Analysis and Research on the Big Data Security Based on Cloud Platform

Bo Yang[1, a]

[1] Beijing Earthquake Administration of Beijing Municipality, Beijing, 100080, China

[a]231373693@qq.com

Keywords: data security; encryption model; cloud big data; code information

Abstract. Objective: The purpose of this paper is to build a large data security encryption model in the cloud, which can make users to upload and download data more reliable. Model: This paper uses the data security storage model to analyze the security model of the cloud big data. Process: We used the mathematical code, and storage the cloud big data into the memory, then, the code information is composed of element equation, each memory is unable to obtain the code, so to achieve the purpose of encryption and protection of data. Conclusion: Cloud technology and big data security is still in its infancy, which is the trend of future development, this paper provides a theoretical reference for the security encryption mode and it has a great significance.

Introduction

In the era of big data[1,2,3], information and knowledge are also important resources of enterprises. Now many developed countries in the effort to become a knowledge-based enterprise, and constantly using advanced information technology to tap the knowledge from large data in the accounting, and enhance the core competitiveness of enterprises. China in the eighteen major is a clear to the level of information technology to improve the level of a comprehensive well-off society into one of the goals, so as to ensure sustained economic development, comprehensive national strength has been enhanced. As Chinese enterprises, how to adapt to the complex economic environment, realize the value of enterprise's value, is also the focus of the current enterprise information construction [1]. The core of enterprise information is accounting information, due to the high cost, low efficiency, long construction cycle, technology, and other factors, the existing accounting information system is difficult to obtain large amounts of accounting data from the external enterprise, and found the knowledge, to provide scientific basis for the management of enterprise managers. Therefore, how to obtain and excavate the valuable knowledge hidden behind the big data in accounting, and promote the sustainable development of enterprises, is the common problem of the academic and business circles.

With the development of networking technology, cloud computing, it has low cost, large storage space, processing speed and other advantages, around the world with information related research began to rely on it, our country also launched a cloud computing information planning in 45 countries, focusing on three areas of cloud computing platform of big data services based on. In the construction of accounting information system [2], this paper also proposes a large data analysis platform based on cloud computing, which uses cloud computing technology to obtain, cluster and analyze the accounting data. It not only overcomes the problems of traditional accounting information system, but also greatly improves the efficiency of large data analysis.

Research on security related technologies for large data storage

In the traditional data relationship, Owner (Data) is a data provider; users only need to submit a user name and password can be related to the operation. But in cloud computing, data owners and cloud service providers (Service Provider Cloud) these two roles are often separated [3]. The role of cloud service providers are mostly borne by commercial organizations, these agencies in the user's trust region, so the traditional authentication methods cannot meet the needs of cloud storage security

access, cloud storage access need to take additional verification mechanism, document [4] overview of the access process of cloud storage, and the right to authentication, encryption, decoding and other operations to do a simple introduction, in this paper, a reliable security access model is proposed on the basis of document [5-7], as shown in figure 1. Users send requests to the data owner to get the real-time issue of the key, certificates, access to the cloud. This model can provide a safe and reliable access, but the defect is that when the user needs to operate the data, the data owner must be in an online state. Once communication is limited, the scheme is unable to guarantee secure access.

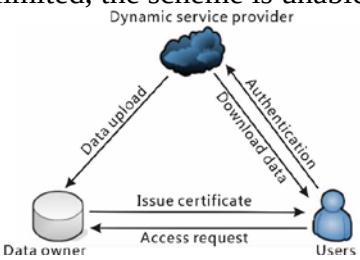

Fig. 1 A reliable security access model

In view of the above problems, the paper presents an optimized solution for document [8]. The scheme uses the access control method based on the user's ability and the related encryption strategy. The model is shown in Figure 2. Each data owner has a user's "ability to table", which stores a user's operating privileges for certain files, and then uploaded to the cloud server, along with the encrypted file, when users need access, cloud server according to user identity to operate. If the user is in the range of the ability table, it is not trusted users, directly refused access; if the ability of the table, then feedback user information, including the key to decrypt the file. Because the data owner has the message encryption, it will not leak to the cloud. In this scheme, the data owner can be in an offline state for most of the time, and can only be updated with the new user registration.

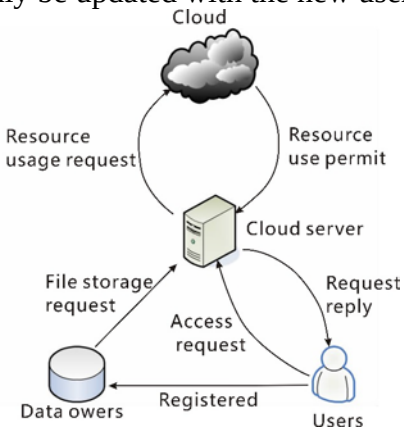

Fig. 2 Model based on user access capability

Similar to the literature [9], the [10] also has data from the "eternal online" in the liberation of imprisonment. In this paper, an access scheme is proposed, which is combined with attribute encryption and proxy re encryption and a large number of updates are made, and the work of the user's private key is assigned to the cloud. [11] is proposed based on Platform Module Trusted (trusted platform module) access control scheme. The scheme is based on the characteristics of mobile devices, and the security domain of cloud computing is defined from the perspective of the client. The paper proposes an access control strategy based on Merle hash tree, which optimizes the performance of the TPM, and uses the statistical method of dynamic trust degree. However, it can only be used for large lot and private file access control with no specific encryption or certificate issuing mechanism. Literature [12] proposed a control access scheme based on spanning tree, a spanning tree is composed of three minor, respectively allow the identity of the user access "," ban the identity of the user access " [13] and" the identity of the user access selection ", increase the spanning tree of the elasticity and flexibility, extent meet the data have who is convenient and flexible to manage user access to the demand. Document "14" [14] proposed an access scheme based on multiple authentication centers. This scheme assumes a trusted authentication center and multiple independent attributes authentication points, which are verified by the method of user identity and

multiple attribute authentications. The scheme solves the problem of data sharing in multi user identity attributes in cloud storage.

Research shows that the security of cloud storage access depends on the data owner for the user access needs and related feedback. Data owner to maintain the online state can control the security of cloud computing access [15], but a large number of distribution, update the key work will give the host to increase the burden, once the host communication is blocked cannot meet the needs of users to share data. Access control based on the third party cloud server can share the workload of the host and the use of heavy encryption technology in the cloud to avoid leakage of the third parties, but the flexibility and real-time is not high enough, a large number of new users cannot cope with the needs of access [16]. Data owners should be based on the privacy of the data and the user's management mode selection of cloud access control method, in which the security of access to optimize the network efficiency.

Security analysis of large data storage based on cloud platform

Cloud computing market, Ali cloud, Google cloud, Microsoft cloud computing service providers more cloud storage and computing services, which help data owners to solve these massive data storage and management issues, can very well save data owners of the storage space and management costs, however, how to deal with the user data under the premise of protecting data privacy, to help data owners really get the intrinsic value of data, such as correlation, causality, etc., is a problem that has not been fully resolved. The popularity of cloud computing services has been solved. The problem of data computing has been challenged.

In this scheme, the principle of polynomial solution based on finite field is applied to the algebraic cryptography, the data $d \in Z_p$ is randomly divided into K parts $(r_1, r_2, \cdots r_k) \in Z_p$, and the k value is considered as the root of the k polynomial. Each part is stored in a different cloud server, the process of data information is not encrypted, and we call this data processing method for the implicit mechanism. In the scheme, the various cloud servers responsible for storing the data will not reveal any information. Only when all the cloud servers are in collusion can they cause the leakage of the data information and the safe storage.

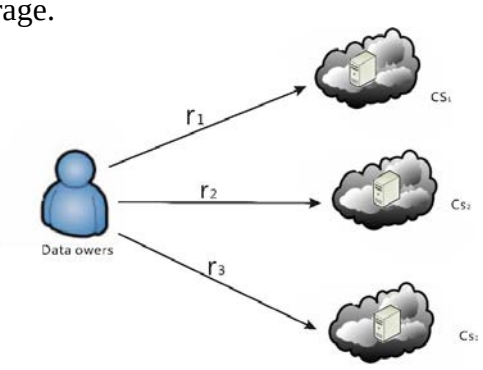

Fig.3 (k,k) data security storage model

According to the fundamental theorem of algebra, each k times equation k, for the k order equation:

$$x^k + a_{k-1}x^{k-1} + a_1 x + a_0 = 0 \tag{1}$$

On the k root, it represents $(r_1, r_2, \cdots r_k) \in \{\text{Complex set}\}$, according to the basic definition of algebra, another way of writing it:

$$(x - r_1)(x - r_2)\cdots(x - r_k) = 0 \tag{2}$$

In cryptography, we use the finite field Z_p, where p is a prime number, according to the nature of the one dollar polynomial we use data $d \in Z_p$ instead (1) in the a_0, then we get the next type:

$$x^k + \sum_{i=1}^{k-1} a_{k-1}x^{k-1} \equiv 0 \bmod p \tag{3}$$

Among them, $o \le a_i \le p-1$, $o \le d \le p-1$ can be transformed into:

$$\prod_{i=1}^{k}(x - r_i) \equiv 0 \bmod p \tag{4}$$

r_i is the data sub block; it is obvious that it is independent of the variable x, so it is very obvious:

$$\prod_{i=1}^{k} r_i \equiv d \bmod p \tag{5}$$

Overall, the (k,k) scheme requires all of the k partitions to reshape the data, and the second scheme is extended to the first scheme, and adding redundancy to form (k,n) partition scheme, of which, $k \le n$ and $k \ge 2$, in this scheme, we only need to n in k partition to reshape the data. Similarly, when we analyze the security of the security, we can see that the stolen data information is stolen, in the (k,k) scheme, the d can recover the data information from the k. If you steal one of the k -1 information, then there is the possibility of d to obtain the entire data information 1/p. In the scheme, the d the whole data information can be learned by the k, which is known to be redundant information. Below our scheme will be mainly from the perspective of security privacy design, to steal the entire data information C' and A in each element of the d, you must know that every element in C' and A is in the information, or you can't get the full data d. This is meaningful in protecting the privacy of data information.

Conclusion

Cloud computing and big data is currently the world's most anticipated technological revolution, and data security is not only related to the development of cloud computing technology, but also to the privacy and interests of each user. This paper is based on 3 aspects: access control, data encryption and data integrity detection. For the research point, this paper summarizes the classical techniques and related schemes for cloud computing security. Future security for cloud computing and large data storage is still a lot of problems need to be resolved: In the access control, part of the verification scheme based on the third party is credible, but in the actual file storage operation, such authentication center is not completely credible, and its security and credibility is still required to be further improved; (2) in terms of encryption security, due to large data size, will undoubtedly increase the complexity of the algorithm, and the network will increase the complexity of the algorithm, which is easy to cause congestion. (4) Cloud storage security is not only a technical problem, but also includes the system of standardization, monitoring mode and other issues. How to establish a complete set of cloud storage mechanism let the different terminals and safer and more convenient to share data, each link can be accountability, which is a topic worthy of the probe.

References

[1] Hashem I A T, Yaqoob I, Anuar N B, et al. The rise of “big data” on cloud computing: review and open research issues. Information Systems, 2015, 47: 98-115.

[2] Zuech R, Khoshgoftaar T M, Wald R. Intrusion detection and Big Heterogeneous Data: a Survey. Journal of Big Data, 2015, 2(1): 1-41.

[3] Yang S, Zhang X, Diao L, et al. CAPER 3.0: A Scalable Cloud-Based System for Data-Intensive Analysis of Chromosome-Centric Human Proteome Project Data Sets. Journal of proteome research, 2015.

[4] Ivan C, Popa R. Cloud based Cross Platform Mobile Applications Building and integrating cloud services with mobile client applications. Advances in Computer Science: an International Journal, 2014, 3(2): 69-77.

[5] Assunção M D, Calheiros R N, Bianchi S, et al. Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing, 2015, 79: 3-15.

[6] Öksüz A, Walter N, Compeau D, et al. Sync&Share North Rhine-Westphalia: a case on a university-based cloud computing service provider. Journal of Information Technology Teaching Cases, 2015.

[7] Chen C L, Yang T T, Shih T F. A secure medical data exchange protocol based on cloud environment. Journal of medical systems, 2014, 38(9): 1-12.

[8] Dong X, Li R, He H, et al. Secure sensitive data sharing on a big data platform. Tsinghua Science and Technology, 2015, 20(1): 72-80.

[9] Gonidis F, Paraskakis I, Simons A J H. Rapid Development of Service-based Cloud Applications: The Case of the Cloud Application Platforms. International Journal of Systems and Service-Oriented Engineering (IJSSOE), 2015, 5(4): 1-25.

[10] Bhargava B, Khalil I, Sandhu R. Securing Big Data Applications in the Cloud. IEEE Cloud Computing, 2014 (3): 24-26.

[11] Abolfazli S, Sanaei Z, Tabassi A, et al. Cloud Adoption in Malaysia: Trends, Opportunities, and Challenges. Cloud Computing, IEEE, 2015, 2(1): 60-68.

[12] Liu B, Madduri R K, Sotomayor B, et al. Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses. Journal of biomedical informatics, 2014, 49: 119-133.

[13] Sinnott R O, Voorsluys W. A scalable Cloud-based system for data-intensive spatial analysis. International Journal on Software Tools for Technology Transfer, 2015: 1-19.

[14] Sahoo S S, Jayapandian C, Garg G, et al. Heart beats in the cloud: distributed analysis of electrophysiological ‘Big Data’using cloud computing for epilepsy clinical research. Journal of the American Medical Informatics Association, 2014, 21(2): 263-271.

[15] Yu Y, Mu Y, Ateniese G. Recent advances in security and privacy in big data. 2015.

[16] Wang L, Ranjan R, Kołodziej J, et al. Software Tools and Techniques for Big Data Computing in Healthcare Clouds. Future Generation Computer Systems, 2015, 43: 38-39.