Research About Data Access Control Based on Data Sensitivity for Cloud-Fog Cooperative

Ruixia Li*, Zhenwei Chen and Wei Peng
School of Electronic and information Engineering, West Anhui University, Lu'an 237012, China
*Corresponding author

Abstract—**Cloud computing provides flexible services for users because of its sharing and openness, so traditional access control model cannot achieve flexible and dynamic access control in cloud computing environment. In this paper, a cloud-fog collaboration model is introduced. By adding a fog server, not only the distance between the terminal device and the cloud computing and storage is shortened, but also the burden of the cloud server is lightened. At the same time, the fog layer can determine whether the data need to be encrypted before being transmitted to the cloud by judging the sensitivity of the data, as well as the time and energy consumption required for encryption, so as to ensure the safety and efficiency of access data. We design the model in detail, introduce the basis of sensitivity judgment, and make a security comparison with the existing access control. The results show that the cloud-fog collaboration model can provide a good guarantee for users to access data.**

Keywords—cloud computing; fog computing; data sensitivity; access control

I. INTRODUCTION

In recent years, the Internet of Things has connected various kinds of intelligent devices to the network (such as wearable devices, vehicles, wireless sensors, etc.). Cisco predicts that the number of intelligent devices connected to the Internet will exceed 50 trillion by 2020, with an average of 6.58 smart devices per person [1-2]. With the increase of users and devices, massive data will be generated. Cloud computing is regarded as a good solution for massive data storage and processing. It provides users with scalable software, platforms and infrastructure at a low price. However, with the development of big data technology, cloud computing is also facing more and more challenges. On the one hand, it is necessary to solve the contradiction between the increasing application demand of massive data and its inherent shortcomings (such as lack of mobility support, unreliable delay, many Internet of Things application delay is generally less than tens of milliseconds), on the other hand, a large number of increasing production of connecting devices. The amount of data generated increases exponentially, and data transmission requires extremely high communication bandwidth. However, these requirements are far beyond the service level of cloud computing. Therefore, there is an urgent need to find a new solution. Fog computing can shorten the distance between terminal devices and cloud computing and storage by introducing an intermediate fog layer [1-2] between cloud and edge devices. It can directly provide new services and applications. The prominent feature of fog computing is to support mobile, closer to users and widespread geographical distribution.

Access control is an important technology in big data technology. Because of the inherent open and sharing characteristics of cloud platforms, data storage in the cloud will lead to data loss of control by data owners, and their security and privacy will face potential threats from many aspects in complex environments. The data stored in cloud platform and users are dynamic, and even the user's rights are dynamically adjusted. Therefore, access control can ensure the availability and reliability of data.

In recent years, scholars have carried out a lot of research on access control model of large data. The typical model is cloud computing from the aspect of system architecture. Secondly, with the wide application of the Internet of Things and the rapid development of 5G [4], the application of various new service modes and terminal devices has produced a large amount of data. Such modes are diversified and data sources are heterogeneous. The traditional data access control security mechanism in cloud computing mode cannot meet the needs of practical applications. Therefore, fog computing has become a necessary complement to cloud computing [5]. The cloud-fog collaboration architecture model satisfies the needs of such applications very well.

In summary, in the complex environment of many heterogeneous networks, the access control of big data only using cloud computing architecture mode cannot guarantee security and efficiency, and cannot achieve good user experience. Based on cloud-fog collaboration architecture mode, this paper proposes an access control model based on data sensitivity, energy consumption and time. The model first determines whether the semi-trusted fog node needs to be encrypted in advance according to the sensitivity of the data, and then transmits it to the remote cloud for storage. At the same time, the energy consumption and time attributes of fog node encryption are taken into account, thus forming dynamic fine-grained access control to ensure data security and user privacy.

II. PRELIMINARIES

A. *Cloud-fog Collaboration Architecture Model*

Cisco first introduced the concept of fog computing in 2012. Fog computing is an extension of cloud computing. It adds fog layer between terminal devices and clouds to deploy computing and storage services closer to users. HP defines it as fog computing, which is used in heterogeneous wireless networks. A large number of distributed devices cooperate with each other without interference to complete computing and storage tasks. Providing this service to users for a fee [6]. Table 1 is a comparison between cloud computing and fog computing. Through comparison, it can be seen that fog computing adds a network and computing intermediate service layer between users and cloud servers, which can satisfy rich application scenarios than independent cloud computing.

TABLE I. COMPARISON BETWEEN CLOUD COMPUTING AND FOG COMPUTING

Parameter	cloud computing	fog computing
Deployment	Network core	Network edge
Ownership	Commercial entity	Commercial entity and individuals
Hardware	Adequate and scalable	Currency
Framework	Centralized	Distributed
Accessment	Fixed and wireless	Wireless based
Target user	General Internet users	Mobile user
Service	Virtualization	Virtualization
Time delay	High	Low
Mobility	Not applicable	Applicable
Location awareness	Not applicable	Support

B. *Cloud-fog Collaboration Architecture Model*

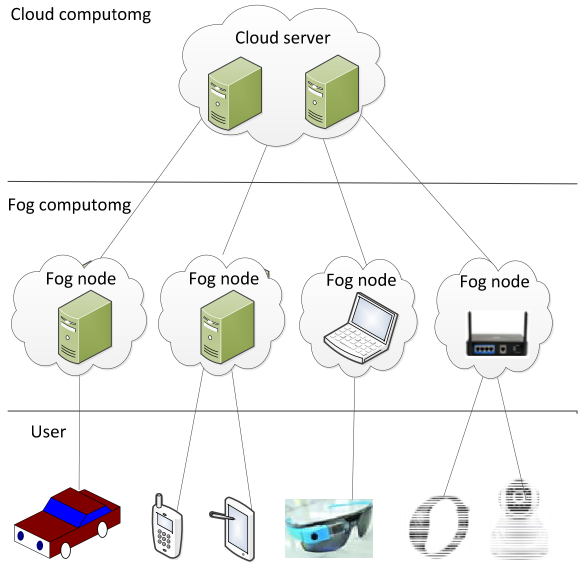

FIGURE I. CLOUD-FOG COLLABORATION MODEL

Figure 1 is a cloud-fog collaboration model, which includes users and terminal devices, intermediate fog layer and remote cloud computing. Users' equipments are connected to the middle layer fog equipment, fog equipment can communicate with each other, and fog equipment is connected to the cloud. As we all know, there are three main difficulties in the application of intelligent devices: battery capacity, storage space and computing resources. In addition, bandwidth resource is also an aspect that affects users' experience of quality of service. Therefore, the fog layer can pre-process data from users, so as to alleviate overloaded cloud device data center, reduce service delay and improve user experience [7].

Fog computing will collect and process data first. Obviously, if there is no proper security and privacy protection mechanism, fog computing cannot be used. Therefore, fog computing will suffer from the same classical security and privacy problems as cloud computing. Therefore, it is particularly important to ensure security in access control. By allowing or restricting user access to the system, the normal access of effective users is guaranteed, the attacks of unauthorized users are prevented, and the security problems caused by failure operation of effective users are solved.

In traditional access control, users store data in trusted servers. The trusted server then checks whether the requested user can access to the data. In cloud-fog collaboration mode, access control can be divided into two levels. Firstly, the fog node can directly transmit the data with low sensitivity to the cloud in the form of ordinary text according to whether the data will pose a serious security threat because of high sensitivity. Otherwise, the data with high sensitivity can be encrypted in the fog layer first, and then encrypted in the form of cipher text. Transfer to the cloud. Next, we will introduce the evaluation basis and methods of data sensitivity.

III. DATA SENSITIVITY

A. *The Judgment Basis of Data Sensitivity*

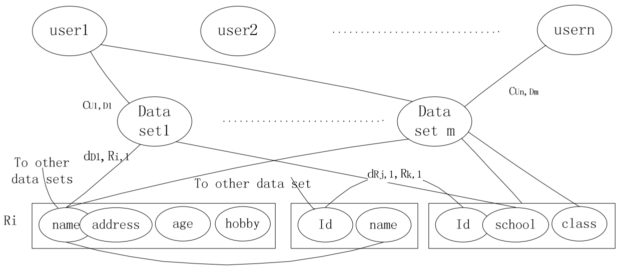

FIGURE II. DATA SENSITIVITY MODEL

The data contains sensitive information concerning personal privacy. The article [8] considers that the following three principles are related to data sensitivity.

- Data access: The more data is accessed, the more sensitive it is. Because with the widespread use of data, it also increases the probability of abuse.

- Connectivity (or similar data): The more a data attribute is associated with other attributes, or with other databases, the more sensitive it is.
- Data quality: The higher the data quality of a data set, the more sensitive it is to missing or damaged or erroneous data.

Based on the above three factors, we use Shannon Entropy [9] to estimate the sensitivity of the data. Assuming that the data set stored in the fog node is modeled as a graph, the node of the graph is the user, the data set and the data item in the data set: as shown in Figure 2.

U_i denotes the user node i, D_i denotes the user's request i, R_i denotes the data set i, $R_{i,\ j}$ denotes the data set R_i's item j. $C_{U_i,\ D_j}$ denotes the number of times user U_i requests D_j, the i request for a particular data item k in data set R_j is $d_{D_i},\ R_{j,\ k}$, and the j data item in a data set i to an edge of the n data item in data set R_m is $d_{R_i,\ j,\ R_m,\ n}$. In this paper, we calculate the sensitivity of a given node by deleting its influence on the system.

B. *Data Sensitivity Definition*

Definition 1: According to the first principle, the probabilistic mass function of data C_{U_i,D_j} in the data sensitive graph is calculated as follows:

$$p(U_i,\ D_j) = \frac{\sum_{i=1}^{m} C_{U_i,\ D_j}}{\sum_{j=1}^{k}\sum_{i=1}^{m} C_{U_i,\ D_j}} \tag{1}$$

In the above formula, $\sum_{i=1}^{m} C_{U_i,\ D_j}$ represents the number of requests to D_j made by all users, and $\sum_{j=1}^{k}\sum_{i=1}^{m} C_{U_i,\ D_j}$ represents the number of other requests made by all users. Request D_j of all nodes in the sensitive graph is used to calculate the entropy.

Definition 2: According to Principle 2, the probability mass function of the connection is calculated as follows:

$$p(R_i) = \frac{\sum_{j,m,n=1,1,1}^{j,m,n=a,b,c} d_{R_i,\ j},\ R_{m,\ n}}{\sum_{i,j,m,n=1,1,1,1}^{i,j,m,n=d,a,b,c} d_{R_i,\ j},\ R_{m,\ n}} \tag{2}$$

A data set can be used as a join point for another data set. The symbol attribute shown in Figure 2 connects multiple data sets. Such nodes are very sensitive data. In the above formula, $\sum_{j,m,n=1,1,1}^{j,m,n=a,b,c} d_{R_i,\ j},\ R_{m,\ n}$ represents the number of arcs or paths connecting node R_i, and $\sum_{i,j,m,n=1,1,1,1}^{i,j,m,n=d,a,b,c} d_{R_i,\ j},\ R_{m,\ n}$ represents the total number of arcs and paths connecting all nodes.

Definition 3: According to Principle 3, the probabilistic quality function can be calculated by the following formula. High quality data means high sensitivity. In our model, we use lost, erroneous and damaged data to represent data quality. $co(R_i)$ represents the correct entry of all data items in R_i and S_i represents all entries of all data items in R_i. $\sum_{Sr} co\ (R_j)/S_j$ is the proportion of correct data in a data set, and $\sum_{i=1}^{n}\sum_{Si} co\ (R_i)/S_i$ represents the correct number of entries in all data sets.

$$co(R_i) = \frac{\sum_{Sr} co\ (R_j)/S_j}{\sum_{i=1}^{n}\sum_{Si} co\ (R_i)/S_i} \tag{3}$$

Definition 4: Formula 1-3 can combine to form entropy. As shown in Formula 4, $H(C_u, d_i)$ refers to the entropy calculated by data access, $H(d_r)$ refers to the entropy calculated by similarity, and $H(N_r)$ refers to the entropy calculated by mass. The calculation of these three entropy is shown in Formula 5-7: Formula 5-7 is derived by substituting Formula 1-3 into Formula 1 respectively.

$$H(x_i) = H(C_u, d_i)gH(d_{ri})gH(N_{ri}) \tag{4}$$

$$H(C_u, d_i) = -\sum_{j=1}^{n} p(C_{Uj},\ D_i)1ogp(C_{Uj},\ D_i) \tag{5}$$

$$H(d_{ri}) = -p(R_i)\,1ogp(R_i) \tag{6}$$

$$H(N_{ri}) = -co(R_i)\,1ogp(R_i) \tag{7}$$

Definition 5: Sensitive values of data sets are determined by removing sensitive graphs, which are calculated by the difference between the sum of all data sets and the entropy of data sets.

$$C(x_i) = \sum_{i=1}^{n} H(x_i) - H(x_i) \tag{8}$$

In formula 8, $C(x_i)$ represents the sensitive value of data set i , $H(x_i)$ represents the entropy of data set i , and $\sum_{i=1}^{n} H(x_i)$ represents the sum of the entropy of n data sets.

IV. ACCESS CONTROL BASED ON DATA SENSITIVITY IN CLOUD-FOG COLLABORATION MODEL

In order to prevent the data involving user sensitive information being used by illegal users, on the one hand, the data is encrypted, so that even if the unauthorized user acquires the data, it cannot be used because it cannot be decrypted; on the other hand, the role and identity attributes of the visitors are limited to ensure the rational use of the data. Existing research on access control mainly restricts the users, categories and operations of access. From the analysis of access users and access strategies, there are mainly role-based access control [10]. The main research contents include the analysis and design of roles and the allocation of privileges. Attribute-based Encryption Access Control[11]. First, we design a reasonable access strategy to meet the access needs of many users with different privileges. Secondly, we focus on privacy protection and protect user data with access control strategy. Finally, we consider the system efficiency to minimize computing and storage overhead while ensuring security and reliability. Identity-Authentication Access Control [12] protects the security and privacy of user data from fingerprints and faces. Act-Based Access Control [13-14] defines a variety of security levels from user roles, time domain attributes and environmental factors to ensure data confidentiality and integrity. In addition, one of the key technologies to protect data security in information security is encryption technology. CP-ABE is an attribute-based encryption strategy, which is the most successful encryption method in fine-grained access control [15]. In addition, KP-ABE is mainly based on Attribute-Based key encryption strategy [16]. Another important feature of practical application is to support some computing outsourcing services and effective search of data ciphertext.

When sensitive data is stored in fog nodes, it also faces security threats. In the cloud-fog collaboration model proposed in this paper, we mainly consider three main issues, privacy protection, energy consumption and efficiency, in order to achieve safe and efficient access control. Firstly, efficiency can be improved by limiting time, that is, to ensure that all tasks can be completed within a specific period of time. Secondly, the quality of encryption is closely related to privacy protection. We use fog server to determine the access mode of requesting data. Before transferring the collected data to the cloud server through the application, the fog server determines its access mode by judging the sensitivity of the information. The data that can be accessed by the service provider with low sensitivity will be stored in the cloud in plain text. For other data with sensitive information, in order to avoid direct access by service providers, it is first encrypted, and then stored in the cloud in the form of ciphertext.

A. *Implementation of Access Control*

Definition 6: Input is a series of data packets D^j , each data packet represents a data set, the data in the data set is sorted according to the privacy weight. Percentage of encrypted data per packet Pr_i^j and its corresponding execution time T_i^j , E_i^j represent the required energy consumption. Output is a percentage of encrypted data per packet.

As mentioned above, if there is a data packet D^j , D_i^j represents some data in the data packet, and Pr_i^j refers to a data packet containing i encrypted data. In the model proposed in this paper, first of all, according to the needs of practical application, formula 8 is used to calculate the data sensitivity, to determine whether the data need to be encrypted in the fog server, and to realize the access of the control part of the data. At the edge level, applications are primarily data owners who determine power consumption and estimate privacy classifications. In the fog computing layer, the data packets are sorted according to their privacy weight, so the output strategy encrypts the data sequentially, and then the energy consumption needs to be evaluated in terms of data type and size. Formula 9 represents the energy consumption cost, $P1$ is the data type, S represents the data size.

$$E^{est} = \alpha^{P1} S \tag{9}$$

The objective function is proposed to abstract the problem in mathematical form:

$$\begin{cases} Pr^0 = MAX[Pr] \\ E^0 = MIN[E] \; while \quad having \quad Pr^0 \\ T \leqslant T_c \end{cases} \tag{10}$$

The formula includes three main variables, Pr^0 denotes the ideal percentage of encrypted data, E^0 is the energy consumption related to the ideal percentage of data, and T denotes the time related to the above two variables. Formula 11 represents the calculation method of the optimal data percentage. m represents the number of data packets. j is the index of data packets. $n(j)$ represents the number limit in a data packet.

$$Pr = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} Pr_j^i \tag{11}$$

$$E = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} E_i^j \tag{12}$$

$$T = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} T_i^j \tag{13}$$

In Formula 10, the sensitivity of user privacy, energy consumption and encryption time are fully considered. The weights of the above three items can be differentiated according to different applications. In this paper, by adding fog layer, the work of encrypting sensitive data in fog computing layer combines the required time and energy consumption to judge. According to the actual application needs, it decides whether to encrypt and then transmit to the cloud, so as to increase the security and reliability of access control.

B. *Usability analysis*

In order to make a clear and intuitive comparison between the proposed model and the existing access control model, we chooses the role-based access control model and the attribute-based access control model to compare from eight aspects. As shown in Table 2, all the indexes of the model in this paper are optimal, which can guarantee the data security and reliability, and take account of both computing time and complexity to achieve flexible data processing. It can meet the requirements of access control with many users, large amount of data, complex running environment and high security requirements.

TABLE I. COMPARISON BETWEEN OF DEFERENT ACCESS CONTROL MODEL

Model	RBAC	ABAC	Our Modle
Time	N	N	Y
Location	N	N	Y
Cloud computing	N	Y	Y
Fog computing	N	N	Y
Flexibility of authorization	N	Y	Y
Extensibility	Bad	Better	Good
Security of Model	Low	Normal	High

V. CONCLUSION

Cloud computing cannot meet the needs of existing access control. In view of this shortcoming, this paper proposes a basic platform combining cloud computing and fog computing, which can improve the security and access efficiency of data from mobility. Secondly, the evaluation index and calculation method of data sensitivity are given to determine the comprehensive data sensitivity, encryption complexity and running time complexity in fog layer. Degree three aspects to determine whether the need for data encryption in the fog layer. Finally, by comparing with role-based access control and attribute-based access control, the results show that the method proposed in this paper can effectively achieve dynamic data security access control. In the future, the model will be further improved by choosing different levels of data in combination with effective encryption algorithm, so as to make a more comprehensive analysis and verification of the model proposed in this paper.

ACKNOWLEDGMENT

This work was supported in part by the scientific research project of the Anhui Provincial Education Department (No. KJ2018A0414) and West Anhui University Project (WXZR201711).

REFERENCES

[1] Muhammad Saad. Fog Computing and Its Role in the Internet of Things: Concept, Security and Privacy Issues [J]. International Journal of Computer Applications, 2018, 180:7–9.

[2] BONOMI F. Connected vehicles, the internet of things, and fog computing[C].The Eighth ACM International Workshop on Vehicular Inter-Networking (VANET). 2011: 13-15.

[3] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Communications of the ACM,2010,53(4) : 50-58.

[4] Li Dongyue, Yang Gang,Qian Bo. Research on Architecture of Internet of Thing [J]. Computer science, 2018, 45(11):27-31.

[5] VAQUERO L M, RODERO-MERINO L. Finding your way in the fog[J]. ACM SIGCOMM Computer Communication Review, 2014,44(5):27-32.

[6] Matei Zaharia, Dhruba Borthakur, J. Sen Sarma, Khaled Elmeleegy,Scott Shenker, Ion Stoica, Job Scheduling for Multi-user Mapreduce Clusters,Tech. Rep. UCB/EECS-2009-55.html, EECS Department, University of California,Berkeley, 2009.

[7] Claude Elwood Shannon, A Mathematical Theory of Communication [J], ACM SIGMOBILE Mobile Computing and Communications, 2008,27,3–55

[8] Ashwin Kumar T.K., Hong Liu, Johnson P. Thomas, Xiaofeh Hou. Content sensitivity based access control framework for Hadoop[J], Digital Communications and Networks,2017,3:213–225.

[9] Keke Gai, Meikang Qiu, Meiqin Liu. Privacy-Preserving Access Control Using Dynamic Programming in Fog Computing[C], 2018 4th IEEE International Conference on Big Data Security on Cloud, 2018,126-132.

[10] SandhuR,CoyneEJ,Feinstein H L,etal. Role G based Access Control Models [J]. IEEE Computer, 1996,29(2):38-47.

[11] EricY,JinT. Attributed Based Access Control (ABAC) for Webservice [C]//Proceedings of the IEEE International Conference on WebServices. Orlando, FL,USA:IEEE, 2005 : 561-569.

[12] Wang Yuding,Yang Jiahai. Data Access Control Model Based on Data's Role and Attributes for Cloud Computing [J]. Tsinghua University (Science and Technology), 2017, 57(11):1150-1158.

[13] Tian Hongliang, Zhang Yong, Li Chao, Xing Chunxiao. A Survey of Confidentiality Protection for Cloud Databases [J].Chinese Journal of Computer, 2017,40(10):2245-2270.

[14] Jin Yu, Wang Fan, Zhao Hongwu, Deng Li. Servey on Trust Mechanisms in the Environment of Cloud Computing [J]. Journal of Chinese Computer Systems, 2016, 37(1):1-11.

[15] Tu Yuanfei, Xia Feng, Yang Geng． Privacy-preserving Ciphertext-Policy Attribute-Based Encryption in Hybrid Cloud [J]． Microelectronics and Computer， 2016,33(10):53-58.

[16] Goyal V, Pandey O, Sahai A, Waters B. Attribute-Based encryption for fine-grained access control of encrypted data. In: Proc. Of the 13th ACM Conf. on Computer and Communications Security. New York: ACM Press, 2006. 89−98.