International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org Volume 3 Issue 8 ǁ August 2014 ǁ PP.24-30

The Architecture of Cloud Storage Model Based On Confusion Theory

Xirui Quan[1], Genqing Bian[1,2], Bilin Shao[2]

([1.]School of Information and Control Engineering, Xi'an University of Architecture and Technology, China;
[2.]School of Management,Xi'an University of Architecture and Technology, China)

ABSTRACT : *Aiming at the problem of data security in cloud storage system,the paper proposes a cloud storage model based on confusion theory. Data confusion algorithm and slicing technology are utilized to slice the data in accordance with characteristics of data attributes in cloud storage system. Confusion strategy is formed by adopting confusion algorithm to confuse the relation of among data slicing. Thus, it is on this premise that data is available to implement effective protection of data via slicing relationship confusion. Experiment shows that the model not only can guarantee data security in cloud storage system, but also reduce system complexity and computing cost.*

KEYWORDS: *Cloud Storage Model, Confusion Algorithm, Slicing Technology, Data Security*

I. INTRODUCTION

Cloud storage is a new concept extended and developed on the basis of cloud computing. And it is a cloud computing system and a storage center of data which is configured with supreme large storage space and integrated with high performance service processing. On one hand, it could be a flexible extended storage resource of cloud computing resource pool[1]; on the other hand, as the administration center of resource which is provided by cloud service, it is the core of cloud computing service center and intensive administration of tenants'data information[2]. Thus, cloud storage is simplify as cloud computing with store function. And it is also an open network environment oriented supporting multiple types which supports massive information management and considers data security and reliability[3].As the indirect service provider which provide cloud services to cloud computing, cloud storage has massive resources pool, the core problem of cloud storage is management and deployment of these resources. According to the comprehensive analysis of cloud storage technology architecture and core functionality, massive resources storage needs to be allocated on demand in real time, security isolation among the data needs to be taken seriously and the unauthorized user needs to be prevented access to the system.[4]. From architecture to analyze, cloud storage system is the cloud computing architecture's storage layer, so cloud services data protection can start from the secure storage. To the end, confusion theory is introduced to establish cloud storage model based on data slicing to protect user data security and improve storage efficiency.

II. CONFUSION THEORY

Confusion theory was first proposed to protect the trademark rights, then gradually be introduced into the areas, such as the protection of intellectual property rights, software confusion, data confusion algorithm and so on. Confuse algorithm's nature is through the obfuscation transformation to transfer the original program,

protecting the safety of original program under the promise that guaranteeing the output would be unchanged, in order to avoid the bad behavior such as decompiled[5][6], the basic principle of obfuscation is shown in Figure 1

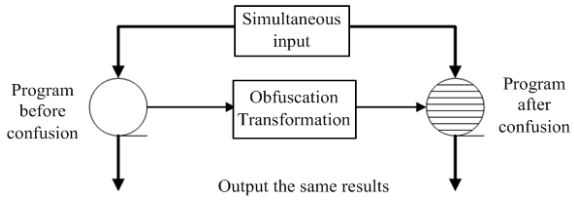


Figure 1 the principle of obfuscation transformation

Confusion algorithm is utilized to process the application, the functions between the treated application and the untreated application are the same, so the application's structure being processed under the confusion dealing has changed a lot, make the attackers cannot have a good understanding with them[7]. In practical, even the attacker finally cracked the application under the confusion dealing, he must have paid a high contribution, since it can be thought to have a protection to the software.Suppose *M* is a transformation from original program to the target, if it has the same observable behavior, then M can be recognized as an Obfuscation Transformation. If all of the following conditions be satisfied, *M* can be a legal Obfuscation Transformation: if it cannot be stop or stop at a wrong state, then stop or not stop will be fine; otherwise, it must be stop and output the same results[8].

According to the different objects, obfuscation transformation can be divided into four sections: morphology transformation, the transformation of control flow, data transformation and class structure transformation [9]. The class structure transformation includes class merger, class splitting and type hiding [10].
Class merger: two or more classes in program merged into a class, thus destroying the structure of the original class, achieve the goal that hidden the system design. Assume that classes c_1、c_2 should be merged into one class c_t, firstly merge the functions and filed in the two classes into class c_t, and then rename the same variables or functions.

$$\mu_f : Fieds\ (c_1) \cup Fieds\ (c_2) \to Fieds\ (c_t) \quad (2\text{-}1)$$

$$\mu_m : Methods\ (c_1) \cup Methods\ (c_2) \to Methods\ (c_t) \quad (2\text{-}2)$$

Class splitting: using class c_1 to replace c_2

$$\mu_{split} : Members\ (c) \to 2^{\{c_{t,1}, c_{t,2}\}} \quad (2\text{-}3)$$

Type hiding: there are *n* interfaces in class *c* , $i_1, i_2, \ldots, i_n$ used to declare the type replaced the *c*, assume P_r is a application.

$$\mu_{hidding} : Methods\ (c) \cup publicMeth\ ods\ (P_r) \cap ins\ tan\ ceMethods\ (P_r) \to (i_1, i_2 \cdots, i_n) \quad (2\text{-}4)$$

III. DATA SLICING ALGORITHM

The aim of slicing data in cloud storage system is decomposing the whole logical structure of data, save data as the slicing type, which is prone to confusion processing of data relationship, then the data can be storaged and managed securely.The original semantic of data should not be changed and the data should not be repeated stored while the data has been sliced. Tenant Data Privacy is a set of attribute constraint of data and vertical slicing of data recording, each *TDP* stands for a data slicing, express *TDP* with formula (3-1).

$$TDP = \{Da\ Pri Id, Da\ Pri\} \quad (3\text{-}1)$$

When the data is single data, *DaPriId* refers to confusion value which is corresponding to user data; when *TDP* is combination form of user data, *DaPri* merely stands for a subset of *TDP* and require any two *DaPri* of *TDP* have no intersection. Because of single data is special case of multi-data, the multi-data cases mainly be discussed . Non-compatible slice and compatible slice need be examined while slicing data. Non-compatible slice refers to combination of some data in data slice will result in data leakage, compatible slice refers to combination of data slicing shall not result in data leakage, flow of data slicing algorithm is shown as Figure 2.

Pseudo code of data slicing algorithm:

Begin:

step1:input *x*;(*x* is privacy demand)

step2: if(*x*==slicing is corresponding)

{If slicing is belong to non-Compatible Slice, return to step 1,

again input privacy demand }

Step3: Implement slicing combination strategy;

if(whether slicing is compatible or not)

{If slicing is non-compatible Slicing, return to step 3,

implement slicing combination strategy; }

Largest data slicing;

if (whether there is privacy data which is not sliced or not)

{ There is privacy data which is sliced, return to step 2; }

Found *LST*;

End.

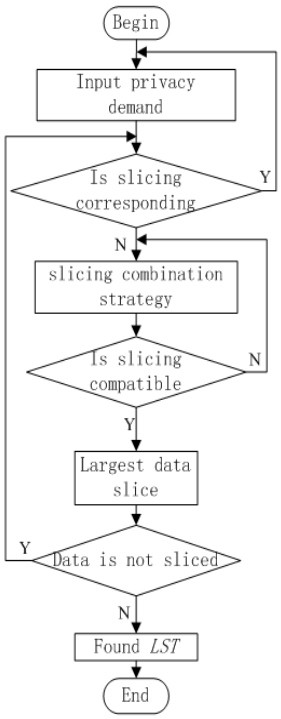


Figure. 2 flow of data slicing algorithm

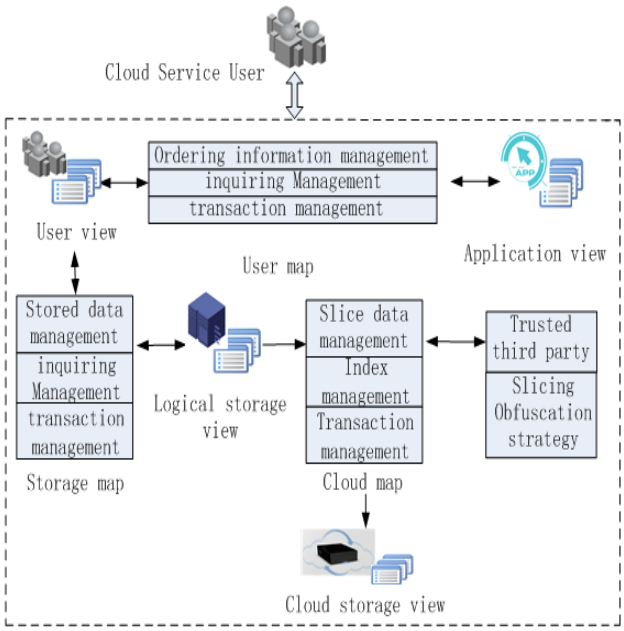


Figure.3 Cloud storage model based on confusion theory

IV. THE ARCHITECTURE OF CLOUD STORAGE MODEL BASED ON CONFUSION THEORY

At present encryption scheme and confusion method to protect user data is a hot research area in data outsourcing field, because data protection based on encryption of data require data which has been encrypted to be decrypted to go on manipulation, which reduces the data processing efficiency. So this paper proposes a data slicing storage using data relationship confusion algorithm. Firstly, according to the secrecy degree of user data cut data attribute into logical storage table of data protection, next hid the relationship among all data slicing after slicing data with confusion algorithm, at last store sliced data as data slicing in logical block table. Cloud storage model is shown as Figure 3.

4.1 Data slicing confusion algorithm

In the light of separating character of data slicing, confuse the relationship of data slicing, and save confusion strategy at a believable third place to implement data slicing confusion and protect from data leaking. Data slicing sigh *DaPriId* refers to each slicing of data, *DaPriId* could be expressed as formula (4-1).

$$DaPriId = E_{key}\left(Id_{TTP} \oplus a^{DSId}\right) \quad (4\text{-}1)$$

Where, E stands for multiplicative homomorphic encryption function in data slicing strategy; *key* is corresponding secret *key*; Id_{TTP} is used for marking every data recording according data slicing strategy; *DSId* is the serial number of data slicing; a is the generating element of cyclic group in data privacy strategy.

The set of data slicing is obtained with slicing algorithm, confusion strategy mainly is used to confused relationship among data slicing and coordinate with believable third party, pseudo code of relationship of data slicing confusion algorithm are following:

Begin:

Input x(x is the set of data privacy slicing)

A: Preset confusion strategy;

According to preset confusion strategy cut x into y (y is data slicing)

Build *DaPriId*;

Build logical view;

if (whether there is relevant mapping or not)

{ If thre is not relevant mapping, return **A**, again get y in x;}

End.

4.2 Reconstructing after data slicing confusion

Store the user data by slicing technology. All the data can be reconstructed combining with the data global identifier Id_{TTP}, confusion strategy and record of each data slicing *DaPriId* when users need to use the data. Steps are as follows:

Step1: Input Id_{TTP} of each marked data record;

Step2: Get the data confusion strategy according to the trusted third party and confusion refactoring module;

Step3: Get the slicing data in sequence combining with confusion strategy and each data slicing *DaPriId*;

Step4: Return all of the data.

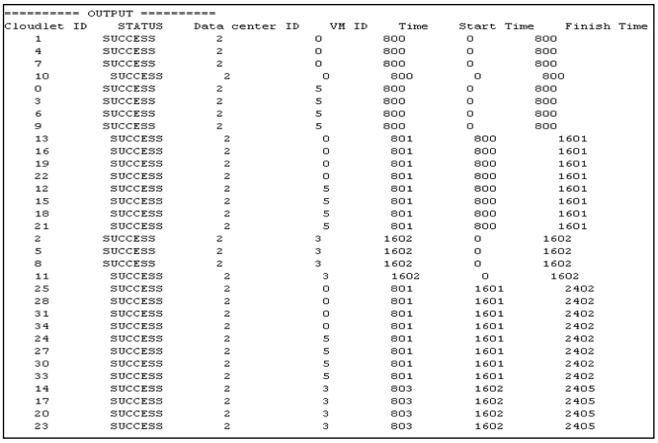
The operability and availability of data mainly be considered about data reconstructing. The data can be reproduced by refactoring algorithm combining with data slicing when the user needs to use the stored data. The pseudocode of refactoring algorithm after confusing data slicing are as follows:

Step1: Input x (x is the Id_{TTP} of data record);
Step2: Get y (y is the confusion strategy stored in the third party);
Step3: Restore slicing data according to y and *DaPriId*;
Step4: Output z (z is reconstructed data);

In conclusion, the paper, according to the characteristics of user data in cloud storage system to slice the data. And generate the confusion strategy by using confusion algorithm to confuse the relation of the slicing data. Then store it in the trusted third party . Thus build the cloud storage model based on the confusion theory and improve the data processing efficiency and security.

V. THE EXPERIMENT AND ANALYSIS

In the virtual environment, the paper realizes to data slicing algorithm of structuring model and refactoring algorithm after the confusion. The running time of the data slicing algorithm is as shown in Figure 4.



```
========== OUTPUT ==========
Cloudlet ID    STATUS    Data center ID    VM ID    Time    Start Time    Finish Time
     1         SUCCESS         2             0       800        0           800
     4         SUCCESS         2             0       800        0           800
     7         SUCCESS         2             0       800        0           800
    10          SUCCESS          2             0       800         0           800
     0         SUCCESS         2             5       800        0           800
     3         SUCCESS         2             5       800        0           800
     6         SUCCESS         2             5       800        0           800
     9         SUCCESS         2             5       800        0           800
    13          SUCCESS          2             0       801         800          1601
    16          SUCCESS          2             0       801         800          1601
    19          SUCCESS          2             0       801         800          1601
    22          SUCCESS          2             0       801         800          1601
    12          SUCCESS          2             5       801         800          1601
    15          SUCCESS          2             5       801         800          1601
    18          SUCCESS          2             5       801         800          1601
    21          SUCCESS          2             5       801         800          1601
     2         SUCCESS         2             3      1602        0          1602
     5         SUCCESS         2             3      1602        0          1602
     8         SUCCESS         2             3      1602        0          1602
    11          SUCCESS          2             3       1602         0          1602
    25          SUCCESS          2             0       801        1601          2402
    28          SUCCESS          2             0       801        1601          2402
    31          SUCCESS          2             0       801        1601          2402
    34          SUCCESS          2             0       801        1601          2402
    24          SUCCESS          2             5       801        1601          2402
    27          SUCCESS          2             5       801        1601          2402
    30          SUCCESS          2             5       801        1601          2402
    33          SUCCESS          2             5       801        1601          2402
    14          SUCCESS          2             3       803        1602          2405
    17          SUCCESS          2             3       803        1602          2405
    20          SUCCESS          2             3       803        1602          2405
    23          SUCCESS          2             3       803        1602          2405
```

Figure 4.The running time of slicing algorithm

As shown in Table 1, different types of data privacy compatibility were selected for the experiment and the data with different attributes on the basis are selected. The running costs of slicing algorithm is as shown in Figure 5.

Table 1 Data privacy compatibility

Data type	Data privacy compatibility	Data privacy non-compatibility
Type 1	20%	80%
Type 2	50%	50%
Type 3	80%	20%

The experimental result of data slicing shows: the running costs of slicing algorithm before the data confusion is related to the data compatibility and the quantity of data attributes. The data compatibility is

inversely proportional to non-compatibility. Multiple slice combination are allocated in the slicing algorithm computations allocation while the non-compatibility degree is higher (compatibility degree is lower). So it's running cost is larger than the data non-compatibility experiment.

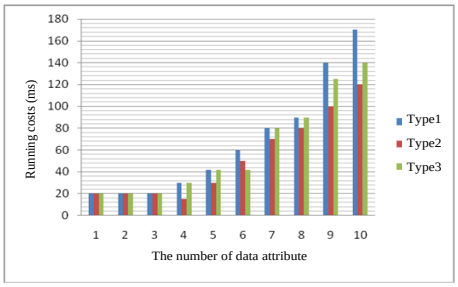


Figure 5 Running costs of slicing algoritm

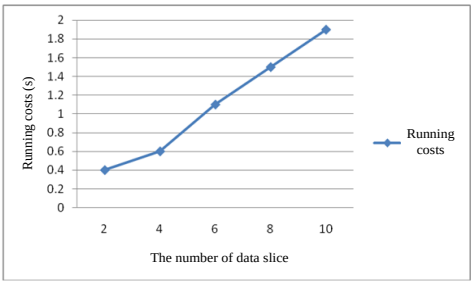


Figure 6 Running costs of reconstruction algorithm

After the data sliced has been storaged, the data can be reproduced by trusted third party confusion strategy and data slicing mark. The running costs of reconstruction algorithm is as shown in Figure 6.The experimental results of the reconstruction algorithm show: obtaining the slicing relationship confusion strategy from the trusted third party to reconstruct the data privacy according to the slice mark. The more slices, the largger running costs.

VI. CONCLUSION

According to the data security problems of multi-tenant application in the cloud storage system, the paper structures the cloud storage model based on the theory of confusion. Data confusion algorithm and data slicing technology are utilized to protect user data by the confusion of the data privacy slice relationship. Experiments shows that the model has more advantages in ensuring safety, reliability and usability of user data in the cloud storage system.

VII. ACKNOWLEDGMENT

This research project is supported by the National Natural Science Foundation of China under Grant No. 61272458.

REFERENCES

[1] Wang H. Proxy provable data possession in public clouds[J]. Services Computing, IEEE Transactions on, 2013, 6(4): 551-559.

[2] Parakh A, Kak S. Space efficient secret sharing for implicit data security [J] . Information Sciences, 2011, 181(2):335-341.

[3] Fugkeaw S. Achieving privacy and security in multi-owner data outsourcing[C]//Digital Information Management (ICDIM), 2012 Seventh International Conference on. IEEE, 2012: 239-244.

[4] Deng Q. The research of cloud computing security mechanism based on Hadoop [D].Nanjing University of Posts and Telecommunications ,2013.

[5] Shi Y. & etc. The research of confusion algorithm[J]. The Journal of Tongji University (Natural Science),2005,06:813-819.

[6] Jiang H.& etc. Code confusion technology research based on control flow[J]. Application Research of Computers, 2013,03:897-899+905.

[7] Parameswaran R, Blough D M. Privacy preserving data obfuscation for inherently clustered data[J]. International Journal of Information and Computer Security, 2008, 2(1): 4-26.

[8] Zhao Y. J.& etc. Evaluation of Code Obfuscating Transformation[J]. Journal of Software,2012,03（3）:700-711.

[9] Itani W, Kayssi A, Chehab A. Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures[C]//Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on. IEEE, 2009: 711-716.

[10] Wu S. Research of software protection based on binary code obfuscation[D].University of Electronic Science and Technology of China,2013.