highly redundant, difficult to tamper with, and difficult to forge.

Therefore, according to the demand for data sharing between enterprises and the technical characteristics of the blockchain, this paper proposes an enterprise data sharing model EntDSM based on the alliance chain.

The main contributions are as follows:

1) Explore the feasibility of combining data sharing between enterprises with blockchain technology, and implement an enterprise data sharing model based on the alliance chain.

2) It implements data on-chain and data retrieval through the EThereum platform and smart contracts, making it easy for business users to find the data they need.

3) And adopt the idea of federal learning to achieve the purpose of invisible data availability, that is, the data transmitted between enterprises is not the data itself, but the parameters of the model required to process the data, so that the data is not local, data sharing can be achieved, effectively This protects the data privacy of enterprise users, and also reduces the risk of data leakage and data abuse.

4) Finally, all data sharing records are stored on the chain, which is convenient for the supervision department to audit and supervise, and to avoid unnecessary losses brought to the enterprise by illegal operations.

The rest of the paper is organized as follows, the section II is related work, and the section III introduces the system model. The section IV analyzes the security of the system model, and finally gives conclusions and future work directions.

II. RELATED WORK

There have been some researches on data sharing based on blockchain. [7] describes a decentralized personal data management system to ensure that users own and control their data. T_{access} and T_{data} transactions are used to transfer instructions, such as storing, querying, and sharing data. Literature [8] proposed a distributed electronic medical record management system based on Ethereum blockchain technology-MedRec. It integrates distributed medical data stored in databases of different organizations on the chain, and verifies, authorizes, and shares patient data through on-chain smart contracts. Liang Xueping et al. [9] proposed a user-centric health data sharing solution, using a decentralized permission chain to protect privacy, and using blockchain-based membership services to strengthen identity management. [10] proposed a framework for managing and sharing cancer patient care EMR data to ensure privacy, security, availability and fine-grained access control of EMR data. Xia Qi et al. [11,12] proposed a system based on blockchain technology to solve the problem of sharing medical data in a environment of lack of trust (MeDShare), providing a data source for medical data shared in cloud storage between big data entities, audit and control services. Document [13] proposed a personal health information sharing (BSPP) scheme based on blockchain security and privacy protection for the diagnosis and improvement of electronic medical systems. [14] proposed a blockchain-based information management system, MedBlock, to share patient information. The improved consistency mechanism achieves the consistency of EMRs without increasing energy consumption and network congestion. [15] proposed a blockchain-based framework for the secure, interoperable and efficient access of medical records by patients, providers and third parties, while protecting the privacy of patient sensitive information.

Most of the above researches focus on access control and data sharing in the field of medical and health. In addition, there are also some data sharing researches on blockchain technology in other fields. Enigma [16] implemented a secure multi-party computing platform based on blockchain technology. Its data is stored in a distributed manner, and no one node can completely access the data. [17] proposed a blockchain-based IoT design, which brought distributed access control and secure data sharing. Huang Longxia et al. [18] proposed a blockchain-based data storage sharing mechanism and introduced an efficient public auditing scheme to provide users with secure data sharing services. Reference [19] stored all the data on the blockchain and set up two smart contracts on the chain. T_s is used to upload data from the data producer to the blockchain, and T_q is used for data consumption. The user queries the data, and finally performs authorization and data sharing through smart contracts. [20] proposed a blockchain-based data sharing method that enables data owners to control the anonymization process, thereby improving the security of services. Literature [21] proposed a solution for the data storage and sharing of the decentralized storage system by combining the framework of the decentralized storage system, the interstellar file system, the Ethereum blockchain and ABE technology. Reference [22] proposed a data access control and sharing model using blockchain, using ABE to control and share enterprise data.

In the above research, in order to avoid the huge data on the blockchain, two solutions are usually used: one is to apply parallel blockchain technology [23], such as the double-chain structure described in reference [24]; the other is to use The blockchain is associated with a distributed hash table [25], enabling off-chain storage. These studies mainly focus on individual-to-enterprise data sharing, and lack of research on data sharing scenarios between enterprises. At the same time, these studies have not made in-depth exploration on data consumer query and retrieval. If companies cannot find the data they need, they cannot share data. Therefore, research on data retrieval is also very important. Finally, although the data is encrypted in the process of saving or transmitting, when using the data, the data requester can obtain the original data, which brings the risk of data leakage and data abuse.

III. SYSTEM MODEL

The EntDSM model proposed to solve data sharing between enterprises includes two roles:

Supervision department: It is composed of relevant government departments. It is mainly responsible for reviewing the qualifications of on-chain enterprises and whether the data sharing behavior and content between enterprises are legal and compliant.

Enterprise users: Enterprise users are divided into data owners and data requesters. The data owner refers to the company that owns the data and can publish the data he wants to share on the chain; the data requester refers to the company that requests the data and can find the data he needs on the chain through the data retrieval function.

The blockchain-based enterprise data sharing model is shown in Figure 1.

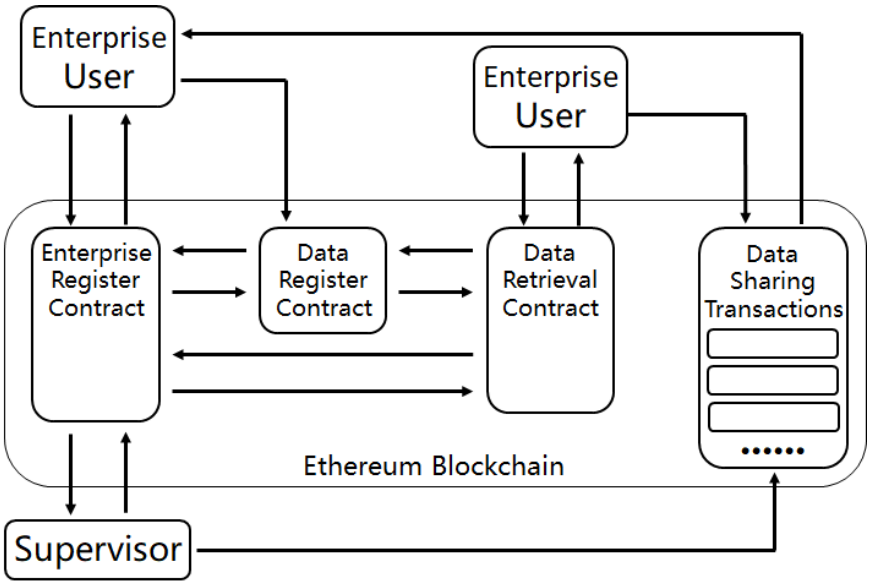

Figure 1. Enterprise Data Sharing Model

The data sharing process involved in Figure 1 is as follows:

1) The supervisory authority deploys two smart contracts on the Ethereum blockchain, namely the enterprise registration contract, the data registration contract and the data retrieval contract.

2) The enterprise user first completes the registration using the enterprise registration contract.

3) The supervisory authority reviews the qualifications of the enterprises that have submitted registration applications through other methods outside the chain, and modifies the status of the enterprises on the chain.

4) An approved business can associate multiple addresses in the business registration contract.

5) Successfully registered business users can use the associated address to upload data related information in the data registration contract.

6) Successfully registered business users can find the required data through data retrieval contracts.

7) The data requester can send the model and parameters for processing the data to the data owner in the form of a transaction.

8) The data owner can read the data requester's request from the blockchain, process the data with the model and parameters provided by the data request, and then send the result to the data requester as part of the transaction in the same transaction mode.

9) The supervisory authority may review the data sharing behaviors between enterprises at any time, take accountability for illegal operations, and modify the status of registered enterprises through enterprise registration contracts.

The above process is mainly realized by three only contracts, which are a business registration contract, a data registration contract, and a data retrieval contract. The following describes the relevant interface and algorithm logic of each smart contract.

A. Enterprise Register Contract

The smart contract is deployed by the regulatory authority, and some variables are initialized during deployment.

The supervisorAddr variable of the address type, which stores the address of the regulatory agency.

Companies variable of structure array type, which is a list of company information. The enterprise information is defined by a structure containing the enterprise name (string), the enterprise address (address), and the enterprise status (enum).

Enterprise registration contracts have two main functional interfaces:

comInfoReg(comName): Enterprise information registration method. This method can only be called by enterprise users. Each time a business user calls this interface, a business name is passed in, and this method associates the business name with the address and saves it to the Companies business list. The input of the algorithm1 is a string——the name of the enterprise, and a boolean value is output to indicate whether the enterprise is successfully registered. New users will register successfully, and

registered users will fail to register again.

changeComState(comName,newComState): This function can only be performed by the regulatory authorities，otherwise it will throw an exception .When the registration of a new business user is completed, the regulatory authority first needs to review its qualifications, and then modify the status information of the business. Enterprise users can perform other operations only when the status information of the enterprise is "Audited".This function is very important ， because subsequent operations require status information of enterprise , such as data registration and data retrieval.

Algorithm 1 comInfoReg

```
Input: comName
Output: bool
1  if msg.sender is supervisorAddr then
2      throw;
3  end;
4  if mag.sender is in Companies.comAddr then
5      return false;
6  end;
7  if comName is not in Companies then
8      Companies.push(comName, mag.sender, ToAudit);
9      return true;
10 end;
11 if comName has exist then
12     Companies[comName].comAddr.push(msg.sender);
13 return true;
14 end;
```

Algorithm 2 changeComState

```
Input: comName
Output: bool
1  if msg.sender is not supervisorAddr then
2      throw;
3  end;
4  if comName has exist && Companies[comName].state
not equal newComState then
5      Companies[comName].state <= newComState;
6      return true;
7  else
8      return false;
9  end;
```

B. Data Register Contract

The contract is deployed by the regulatory authorities to save data descriptions of successfully registered companies. The data description is a 6-tuple,

DataDescription = (dataOwner, dataDomain, dataName, dataAttr, dataSize, dataHash).

Their meanings are:

- dataOwner: address information of the data owner;
- dataDomain: data source domain;
- dataName: the name of the data;
- dataAttr: various attributes of the data;
- dataSize: data size, that is, the number of data items;
- dataHash: hash of the original data.

The data registration contract has two main functional interfaces:

updateDatabase(DataDescription): method used by business users to upload data descriptions. An enterprise whose status is reviewed can call this method to upload descriptive information of enterprise data.The input of the algorithm3 is the data descriptions, including the above 6 variables(dataOwner, dataDomain, dataName, dataAttr, dataSize, dataHash). When the contract is called by an authorized user and the data registration is completed, the algorithm returns true, otherwise it returns false.

Algorithm 3 updateDatabase

```
Input: DataDescription
Output: bool
1  msg.sender has audited then
2      database.push(DataDescription);
3      return true;
4  else
5      return false;
6  end;
```

getIndexList(caller,keyword): gets a list of index positions containing data descriptions of keywords and returns them to the method caller.

C. Data Retrieval Contract

The user can obtain the data description of the required data through the data retrieval contract, and then can send data request transactions to the data owner through the information in the data description. The result variable in the contract is used to store search results.Similarly, data retrieval contracts can only be called by authorized users. The input of algorithm4 is the caller keyword, and the returned keyword list is called by algorithm5. Algorithm5 compares the information in the data registration module with the keywords and returns the search result.

Algorithm 4 getIndexList

```
Input: caller keyword
Output: indexList
1  if caller hasn't audited then
2      throw;
3  end;
4  for i := 0 to database.length do
5      for item in database[i] do
6          if item.contains(keyword) then
7              indexList.push(i);
8              continue;
9          end;
10     end;
11 end;
12 return indexList;
```

Algorithm 5 search

Input: keyword
Output: result
1 indexList <= DataRegister.getIndexList(msg.sender, keyword)
2 for i = 0 to indexList.length do
3 result[i] <= DataRegister.Companies[indexList[i]];
4 end;
5 return result;

IV. Case Study and Security Analysis

In this section, through specific case analysis, we demonstrate how to share data in our EntDSM, and analyze its advantages. The whole process includes business registration, data registration, data retrieval and data sharing. Suppose EC is an e-commerce enterprise and BK is a bank. EC and BK want to share data to promote each other's business. It is not feasible to use traditional enterprise data sharing methods, because BK's industry makes its data very sensitive and cannot be directly exposed to the EC for processing. If EntDSM is used, the EC can complete the data processing without obtaining BK data. The specific steps are as follows.Our model is easy to extend. This example only introduces the situation of two users for the sake of concise description. The participation of multiple users in data sharing can also be achieved through the following process.

Enterprise registration. The process of enterprise registration is described in algorithm *comInfoReg* and algorithm *changeComState*. Specifically, EC and BK first need to call the *comInfoReg*(comName) method to register and obtain the initial state *comState.ToAudit*, then the supervisor verify the enterprise information through other ways to ensure the registration is true and effective, and call the *changeComState*(comName, newComState) method to modify the enterprise state to *comState.Audited*. Enterprise registration contracts save enterprise information on the blockchain, but only the supervisor can view it. Through the enterprise registration process, some malicious nodes can be shielded.Of course, supervisor can dynamically manage the enterprise state. If some companies violate the regulations, they can temporarily cancel their authorized user status.

Data registration. After BK successfully completes the enterprise registration, it calls the *updateDatabase* (DataDescription) method described in Algorithm *updateDatabase* to upload its data description information. The data description contains only metadata and no raw data. Therefore, the raw data is isolated from the blockchain and other enterprise users, avoiding the risk of data leakage. Moreover, even though the raw data may continue to grow, the metadata will not change, so that the blockchain will not become too heavy and become a slow database.

Data retrieval. After the EC passes the audit, it can deploy a *DataRetrieval* contract and call the *search*(keyword) method described in the algorithm *search* to find the data description information uploaded by BK in the *DataRegister* contract. Through the metadata information in the data description and the EC's own needs, the EC can design a model for processing the data.

Data sharing. After the EC completes the design of the data processing model, the model can be encrypted by the encryption function and the public key of the data owner in the data description, then sign the message, and send it to the data owner BK in the form of a transaction. After receiving the transaction, BK can first verify whether the signature is correct, and then use its private key to decrypt the message to obtain the data processing model sent by the EC. BK can now review the model and judge its security. If the model is secure, the data can be processed locally and the results returned to the EC in the same way.

During the entire data sharing process, the EC cannot obtain the raw data of the BK, but the EC can still process the data of the BK. The effect of making the data invisible is achieved. At the same time, data sharing is recorded on the blockchain, which greatly facilitates the audit and supervision work of the supervisor.

In our model, we combine the Ethereum blockchain, off-chain storage technology, and federated learning ideas to implement a data sharing model between enterprises. Next, we will discuss its benefits, security, and privacy.

(1) Data Owners Control Their Own Data

In this model, the data requester can only see the related description of the data, and cannot obtain the original data or desensitized data, thereby achieving the purpose of invisible data availability and reducing the risk of data leakage and data abuse. In the traditional enterprise data sharing scheme, we need to assume that the data requester will honestly use the data according to the agreement to ensure data security. In this model, because the data is completely controlled by the data owner, there is no need to worry about whether the data requester is honest. At the same time, there are regulatory authorities on the chain to ensure that the interests of the data owner are not infringed.

(2) Avoid Single Point of Failure

In this model, compared with the traditional third-party data center sharing model, the distributed nature of the blockchain is used to avoid the single point of failure. At the same time, the data on the blockchain has the characteristics of high redundancy, forgery and tampering, which further ensures the reliability and availability of the data. As long as the number of malicious nodes on the chain does not exceed 51%, the model can run normally.

(3) Avoid Protocol Loopholes

With the increasing volume of data and the increasing complexity of data processing, it is difficult for data sharing protocols to be comprehensive.Even though data requester will honestly use the data according to the agreement,vulnerabilities in the agreement may still lead to data abuse and disclosure.In this model，We don't need to worry about this problem. To ensure the safety,what we need are concise and correct contracts and reliable regulatory authorities.

V. Conclusion and Future Work

Existing enterprise data sharing technologies allow data requesters to obtain data from data owners, whether raw data or desensitized data, will inevitably bring risks to the use of data. And the blockchain technology, which has developed rapidly in recent years and has many excellent features, has brought new solutions to the problem of data sharing among enterprises. The flexible scalability of the blockchain makes it easier for business users to find and discover the data they need. The data on the blockchain is highly redundant, difficult to tamper with, and difficult to falsify, so that the data sharing behavior between enterprises is permanently stored in the blockchain, bringing great convenience to future review and evidence collection. Data can be shared in an invisible way, which further protects the data security of the data owner and avoids the risk of data leakage and data abuse.

Although data can be shared securely between enterprises, each time the data owner shares the data comes with a corresponding cost, and this model does not cover the specific method of how to make up for this cost, which will be our next research direction.

Acknowledgment

This work was supported in part by the Inner Mongolia Natural Science Foundation (No.2019MS06007), Inner Mongolia Science and Technology Plan Project (No.2020GG0187), Inner Mongolia Engineering Laboratory for Cloud Computing and Service Software and Inner Mongolia Key Laboratory for Social Computing and Data Processing.

References

[1] K Schwab, A Marcus, JO Oyola, W Hoffman, and M Luzi. Personal data: The emergence of a new asset class. In An Initiative of the World Economic Forum, 2011.

[2] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.

[3] Buterin V. A next-generation smart contract and decentralized application platform[J]. white paper, 2014.

[4] Wood G. Ethereum: A secure decentralised generalised transaction ledger[J]. Ethereum Project Yellow Paper, 2014, 151.

[5] Androulaki E , Barger A , Bortnikov V , et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains[J]. 2018.

[6] Yuan Yong, Wang Fei-Yue. Blockchain: the state of the art and future trends. Acta Automatica Sinica, 2016,42(4): 481−494

[7] Zyskind, Guy, and Oz Nathan. "Decentralizing privacy: Using blockchain to protect personal data.“ In Security and Privacy Workshops (SPW), (2015) IEEE, pp. 180-184.

[8] Ekblaw A, Azaria A, Halamka JD, MD, Lippman A. A case study for blockchain in healthcare: “MedRec” prototype for electronic health records and medical research data. Technical Report, 5-56-ONC, Massachusetts Institute of Technology, 2016. https://www.healthit.gov/sites/default/files/5-56-onc_blockchainchallenge_mitwhitepaper.pdf

[9] Liang X , Zhao J , Shetty S , et al. Integrating Blockchain for Data Sharing and Collaboration in Mobile Healthcare Applications[C]// The 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC 2017). IEEE, 2017.

[10] Dubovitskaya A , Xu Z , Ryu S , et al. Secure and Trustable Electronic Medical Records Sharing using Blockchain[J]. AMIA. Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2017, 2017.

[11] Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, ‘‘BBDS: Blockchain-based data sharing for electronic medical records in cloud environments,’’ Information, vol. 8, no. 2, p. 44, 2017.

[12] Xia Q , Sifah E B , Asamoah K O , et al. MeDShare: Trust-less Medical Data Sharing Among Cloud Service Providers Via Blockchain[J]. IEEE Access, 2017:1-1.

[13] Aiqing Z , Xiaodong L . Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain[J]. Journal of Medical Systems, 2018, 42(8):140-158.

[14] Kai F , Shangyang W , Yanhui R , et al. MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain[J]. Journal of Medical Systems, 2018, 42(8):136-147.

[15] Dagher G G , Mohler J , Milojkovic M , et al. Ancile: Privacy-preserving Framework for Access Control and Interoperability of Electronic Health Records Using Blockchain Technology[J]. Sustainable Cities and Society, 2018, 39.

[16] Zyskind G, Nathan O, Pentland A. Enigma: Decentralized computation platform with guaranteed privacy. 2015. https://enigma.co/enigma_full.pdf

[17] Shafagh H , Burkhalter L , Hithnawi A , et al. Towards Blockchain-based Auditable Storage and Sharing of IoT Data[J]. 2017.

[18] Huang L , Zhang G , Yu S , et al. SeShare: Secure cloud data sharing based on blockchain and public auditing[J]. Concurrency and Computation: Practice and Experience, 2017:e4359.

[19] L. Yue, H. Junqin, Q. Shengzhi and W. Ruijin, "Big Data Model of Security Sharing Based on Blockchain", 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM), 2017.

[20] Yang M , Margheri A , Hu R , et al. Differentially Private Data Sharing in a Cloud Federation with Blockchain[J]. IEEE Cloud Computing, 2018, 5(6):69-79.

[21] S. Wang, Y. Zhang and Y. Zhang, "A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralized Storage Systems," in *IEEE Access*, vol. 6, pp. 38437-38450, 2018. doi: 10.1109/ACCESS.2018.2851611

[22] Wang XL, Jiang XZ, Li Y. Model for data access control and sharing based on blockchain. Ruan Jian Xue Bao/Journal of Software, 2019,30(6):1661−1669 (in Chinese). http://www.jos.org.cn/1000-9825/5742.htm

[23] Yuan Yong, Wang Fei-Yue. Parallel blockchain: concept, methods and issues. Acta Automatica Sinica, 2017,43(10): 1703−1712

[24] Tsai WT, Blower R, Zhu Y, Yu L. A system view of financial blockchains. In: Proc. of the IEEE Symp. of Service-oriented System Engineering. IEEE, 2016. 450−457.

[25] Maymounkov P. A peer-to-peer information system based on the XOR metric. In: Proc. of the IPTPS. LNCS 2429, Springer-Verlag, 2002. 53−65. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.