A New Model of Data Protection on Cloud Storage

Peng Jing
College of Computer and Information Science, Chongqing Normal University, Chongqing, China
Email:pj93@163.com

Abstract—This paper focuses on studying cloud storage data protection model and implementing encrypted storage of user data in double-key form. User data are encrypted with symmetric encryption algorithm and this secret key is encrypted with asymmetric encryption algorithm. The private key is managed and controlled by users. In this way, users guarantee the security of their own data with the sole private key. Providing effective protection for uploading and downloading of user data and good data storage and transmission support for cloud storage application through the data protection mechanism deployed at client, thus solving the biggest obstacle, i.e. security of cloud data in the application of cloud storage enterprises currently to a certain extent.

Index Terms—Cloud Computing; Cloud Storage; Data Protection

I. INTRODUCTION

Cloud computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the electricity grid.

Cloud computing is a paradigm shift following the shift from mainframe to client–server in the early 1980s. Details are abstracted from the users, who no longer have need for expertise in, or control over, the technology infrastructure “in the cloud” that supports them. Cloud computing describes a new supplement, consumption, and delivery model for IT services based on the Internet, and it typically involves over-the-Internet provision of dynamically scalable and often virtualized resources [1]. It is a byproduct and consequence of the ease-of-access to remote computing sites provided by the Internet. This frequently takes the form of web-based tools or applications that users can access and use through a web browser as if it were a program installed locally on their own computer. NIST provides a somewhat more objective and specific definition here. The term “cloud” is used as a metaphor for the Internet, based on the cloud drawing used in the past to represent the telephone network, and later to depict the Internet in computer network diagrams as an abstraction of the underlying infrastructure it represents. Typical cloud computing providers deliver common business applications online that are accessed from another Web service or software like a Web browser, while the software and data are stored on servers. A key element of cloud computing is customization and the creation of a user-defined experience [2]-[4].

Most cloud computing infrastructures consist of services delivered through common centers and built on servers. Clouds often appear as single points of access for all consumers’ computing needs. Commercial offerings are generally expected to meet quality of service requirements of customers, and typically include SLAs. The major cloud service providers include Microsoft Salesforce, Amazon, and Google.

II. CLOUD STORAGE

Cloud storage is a system integrating various different types of storage devices in the network via application software through clustered application, grid technology or distributed file system for working cooperatively and jointly providing data storage and business access functions.

The concept of cloud storage is an extension of the concept of cloud computing and evolved. If many data management and storage become a cloud computing system’s core processing and computing resources, you need to configure a large number of storage devices in the cloud computing system, the result is the emergence of cloud computing systems into the cloud storage system [5].

A. Structural Model of Cloud Storage System

Cloud storage is a complex system composed of network equipment, storage device, server, application software, public access interface, access network and client program, which provides data storage and business access services through application software with storage device as the core. Figure 1 shows the structural model of cloud storage system.

1) Storage Layer

Storage layer is the most basic part of cloud storage. Storage device can be FC (fiber channel) storage device, IP storage devices such as NAS and iSCSI or DAS storage devices such as SCS1 or SAS. There are often a huge number of storage devices in cloud storage which are distributed in different areas and connected with each other via the internet or FC network.

There is a unified storage device management system above storage devices, which can realize logic virtual processing and multi-link redundancy management of
storage devices and condition monitoring and fault maintenance of hardware equipment.

![Diagram of cloud storage system](image)

Figure 1. Structural model of cloud storage system

2) Basic Management Layer

Basic management layer is the core part of cloud storage as well as the part most difficult to be achieved in cloud storage. Basic management layer realizes cooperative work among multiple storage devices in cloud storage and allows multiple storage devices to provide the same service and higher, stronger and better data access performance through cluster, distributed file system and grid computing etc.

CDN content distribution system and data encryption technology guarantee that data in cloud storage will not be accessed by unauthorized users. Meanwhile, various data backup and disaster recovery technologies and measures can ensure that data in cloud storage will not be lost and guarantee the security and stability of cloud storage itself.

3) Application Interface Layer

Application interface layer is the most flexible part of cloud storage. Different cloud storage operating units can develop different application service interfaces and provide different application services according to the actual business type, such as video monitoring application platform, IPTV and video on demand application platform, network hard disk reference platform and remote data backup application platform.

4) Access Layer

Any authorized user can log on cloud storage system and enjoy cloud storage services via a standard public application interface. In different cloud storage operating units, cloud storage providers different types and means of access.

B. Status of Cloud Storage Data Security Protection

The most prominent problem of cloud storage is data security [6]. Many cloud storage manufacturers had insecurity accidents such as user data stolen. For example, Apple iCloud known for “high security of storage data theoretically” had a serious incident that a hacker sneaked on users’ iCloud accounts and deleted all data of users, causing irretrievable loss of users.

Generally, cloud storage system provides users with data storage by establishing a large-scale distributed data storage cluster which implements centralized management, operation and maintenance [7]. However, to solve the security problem of cloud storage currently, data encryption technology is used. Ways of encryption in general use currently have some problems in security coefficient and performance. Symmetric encryption based on DES and AES etc. is widely used in large data volume transmission due to its rapid encryption and decryption. However, it has such inherent problems as low security and key transmission caused by the use of the same key by encryption and decryption. Asymmetric encryption system represented by RSA algorithm has higher strength and public key-private key system is more flexible. However, the common problem of asymmetric encryption and decryption algorithm is heavy computation. Therefore, asymmetric algorithm does not apply to encryption and decryption of large data volume [8].

Therefore, this paper puts forward a cloud storage-oriented secure data storage system model and strategies and aims at placing the right of data control completely in data owners with storage cloud only providing the space for data storage, thus solving the problem of cloud storage security of private information.

III. DATA SECURITY MODEL

A. Data Encryption Algorithm

1) AES Algorithm

It is also called as Rijndael encryption algorithm which is based on arrangement and substitution operation. Arrangement refers to rearrangement of data and substitution refers to replacement of one data unit with another. AES implements arrangement and substitution operation with several different methods [9-11]. AES is an iterative and symmetric key block cipher, which can use 128, 192 and 256-bit key and use 128-bit (16-byte) block encryption and decryption data. Different from public key cipher, symmetric key cipher uses the same key to encrypt and decrypt data. Bits of enciphered data returned through block cipher are the same as input data. Iterative encryption uses a loop structure and input data are substituted and replaced repeatedly in this loop. This standard is used to substitute the original DES and has been analyzed in multiple aspects and widely used in the world.

![AES encryption and decryption block diagram](image)

Figure 2 is AES encryption and decryption block diagram.
2) Encryption Transformation

Assume that X is 128-bit clear text input of AES and Y is 128-bit cipher text output, then AES cipher text Y can be expressed with the following complex transformation:

\[ Y = A_{k(r+1)} \cdot R \cdot S \cdot A_{k(r-1)} \cdot \cdots \cdot C \cdot R \cdot S \cdot A_{k_1}(X) \]  

(1)

where “*” represents compound operation. \( A_{k_i} \) refers to a transformation of X. \( A_{k_i}(X) = X + K_i \) (Ki is the subkey in the i-th round and XOR operation of bit string). S refers to S box substitution, i.e. substitution of each byte with S-Box. S-Box is a given conversion table. R refers to row substitution and C refers to column substitution.

\[ s'(x) = a(x) \times s(x) \]  

(2)

\[ \begin{bmatrix}
  s_{0,c} \\
  s_{1,c} \\
  s_{2,c} \\
  s_{3,c}
\end{bmatrix} =
\begin{bmatrix}
  02 & 03 & 01 & 01 \\
  01 & 02 & 03 & 01 \\
  01 & 01 & 02 & 03 \\
  03 & 01 & 01 & 02
\end{bmatrix}
\begin{bmatrix}
  s_{0,c} \\
  s_{1,c} \\
  s_{2,c} \\
  s_{3,c}
\end{bmatrix} \]  

(3)

\( \times \) is a special multiplication.

3) Decryption Transformation

Decryption transformation is inverse transformation of encryption transformation, which is not discussed in detail here.

Implementation of AES encryption/decryption algorithm

a. Block encryption

Table I is the comparison table of three different types of AES encryption key block size and the corresponding number of rounds of encryption.

When encryption starts, various bytes input in block are put in a matrix State according to table II. If ABCDEFGHIJKLMNOP are input, the input block will map them to the state matrix State in table II.

### Table I

<table>
<thead>
<tr>
<th>Type of AES</th>
<th>Key Length</th>
<th>Packet Size</th>
<th>Encryption rounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES-128</td>
<td>4 words</td>
<td>4 words</td>
<td>10</td>
</tr>
<tr>
<td>AES-192</td>
<td>6 words</td>
<td>4 words</td>
<td>12</td>
</tr>
<tr>
<td>AES-256</td>
<td>8 words</td>
<td>4 words</td>
<td>14</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

b. S-box transformation

The following transformation \( S[A] \) is implemented for any element A of input matrix:

Any element A is a 8-bit binary number from the perspective of storage. Hexadecimal number \( x \) represented by the first four bits and hexadecimal number \( y \) represented by the last four bits are calculated. For example, when \( A = 11010100 \), \( x = 1101 \) and \( y = 0010 \).

The value of \( S[A] = S[x,y] \) is found in S2Box given by AES algorithm (matrix with 16 rows and 16 columns, where each element is a byte. The specific S-Box is omitted). For example, when \( A = 11010100 \), \( S[A] = S[c,4] = S[1c] = 00011101 \). Or, \( A = b_6 b_5 b_4 b_3 b_2 b_1 b_0 \) is directly transformed into \( S[A] = b_7 b_6 b_5 b_4 b_3 b_2 b_1 b_0 \) with the following formula:

\( b_7 = b_0 + b_6 \mod 8 + b_4 \mod 8 + b_2 \mod 8 + b_0 \mod 8 + c \mod 4 \)

where, \( c = (c_6, c_5, c_4, c_3, c_2, c_1, c_0) = (0, 1, 1, 0, 0, 1, 1) \).

c. Row transformation

In row transformation, the first row of intermediate state matrix State remains unchanged and the second to forth rows are transformed as follows [12]. That is, the state matrix in table III turns into that in table IV.

### Table III

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

d. Column transformation

Column transformation refers to transformation of intermediate state matrix State column by column. Its transformation is the following matrix operation.

\[ \begin{bmatrix}
  s_{0,c} \\
  s_{1,c} \\
  s_{2,c} \\
  s_{3,c}
\end{bmatrix} =
\begin{bmatrix}
  02 & 03 & 01 & 01 \\
  01 & 02 & 03 & 01 \\
  01 & 01 & 02 & 03 \\
  03 & 01 & 01 & 02
\end{bmatrix}
\begin{bmatrix}
  s_{0,c} \\
  s_{1,c} \\
  s_{2,c} \\
  s_{3,c}
\end{bmatrix} \]

With the operation above, the original column is changed into a new column represented by the following formula:

\( s(0,c)' = ((02) \times s(0,c)) + ((03) \times s(1,c)) + s(2,c) + s(3,c) \)

\( s(1,c)' = s(0,c) + ((02) \times s(1,c)) + ((03) \times s(2,c)) + s(3,c) \)

\( s(2,c)' = s(0,c) + s(1,c) + ((02) \times s(2,c)) + ((03) \times s(3,c)) \)

\( s(3,c)' = ((03) \times s(0,c)) + s(1,c) + ((02) \times s(2,c)) + ((03) \times s(3,c)) \)

where + is bitwise XOR operation and multiplication \( \times \) is calculated according to modular multiplication congruence rules introduced below.

Modular multiplication congruence rules used in column transformation are different from the multiplication we generally use. As each element is a byte, this byte can be considered as a formal septic polynomial. That is, \( b_6 b_5 b_4 b_3 b_2 b_1 b_0 \) is considered as \( b_0 + b_4 + b_6 x^3 + b_6 x^4 + b_6 x^5 + b_6 x^6 + b_0 x + b_0 \). For example, \( \{11011001\} = \{d9\} \) can be considered as \( x^7 + x^6 + x^4 + x^3 + 1 \). Column transformation expects to transform a byte into a new one. Therefore, it is necessary to turn the result of multiplication of two formal septic polynomials into a new formal septic polynomial and then it can be reverted to the length of a byte. Here, the congruence multiplication of an eight-time irreducible polynomial is used, i.e. the result of multiplication of two septic polynomials is divided by this eight-time irreducible polynomial and then its residue is taken. In AES, the eight-time irreducible polynomial is \( m(x) = x^8 + x^4 + x^3 + x + 1 \) for example:
\((x^6 + x^4 + x^2 + x + 1) \times (x^7 + x + 1) = x^{13} + x^{11} + x^9 + x^8 + x^6 + x^4 + x^3 + 1\)
\((x^{13} + x^{11} + x^9 + x^8 + x^6 + x^4 + x^3 + 1) \mod (x^8 + x^4 + x^3 + x^2 + 1) \equiv x^7 + x^6 + 1 \mod n\) (7)

(a, φ(n)) ≡ 1 (mod n)  (6)

\(\phi = \left\{ \begin{array}{ll} 1 & \text{if } a \equiv 1 \pmod{n} \\ 0 & \text{if } a \equiv 0 \pmod{n} \end{array} \right.\)

If two positive integers a and n are relatively prime, the Euler function \(\phi(n)\) of n can allow the following equation to be true:

\[ a^{\phi(n)} \equiv 1 \pmod{n} \] (6)

\(\phi(1)=1, \phi(2)=1, \phi(3)=2,...\) is obtained from the definition.

When p is a prime number, \(\Phi(p) = p - 1\).

**RSA algorithm**

RSA is an asymmetric cryptographic algorithm. “Asymmetric” means that this algorithm requires a pair of keys, one of which is used for encryption and the other of which is used for decryption. It can resist all known cryptographic attacks so far and is recommended as public key data encryption standard by ISO.

[Definition:] Euler function \(\Phi(n)\) is a function defined in the set of positive integers. The value of \(\Phi(n)\) equals the quantity of numbers relative prime with n in sequence 0, 1, ..., n - 1.

\[ \Phi(1) = 1, \Phi(2) = 1, \Phi(3) = 2, ... \] is obtained from the definition.

When p is a prime number, \(\Phi(p) = p - 1\).

Encryption transformation: cipher text after encryption of clear text m, 1 < m < n, is:

\[ c = m^e \pmod{n} \] (7)

Decryption transformation: clear text after decryption of cipher text c, 1 < c < n, is:

\[ m = c^d \pmod{n} \] (8)

**B. Cloud Storage Data Protection Model**

Data security protection mechanism is deployed in basic management layer in Figure 1 according to the structural model of cloud storage system. That is, users can encrypt and decrypt data with double encryption [17].

Data key is used to encrypt user data with AES encryption algorithm. As symmetric encryption algorithm is characterized by small mathematical computation and fast encryption, it applies to encryption of mass data and has higher security than DES algorithm.

Data key also requires encryption management. The volume of user data is relatively smaller. Therefore, asymmetric encryption algorithm - RSA algorithm can be used for encryption. RSA key is generated when users register. Users can randomly select digits within a scope as private key. Then, the system will generate the specific public key of the user through this private key. The public key is submitted to the cloud of a reliable party for data encryption and private key is used for decryption. The private key is possessed by users only. In this way, users guarantee their data security with the sole private key [18] - [19].

**1) Encryption Process**

When starting the encryption, the client is extracted from the keystore of the received user data corresponding public key symmetric encryption algorithm. In the encryption process, consists of a symmetric encryption algorithm (specific algorithm based on user need to select) key generator to generate a random key that contains parity information, and will contain parity information via asymmetric key encryption algorithm encrypted. Finally, the encrypted data information processing algorithm and symmetric key encryption algorithm ciphertext together as a data packet stored in the cloud. Repeat the process until all encrypted and transmitted packets, thus completing the entire encryption process.

In the implementation of the data encryption process, the huge amount of data the user data using a symmetric encryption algorithm, while a very small quantity of data is relatively symmetric encryption algorithm is an asymmetric encryption algorithm key for the two encryption key of the ciphertext data stored together in the cloud storage center, the client saves only asymmetric encryption algorithm and the decryption key. This can effectively avoid the presence of symmetric encryption algorithm using the same key and difficult problems brought about by key management, but also solved due to asymmetric encryption algorithms are not suitable for large amounts of data encryption problems caused by the storage efficiency.

**2) Decryption Process**
When the data is decrypted, the decryption side need to use asymmetric encryption algorithm for solving symmetric key to decrypt the encryption algorithm used to restore the key, according to the use of symmetric key encryption algorithm to decrypt the packets, restore all original, thus, completed a packet decryption process. Repeat the process until all the packets decrypted, so get the raw data before encryption.

Using symmetric encryption algorithm and encryption algorithm combining asymmetric encryption and decryption scheme, a symmetric encryption algorithm to solve the key management problem, and asymmetric encryption algorithm to solve the large amount of computation, is not suitable for a large number of problems to encrypt the data. In this solution, there is only a symmetric encryption algorithm key management problems, the user holds the respective asymmetric encryption algorithm key, the corresponding public key stored in the cloud particular warehouse, data is exchanged between the user, the target users download from the cloud asymmetric encryption algorithm and key, and use the key to encrypt data, two kinds of combination of encryption methods to encrypt and ciphertext data transferred and stored in the cloud. Target users to access the data and use their corresponding private key asymmetric encryption algorithm to decrypt the data. Thus, to achieve a double-encrypted data, to ensure data security.

IV. EXPERIMENTAL RESULT

The maximum loading capacity of each node in cloud is defined as 2 concurrent uploading (encryption) and 3 concurrent downloading (decryption). When five nodes have saturated processing capacity, stop adding concurrent client request, allow the system to maintain a steady state and start to record the node state. The monitor program uses NMON to detect the average CPU occupancy rate. Then, continue to add the number of concurrent client and allow the cloud to automatically allocate new nodes. When it extends to 250 nodes, monitoring the cloud state and record data. When a single node is in full load, its average data throughput rate is 4.5MB/S for uploading and 9MB/S for downloading (client nodes and cloud nodes are in the same LAN environment). The overall throughput rate of the cloud respectively reaches 1GB/S for uploading and 2GB/S for downloading. During operation at full load, the average CPU occupancy rate of all nodes is above 95%. When concurrent tasks on nodes reduce to one uploading and one downloading, CPU occupancy rate still remains above 95% and the throughput rate decreases to 3MB/S (uploading) and 13MB/S (downloading) on average.

Figure 4. and Figure 5. show the test result of CPU performance and data transmission rate.

It can be found through the analysis on data above that the self-adaptation feature of cloud reaches the theoretical infinitely high transmitting capacity after the number of concurrent users exceeds 100. Therefore, it is very important to reasonably control the maximum concurrent load of nodes. Ideally, it is best to realize node operation at full load.

![Figure 4. The test results of CPU occupancy rate](image1)

![Figure 5. The test result of data transmission rate](image2)

V. CONCLUSION

The sudden emergence of cloud computing, in all fields are showing good prospects, but in the cloud data storage environments rely on network resources, data security problem is particularly prominent. Users of cloud storage security challenge, and enterprise data can not be easily ported to secure cloud storage medium range of environmental problems, leading to difficulties in the popularity of cloud computing.

In order to effectively address these problems, this paper based on the data stored in the cloud storage service model, through the use of symmetric and asymmetric encryption and a combination of both encryption and decryption programs to achieve a good safety but also to read and write memory efficient data storage policies. With cloud storage related technologies develop and mature, and legal aspects of legislation and relevant system of continuous improvement, cloud storage environments secure storage of data will eventually be conquered, I believe that in the near future, cloud storage, like other applications on the Internet environments, profound influence on our way of life.

ACKNOWLEDGMENT

This word is supported by the Science and Technology Project Affiliated to the Education Committee of Chongqing, China (Grant No. KJ130611).

REFERENCES


Peng Jing received her M.E. degree in software engineering from University of Electronic Science and Technology of China, Sichuan, Chengdu in 2006. Her research interests include digital image processing and information hiding.