Advances in Social Science, Education and Humanities Research (ASSEHR), volume 181
4th International Conference on Social Science and Higher Education (ICSSHE 2018)

Research of Library Information Security Based on Cloud Computing Platform

Yang Fangjing, Liu Yinhong*, Chen Xinyan, Zhang Ronghui
Library of Wuhan University of Technology
Wuhan 430063

***Abstract*—Information security based on cloud computing platform is becoming one of the main concerns and challenges for library in big data era. This paper analyzes the current situation and security problems under the cloud computing platform, and then proposes to use encryption algorithm to ensure data security, furthermore proposes YC-R to improve the previous algorithm. This improved algorithm can better protect data privacy, and compared with previous algorithm, YC-R can shorten the time used in data processing. Meanwhile, it can improve data security capability of library.**

***Keywords*—Library; Big Data; Cloud Computing; Privacy Security; Information Management**

I. Introduction

The arrival of the era of cloud computing is not only a revolution in the field of information technology, but also an effective way to lead social change in the information age. Its core value lies in the mining of big data, making a large amount of data that is difficult to collect and use easy to use through data analysis. With this advantage, cloud computing can be widely used in education, medical, health, insurance, energy agriculture and transportation[1]. At the same time, the era of big data has brought about tremendous changes in the library's operating model, service concept user needs and market environment. How to effectively collect private data such as readers' reading behaviors, identity characteristics, personal hobbies and social relationships, and then integrate, analyze and mine the collected big data to realize the tracking service and demand forecast for readers have become the key to the library to change the service model according to the needs of readers, improve service effectiveness, user satisfaction and market competitiveness.

However, social networks in the cloud computing environment have serious privacy issues while providing diversified and personalized services. For example, in June 2013, former CIA staff Snowden disclosed the “Prism Gate” program. In accordance with what he said, the US National Security Agency and the Federal Bureau of Investigation monitor the social network privacy data of users in various countries through servers and databases of IT companies such as Microsoft, Google, and Yahoo[2]. In addition, there is the "Cambridge Analysis Scandal" that broke out in March 2018. According to reports, Cambridge Analytics obtained 87 million unauthorized Facebook user data by purchasing from third parties, and used this to deliver targeted advertising to help Trump win the 2016 US presidential election[3]. Not only that, but the Brexit referendum was also exposed to Cambridge analysis. Information security and personal privacy protection have become important issues in the era of big data and related to social interests. Therefore, while using big data to improve service quality and user satisfaction, how to take effective measures to protect users' privacy has become an urgent problem. This paper will explore the forms and causes of privacy security issues, and provide a new way to effectively manage library user privacy under the cloud computing platform.

II. The Forms and Causes of Library Security Problems under Cloud Computing Platform

In the context of big data, the enormous economic benefits of personal privacy can lead some organizations or stakeholders to collect, store, and process personal data in a variety of ways. According to the 2018 China Netizens Defrauded and Rights Protection Investigation Report, the most serious fraudulent area that users want to expose is the illegal collection of personal information industry and nearly 70% of the users hope that the 315 evening party will further expose the black-hearted enterprises that illegally collect personal information, which on the other hand confirms the dramatic increase of public's awareness of personal information protection, as shown in Fig. 1.

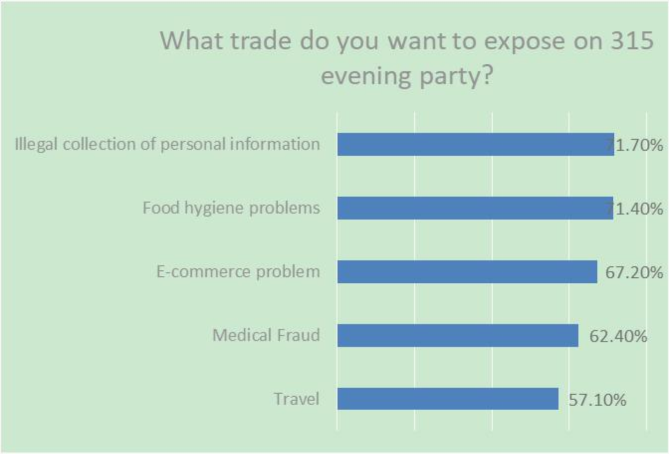

Fig. 1. Public awareness of personal information protection has increased dramatically

Different users of the library under the cloud computing platform have different considerations for information security. The illegal collection of personal information may not only exist at the user's personal level, but may also exist at the library or platform provider level. Individual users are concerned about personal privacy and ways in which privacy

This article is one of the research results of the 2017 Research Working Committee Fund Project No.: 17-YB-08

Copyright © 2018, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

may be compromised in order to better protect privacy. For library or platform providers, they are more concerned with sensitive data security and business continuity operations, and therefore more concerned with data control issues.

A. Library Security Issues

The cloud computing platform needs to provide users with secure and reliable data storage and network services. In this new mode, the user's data is not stored on the local computer, but stored in the remote server, which increases the privacy concerns. Data leakage, data tampering and data loss in the cloud computing platform are embedded in the process of data collection, transmission and storage in the process of running the cloud platform.

1) Security Risks in the Process of Library Data Collection

Data collection is the first step in the process of big data processing. At present, the commonly used data collection methods in the industry include sensor collection and data retrieval classification tools such as Baidu and Google. In the era of big data, smart devices have become essential daily necessities in people's lives. When people use smart devices, smart devices transfer incoming data to the cloud at any time. This process is undetectable or unstoppable by individuals. For example, some commercial companies use cookies to snoop on users' online activities and viewed web pages, use bots to control individual computers and master which services users use in the cloud. Any operation of the user will leave traces on his terminal and can be obtained through certain channels.

2) Security Risks in the Process of Library Data Transmission

At present, the transmission methods of most library data mainly include two methods: wired network transmission and wireless network transmission.

Wired network transmission is the traditional way of data transmission, and data is transmitted through tangible media such as metal wires and optical fibers. In this case, the data is not easily stolen by others, and the confidentiality is strong. Wireless transmission refers to a way of transmitting data using wireless technology. With the development of wireless technology, its flexibility and high cost performance have attracted more people to choose this transmission method. But at the same time, the openness and mobility of wireless networks also make management more difficult. “Wireless networks transmit data through radio waves, and almost any wireless network user can access the data in the coverage area of the data terminal, and anyone can intercept and insert data within this range, and arbitrarily steal data or destroy[5]. ”Data is more vulnerable to malicious attacks by criminals during wireless transmission, such as hacking, password cracking, malicious tampering, etc., resulting in loss of information and leakage of privacy.

3) The Security Risks in the Library Data Storage Process

The data storage process contains two aspects: data storage system and personal data storage.

The data storage system has the ability to manage large-scale data, provide basic support for data processing technology, and is the security guarantee before data is effectively utilized. To ensure the security of data, the data must be protected to distribute in each system. Existing data protection technologies, such as login access control and access control, cannot fully protect data, which makes data more vulnerable to malicious attacks such as data theft, tampering or destruction.

The personal data storage device stores all personal information such as library users' reading behaviors, identity characteristics, personal hobbies and habits, and social relationships, and generates data interaction with the network terminal through automatic transmission[6]. A large amount of complicated data often makes it difficult for readers to completely delete it in a simple way, because the reader not only needs to delete the data stored in the smart device, but also deletes the data uploaded to the cloud through the smart device. This will involve all aspects of the problem, due to the sharing of cloud computing hardware resources, before the storage resources are reassigned to new users, if a complete data erasure is not performed, the new user may restore the data of the previous user through data recovery technology, resulting in data leakage. After the user issues a delete command, the cloud computing platform may be illegally restored if it does not completely delete all the data stored in the platform, bringing out data leakage.

III. THE USE OF KEY TECHNOLOGY IN THE LIBRARY CLOUD PLATFORM

Cloud computing security is a worldwide problem. In addition to the technology itself, the problems in the era of big data contain social, economic, legal and other factors, which make the solution to privacy security issues diversified. According to the security issues and the characteristics of the cloud computing model, combined with the library comprehensive service content, this paper proposes a solution to the library big data encryption under the cloud platform.

Data encryption technology is one of the most important protection methods for information security in cloud computing. The basic process of data encryption is to use an algorithm to process data or files that are originally plaintext. The processed data is an unreadable code, usually called "ciphertext". The original content of the plaintext can only be redisplayed after the corresponding key is entered in the ciphertext[7]. In this way, data is not read and stolen by illegal people. The reverse process of the encryption process is decryption, which is the process of converting the encrypted ciphertext into its original data after being processed by the corresponding key.

In the algorithm of data encryption, it can be classified according to whether the encryption key and the decryption key are the same. If the encryption key is the same, it is called a symmetric password; otherwise, it is called an asymmetric password.

A. Symmetric Encryption Algorithm

Symmetric cipher is a traditional cipher, which has the advantages of low computational cost, fast encryption speed and high security emphasis. The sender preprocesses the plaintext data (original data) using the shared key obtained by the symmetric encryption algorithm, converts the plaintext into ciphertext, and sends it out. After receiving the ciphertext, the recipient decrypts the ciphertext using the same shared key, and restores it to the original plaintext data for reading. Symmetric encryption algorithm uses the same key for the encryption and decryption process, which is a symmetric relationship, as shown in Fig. 2.

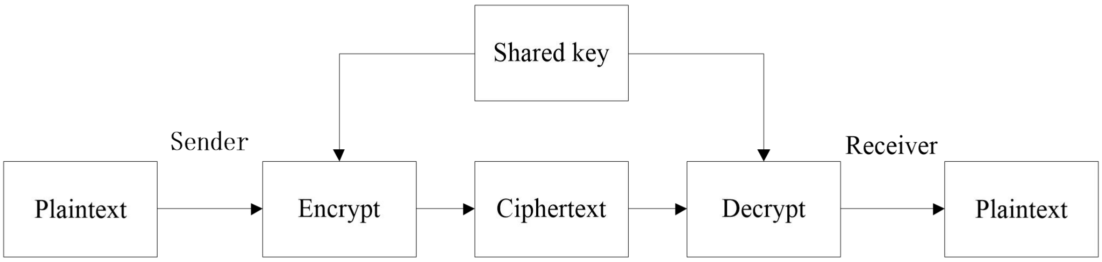

Fig. 2. Symmetric encryption

Suppose the sender (user/reader) A needs to send a message with the plaintext X to the recipient (library/database clerk) B, however, due to the fear that the data is stolen or tampered with during the transmission process, A uses a symmetric encryption algorithm to generate a ciphertext X' after the plaintext X is operated by the shared encryption function Gk, and B obtains the original plaintext X by re-processing the received ciphertext by the shared encryption algorithm Gk to achieve the purpose of data transmission. Among them, the sender(in user/reader) A: X' = Gk(X) and the receiver B(library/database vendor): X = Gk(X').

B. Asymmetric Encryption Algorithm

The asymmetric encryption algorithm is a key technology that is more suitable for library users and platform administrators at this stage. In this algorithm, encryption and decryption use two different keys, and it is impossible to derive another key from one key. The encryption secret key can be disclosed by the library / database business to users / readers, known as the public key; the decryption cipher is only known by the library / database business, known as the private key, and vice versa. If A chooses to publish its key, anyone else can use this public key to encrypt and transmit his message. The private key is kept in secret. Only the private key owner can use the private key to decrypt the ciphertext. Objectively, the identity authentication of the messenger is completed, as shown in Fig. 3.

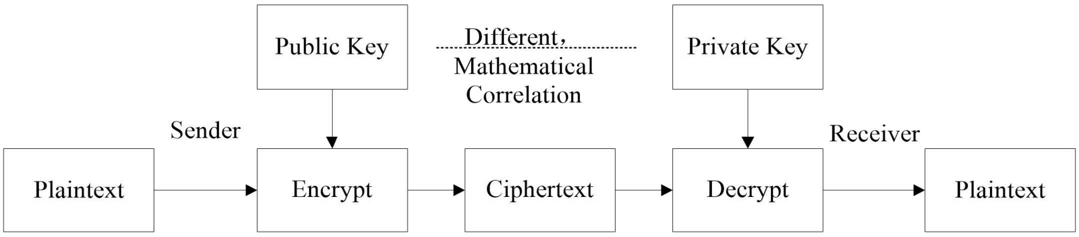

Fig. 3. Asymmetric encryption

In the public key cryptosystem, a key of the key pair is selected as a public key (PK) and shared with other parties. As public information, it uses another key as a private key (SK), which is stored by the user himself as private information. The encryption algorithm E and the decryption algorithm D are also shared with the public key.

When the two communicating parties communicate, the sender A encrypts the plaintext X to be transmitted using the public key $\mathrm{PK_B}$ disclosed by B before transmitting the information (E operation), Receiver B can decrypt the original private text by decrypting the data with its own private key $\mathrm{SK_B}$ (D operation) after receiving the information:

$$\mathrm{D}_{\mathrm{SK_B}}(Y) = D_{SK_B}(E_{PK_B}(X)) = X \tag{1}$$

Different from symmetric encryption, the encryption key is public, but it cannot decrypt the data through it. It can only be decrypted by the corresponding private key, namely

$$\mathrm{D}_{\mathrm{PK_B}}(E_{PK_B}(X)) \neq X \tag{2}$$

The public key and the private key are just the different names of the two keys. They are completely symmetrical, that is, the public key can actually be used as the "private key", and to some extent, the "private key" and the "public key" can be exchanged, namely

$$\mathrm{E}_{\mathrm{PK_B}}(\mathrm{D}_{\mathrm{SK_B}}(X)) = D_{SK_B}(E_{PK_B}(X)) = X \tag{3}$$

Asymmetric cryptography has high encryption intensity and solves the key management problem. Through the special key distribution system, the key does not need to be transmitted in the process of communication, so that even if the user increases substantially, the key will not spread out, and the security is greatly improved. Asymmetric cryptography can not only be used as encryption algorithm, but also be used for digital signature and symmetric encryption key distribution and management, which is suitable for the library cloud platform network openness requirements.

C. *YC-R Encryption Algorithm*

In the traditional symmetric encryption algorithm and asymmetric encryption algorithm, users / readers log in from the client to the library / database vendor cloud platform, without any processing of data, upload or download directly. In this process, there is a risk of data exposure, and the source of the risk may even be a cloud storage service provider. Based on the security status of the library and the characteristics of cloud computing, this paper proposes a YC-R encryption algorithm for the client/reader and cloud segment. Among them, asymmetric encryption/decryption algorithm is implemented in cloud as usual for library cloud computing platform, and YC-R encryption/decryption algorithm is introduced in client side. The whole process is shown in Fig. 4.

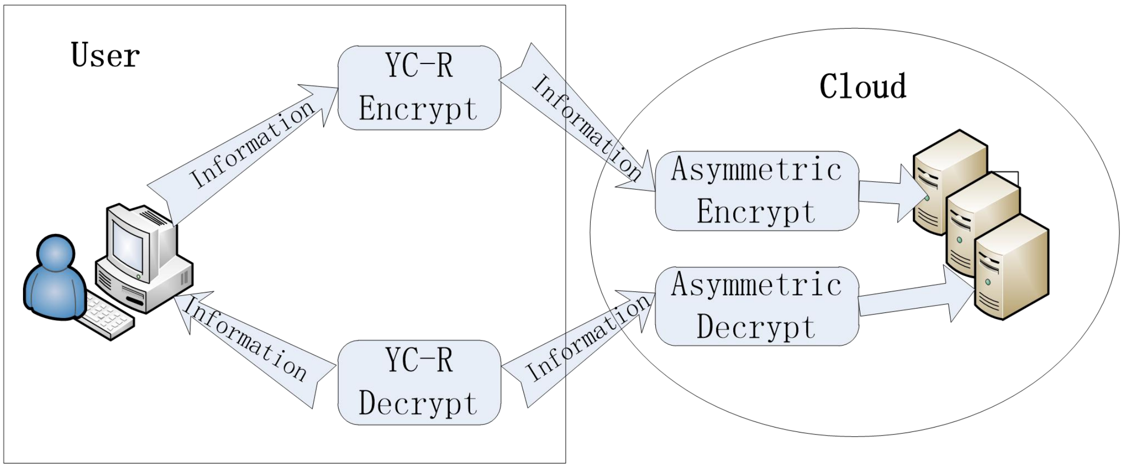

Fig. 4. YC-R strategy overall architecture

User/reader encrypts/decrypts the data through YC-R encryption algorithm embedded in the program before uploading to the cloud to prevent information leakage during uploading. Then asymmetrical encryption / decryption is performed with the help of the operation capability of cloud computing platform, and finally the data is stored safely in the cloud platform of database vendors.

The core of the YC-R algorithm is any piece of information in the user/reader information. Assuming it is INFO, we use a camouflage algorithm to camouflage it. We remember that the value after the disguise is FAKE, the camouflage and the restoration algorithm are as follows:

$$\mathrm{FAKE} = \mathrm{INFO} - \mathrm{K}_1 \mathrm{mod} \mathrm{K}_2 + \mathrm{K}_2 \tag{4}$$

$$\mathrm{INFO} = \mathrm{FAKE} + \mathrm{K}_1 \mathrm{mod} \mathrm{K}_2 - \mathrm{K}_2 \tag{5}$$

Among them, $\mathrm{K}_1 = \mathrm{random}(1, 2^{64})$, that is a randomly generated constant. $\mathrm{K}_2 = \mathrm{random}(1, \mathrm{S})$, that is, a shaped random number associated with the current file size S. $\mathrm{K}_1\mathrm{K}_2$ are private keys that only the user/reader knows. At the same time, since the MOD function does not have an inverse function[9], it is impossible to crack the FAKE value to obtain the original data by inverse derivation, which ensures the security of the algorithm.

In addition, the YC-R encryption algorithm randomly selects a certain amount of dry data from the interference database, and the interference database matches the real database type. It is possible to segment real data, insert interference data into real data, and achieve the purpose of hiding real data by introducing a large number of interference items. For the hiding of real data, the adjustment is made by changing three parameters $\mathrm{m}, \theta, \sigma$, that is, the insertion process of interference: For every m bytes of real data, there is a possibility of θ, inserting n bytes of interference data, and θ, σ can decide how much to insert interference data. $\theta\,(0 \le \theta \le 1\ )$ indicates the possibility of inserting m bytes of real data into n bytes of interfering data, σ indicates the upper limit of inserting n bytes of interference data, $\mathrm{n} = random(0, \sigma)$, that is, m tends to 0, θ tends to be infinite and σ tends to be infinite. The more interference items are introduced, the better the real data hiding and the most system resources consumed.

D. *Performance Analysis of YC-R Encryption Algorithm*

In order to analyze the performance of the YC-R encryption algorithm more intuitively, we compare it with the traditional asymmetric encryption algorithm. The test results are shown in Fig. 5.

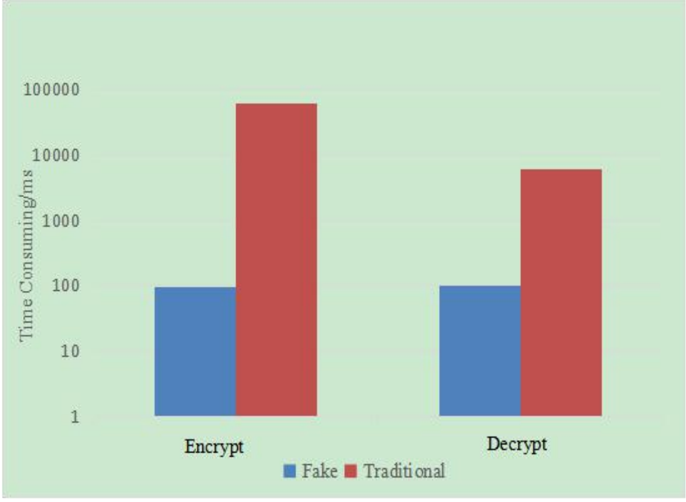

Fig. 5. Comparison of improved algorithms with traditional asymmetric algorithms

The results show that the YC-R encryption algorithm consumes much less time than the traditional asymmetric encryption algorithm, and can quickly complete the encryption/decryption of the data, so it can be applied to the reader's privacy document.

In addition, we analyze the introduction of the interference term, set m=5, θ =0.6 and σ =3, and compare the introduction of the interference term with the filter consumption time, as shown in Fig. 6.

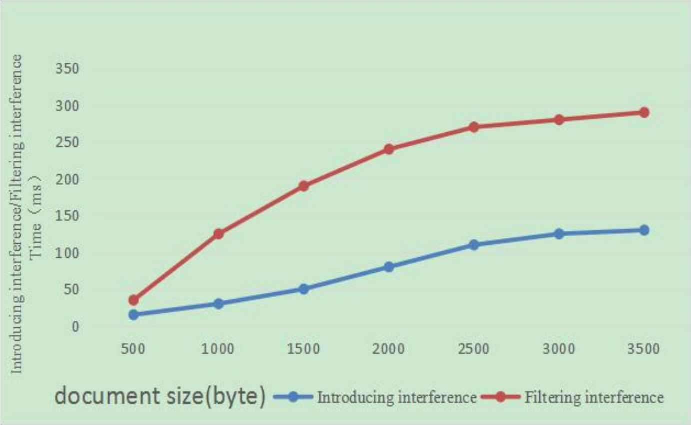

Fig. 6. Effect of document size change on the introduction of interference

Through data analysis, it can be concluded that the time taken to introduce interference and filter out interference increases with the increase of the document, and at the meantime, the time spent filtering out the interference is significantly more than the time taken to introduce the interference. The introduced interference data and the real data occupy the storage space. When more interference is introduced, more storage space will be consumed. Therefore, when the interference is introduced in the face of an appropriate number of user privacy data, the effect is the best. Another feature of the cloud platform is that the storage space is very powerful and relatively inexpensive, making up for the extra storage space required by the YC-R algorithm.

IV. Conclusion and Deficiency

This paper analyzes the existence and causes of library security problems under the cloud computing platform, that is, leakage; tampering and loss mainly exist in data collection, transmission and storage. Combining the particularity of the library under the cloud platform, improved data encryption/decryption architecture is proposed. The improved algorithm YC-R algorithm improves the privacy of data, and shortens the data processing time as much as possible compared with the traditional encryption algorithm, and solves the data privacy security problem from the technical level. There are still some limitations in the discussion of this article: (1) When the YC-R algorithm faces massive user privacy data, it needs to consume a large amount of storage space, and the usage time also increases accordingly. Therefore, it is necessary to consider the hierarchical processing of library user privacy, and prioritize user data with high privacy level. (2) Cloud storage security is not only a technical issue, but also includes issues such as system standardization, regulatory mode, and legal level, including the establishment of a complete cloud storage architecture, allowing different terminals to share data more securely and conveniently. Only by realizing the combination of strategy, technology and human factors, and providing information security to the library construction, maintenance and operation of the cloud platform, can we provide readers with better cloud computing services.

References

[1] China Telecom Network Security Lab. *Cloud Computing Security Technology and Application* [M]. University of Electronics Industry Press, 2012: 7-17.

[2] Feng Dengguo, Zhang Min, Zhang Wei & Xu Zhenyun. *Research on Computing Security*[J].Journal of Software,2011,22(1):71-83.

[3] Barroso LA, Dean J & Holzle U. *Web search for a planet: The Google cluster architecture.* IEEE Micro, 2003,23(2):22–28.

[4] China Information Network, 2018 Chinese Internet users defrauded and rights investigation report (full text) [EB/OL]. [2017-12-12]. http://www.askci.com/news/chanye/20180326/083922120424.shtml.

[5] Li Yuan & Wang Yanhong. *Wireless Network Security Threats and Countermeasures*[J]. Modern Electronic Technique,2007(5):91-94.

[6] Ma Xiaoting. *Study on the Privacy Protection of Library Personalized Service Readers in the Age of Big Data*[J]. Library Tribune,2014 (2):84-89

[7] Deng Qian. *Research on Cloud Computing Security Mechanism Based on Hadoop*[D]. Nanjing University of Posts and Telecommunications Master's Thesis, 2009: 6-17.

[8] Ge T,Zdonik S.Fast,Secure Encryption for Indexing in a Column -Oriented DBMS[C]. Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, 2007:676–685.