Research on Decentralized Access Control in Big Data Environment

Nianzu Liu¹, Weichen Ma², Danjie Shen³*, Bing Guo¹

¹ College of Computer Science, Sichuan University, Chengdu, China, 610065
² School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China, 611756
³ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China, 610054
*Corresponding author e-mail: 2019223045144@stu.scu.edu.cn

Abstract. With the rapid increase of network data, there are many challenges of security about data sharing in big data environment. The traditional attribute-based access control model generally adopts the centralized method. In the process of practical operation, once the decision node is attacked, the method will bring security risks to data resources. In order to improve the security of centralized access control and adapt to both dynamic and fine-grained characteristics of data access control in big data environment. The decentralized access control method for big data is proposed by combining the blockchain technology with attribute-based access control model.

1. Introduction
The development of big data technology has promoted data circulation and value transfer, while bringing new development opportunities, it also faces severe challenges of data security. Data sharing in the big data environment is distributed and dynamic which makes the management scenarios and security requirements of big data more complicated. Therefore, traditional access control is difficult to meet the above needs. Meanwhile access control of big data faces many challenges as follow: the increased difficulty in formulation and authorization management of access control policy, the trouble about describing objects of access control, the complexity in subject set of the accessed data, the problem on protecting personal privacy in the data objects, the lack of security considerations for the big data analysis process and so on. Once there are loopholes in the process of access control, resulting in illegal access to data, it will cause economic losses to users participating in data sharing and will also affect user safety. Blockchain has the characteristics of distributed, transparent transactions, difficult to tamper with and a trusted mechanism that does not require third-party endorsement, which coincides with the demands of distributed deployment, audit mechanism and trust mechanism that need to be met by access control in the big data environment [1]. In the context of big data, there exists much improvement room and broad application scenarios for blockchain-based access control mechanisms.

2. Knowledge of blockchain technology and attribute-based access control

2.1. Introduction to Blockchain
Blockchain technology was first applied in the Bitcoin system and it has the characteristics of
decentralization, collective maintenance, security, credibility and traceability. There is no single official definition of blockchain, which is mainly interpreted as a distributed database and originated from the emergence of Bitcoin. Blockchain, which was proposed by Satoshi Nakamoto in 2008, is a series of data blocks related by cryptography technology. And each data block contains certain bitcoin transaction information to verify the validity of the data information and to calculate the next data area [2-3]. Blockchain technology is not a new type of technology in essence but it can be understood as a new technology formed by the clever combination of many current existing technologies such as hash functions, P2P networks, and consensus mechanisms.

2.2. Introduction to smart contract
The original definition of smart contract which was proposed by Nick in 1994 is a set of commitments defined in digital form, including agreements on which contract participants can execute these commitments. However, due to the limitations of early technology and usage of scenarios, smart contract has progressed slowly for a long time. It was not until the emergence of the blockchain which is the underlying technology of Bitcoin, that people discovered that the decentralized and trusted execution environment of blockchain perfectly fits smart contracts. Smart contracts also provide programmability for the blockchain and expand the application prospects of the blockchain.

2.3. Access control based on attribute
Attribute-based access control uses entity attributes to uniformly describe subjects, objects, permissions and authorization constraints, and uses attributes or attribute groups to distinguish different entities. The relationship among the entity attributes is used for formal modeling and their attributes are used as the basis for authorization decisions [4]. Therefore, in the attribute-based access control model, the operation authority of the subject to the object is not only based on the corresponding attributes of the subject and the object, but also restricted by the system environment. In addition, this model can easily convert with traditional access control mechanisms and provide fine-grained access control for complex information systems. The determination of the operation authority is mainly based on the evaluation of the attributes of the visiting subject, the object attribute, the operation attribute and the environment attribute according to the strategy, and then the access authority can be determined.

3. Problems of access control in big data environment
The aggregation of many distributed data sources coming from different institutions, companies and organizations that are not possible to share directly with others because of their different security protection of data. Big data resources have the characteristic of discretization, and it is necessary to implement effective access control to sharing of big data resources in a distributed and complex environment. And its distribution is reflected in the following aspects. (figure 1)

3.1. Distributed environment
It is not advisable to have only one centralized authority judgment node in the big data environment, and meanwhile the centralized authority judgment does not conform to the application scenarios of big data. The circulation and sharing of big data is an application process involving multiple parties and group intelligence perception, which requires the diversity of participants [5]. Only by sufficient interaction can the value of big data be fully mined and the utilization rate of big data resources be improved. Moreover, the traditional single-authority judgments made by third-party institutions may result in user-unknowable behaviors beyond authority and problems about transparency of judgment authorization. In addition, when malfunctions happen to a single-authority node, it will cause the entire system of big data to stop operating. So there is also a problem of single point failure.

3.2. Granularity of access control
Traditional role-based access control, which is considered a convenient method of management, using
a matrix method to assign permissions statically fails to meet the demands of dynamic and complex relationships of access control in the big data environment due to the complicated relationship among subjects participating in data sharing. Once the relationship of the subjects changes, the structures of permission assignment need to be allocated.

3.3. Dynamic access control
Big data is generated dynamically and increases rapidly. So it is necessary to perform timely access control management on dynamically generated new data [6]. Traditional centralized access control mechanisms is generally used to manage static resources, and have weak dynamic expansion capabilities, low flexibility and lagging data access control management. And it is also unable to perform efficient access control management on newly generated data resources.

3.4. Policy formulating
The centralized access control policy, which can not meet the security requirement of data sharing and may lead to inefficient management, is uniformly formulated and maintained by the administrator [6]. Therefore, the sharing and circulation of big data requires their owner to formulate and maintain the access control policy, and then realize the distributed development of the owner-driven access control policy. In addition, open sharing is obviously conducive to the mining of the value of big data, which requires that the policy of big data resources need to be trusted, publicized and more convenient for users to query and verify authenticity of policies. And this will efficiently promote circulation and sharing of big data resources which own the permission access.

4. The advantages of blockchain technology applied to access control
1) The right to management and use of resources is truly in the hands of the resource owner, and the policy information stored on the blockchain is visible to all subjects [7]. The credible disclosure of the policy is more conducive to promoting the sharing of resources of big data and is helpful to mining and leveraging the resource value of big data.
2) Based on smart contracts, automated and credible access control to resources of big data can be realized, without the participation by security administrators. Access control is based on the policies issued by the resource owner, and the process of judgment is open and transparent.
3) The blockchain is based on the distributed shared ledger technology, which can effectively ensure that the source and storage of the made policy is reliable and credible [8]. Through the consensus mechanism of distributed nodes, the anti-attack ability of the access control system is improved, and the occurrence of single point of failure is effectively prevented, as a result, the availability of the system is guaranteed.

Figure 1. Challenges facing access control
Blockchain is a data management mode that does not delete data but just adds data. Transaction data is always stored on the blockchain, and the data stored on the blockchain cannot be tampered with, therefore, it can realize the tracking control of resources of big data in the process of sharing and circulation and facilitate system auditing verification.

5. Decentralized access control model for big data

5.1. Model description
The decentralized access control model combines the attribute-based access control model with the blockchain, and uses smart contracts to realize the key modules such as policy enforcement points, policy decision points, policy management points, and policy information points in the attribute-based access control model, then cooperates with all resources stored in the off-chain database, including subject attributes, object attributes, environmental attributes, policy information and data resources, accordingly, the automatic management of attribute information and access policy information of data resources, as well as decentralized access control decision-making and resource security management are implemented. The smart contracts on the chain include: PEPC(Policy Enforcement Point Contract) is responsible for converting the access request sent by the user into an attribute-based access request, at the same time, PEPC executes the PDPC decision result, that is, denies or allows this access request. PDPC(Policy Decision Point Contract) makes authorization decisions based on access control rules. PAPC(Policy Administrate Point Contract) manages access control rules and provides rule support for policy decisions of PDPC. PIPC(Policy Information Point Contract) manages some attribute information (such as subject attribute, object attribute, authority attribute, environment attribute, etc.), and provides attribute support for PEPC. (figure 2)

5.2. Access control process
1) During the process of attribute uploading, the administrator is responsible for uploading the attribute information of the access subject to the attribute information database in advance.
2) During the process of data uploading, the data uploader uploads the data access policy to PAPC while uploading the data resource, PAPC manages the access policy, and generates the hash value of the data then stored them on the blockchain, and the original data is stored in the local database.
3) During the process of data request, the data requester will send the data request to the PEPC, and the PEPC will convert the access request into an attribute-based access request, and then submit
the request to the PDPC. $access(S,R,E,P)$ means whether the access subject $S$ can carry out $P$ operation on the object resource $R$ under the environment $E$.

$$access(S,R,E,P) \rightarrow access(A(S),A(O),A(P),A(E))$$ (1)

The PDPC initiates an access policy query of the object to the PAPC and an attribute information query to the PIPC, then after verifying the attribute set of the subject and the acquired access policy, the result is fed back to the PEPC, finally, the PEPC executes the result obtained by the PDPC, and denies or accepts the access request. $A$ means the function which is to obtain the corresponding object attributes, and the $f$ function represents to determine whether these object attributes comply with the access policy of the corresponding resource according to the subject attribute, object attribute, operation, environment attribute, etc., and give the allow access policy or deny access policy.

$$f(A(S),A(O),A(P),A(E)) \rightarrow \{\text{permit}, \text{deny}\}$$ (2)

5.3. Model analysis

The use of smart contract technology realizes the dynamic management of attributes and strategies, and totally transfers the centralized decision-making institutions of traditional access control to automatic execution on the blockchain, which ensures the transparency and security of the decision-making process, and implements the distributed access control and credible verdict. At the same time, the way of storing the data summary hash on the chain and storing the original data in the local database solves the storage bottleneck of the blockchain and realizes the efficient sharing of resources of big data.

Combining with the corresponding requirements of access control in the environment of big data, the access control model proposed in this paper is evaluated in terms of security, fine-grained, flexibility, efficiency of policy management, and decentralization, and compared with the traditional role-based access control model and the traditional attribute-based access control model, and the results are as follows. (table 1)

| Characteristic          | Type of model                                      |
|-------------------------|---------------------------------------------------|
|                         | Role-based access control model | Attribute-based access control model | Model of this paper |
| security                | ●                                                | ●                                  | ●                  |
| Fine-grained            | ○                                                | ●                                  | ●                  |
| Flexibility             | ○                                                | ●                                  | ●                  |
| Policy management       | ○                                                | ○                                  | ●                  |
| efficiency              |                                                  |                                     |                    |
| Decentralization        | ○                                                | ○                                  | ●                  |

6. Conclusion

The paper proposes a decentralized access control mechanism for resources of big data. The access control model is extended on the basis of the attribute-based access control model. By storing attribute information on the chain and realizing the key modules such as attribute decision-making agency, policy enforcement points, strategic decision points by using smart contracts, the decentralized decision-making method is implemented which further enhance the security of the access control process and increase the credibility of access decisions. Compared with the traditional access control model, this scheme has higher anti-attack and anti-collusion, and ensure that data will not be maliciously accessed. This model implements a safe, reliable, and transparent new access control architecture, which can effectively promote the safe circulation and sharing of big data.
References

[1] T.Jiao, D.R.Shen, et al 2019 BlockchainDB: querable and immutable database. Journal of Software 30(09) 2671-2685.

[2] Fay, Thomas, and D.Paniscotti Systems and methods of blockchain transaction recordation U.S. Patent Application No.15/086,801.

[3] Tama, B. A., Kweka, B. J., Park, Y., Rhee, K. H. 2017 A critical review of blockchain and its current applications. In: 2017 International Conference on Electrical Engineering and Computer Science (ICECOS) IEEE. pp. 109-113.

[4] L.Fang, L.H Yin,Y.C Guo. 2017 A survey of key technologies in attribute-based access control scheme.Chinese Journal of Computers. (07),1680-1698.

[5] Z.Y.Lin. 2018 Experience Sharing of Big Data Technology Principles and Application Course Construction. Big Data, 4 (06): 29-37.

[6] X.H.Cui, G.Chen, et al. 2016 Research on cloud data access control technology in big data environment. Modern Electronics Technique(15),67-69.

[7] A.Liu, X.H.Du, et al. 2019 A blockchain-based access control mechanism for big data. Journal of Software (09),2636-2654.

[8] X.L.Wang, X.Z.Jiang, et al. 2019 Model for data access control and sharing based on blockchain. Journal of Software 30(6): 1661-1669.