

Isolated Distributed Architecture for Cloud-Storage Services Based on Structured Strategy

Ying Gao and Jianlin Zheng

College of Computer Science and Engineering, South China University of Technology, China
gaoying@scut.edu.cn, helinz@qq.com

Abstract. Cloud Security threats have greatly hindered the development of Cloud Computing and the promotion of cloud application, so how to protect the data stored in the cloud is not only the core issue of security, but also the biggest challenge of development. As these problems above, a trust-control architecture based on the concept of Cloud Storage Service was creatively proposed. Based on the structured strategy and isolated distributed storage-framework, the architecture of cloud services can eliminate the concerns about cloud security and achieve a high degree of safe, reliable and available cloud storage. It's rather a simple and efficient way proved by experiments.

Keywords: Trust Mechanism, Security, Structure, Storage, Cloud, Service.

1 Introduction

To promote Cloud Computing is very hard, for customers, trust is considered as the key issue of cloud computing. People will trust their money into banks, because banks are state-owned and behind the legal guarantee of the Government, but the security of cloud providers' data centers get no third-party of any credibility in the institutional guarantee [1]. Thus users are afraid of sharing their data in cloud centers. Cloud Security has become the biggest problem encountered in the development of Cloud Computing; therefore, how to provide a trusted mechanism to meet the security needs in the cloud environment is now an urgent problem to be solved.

2 Ideas of Structured Isolated Distributed Cloud-Storage Services

2.1 The Way to Solve Cloud Security

Relative to technology, cloud computing is essentially more of a model or a strategy on the innovation based on existing technologies. Cloud computing has changed the serving way, but hasn't overturned the traditional safe mode [2].

The difference is the change of safety measures, the location of deployment of safety equipment and the main body of the responsibility for security. Similar to the traditional way, the way to solve the issues of cloud computing security is also the combination of these three elements of the strategy, technology and people [3].

2.2 The Concept of Cloud Storage Service

The Cloud storage is a extension of cloud-computing concept in the infrastructure layer, however, it's currently mainly used as an attachment of cloud-computing platform, not as an independent standardized cloud service. In addition, cloud providers have built their own data centers; data just exists in respective centers.

Cloud computing means that the IT infrastructure runs as a service, and the service can be anything from the rental of the original hardware to using third-party APIs. Similarly, storage facilities should also be used as a basic resource of the cloud environment. The future of cloud storage should be cloud-storage services.

As the development of next-generation cloud storage, Storage As a Service is a change to existing way, it is a special form of architecture services. This cloud storage service is also transparent to users. It doesn't refer to a specific device or a cloud provider's cloud storage centers, it means the aggregates consisting of storage devices distributed in different physical geographical or various providers' storage-centers and users can use the services provided by a number of different cloud-storage centers.

The core of cloud storage service is the combination of cloud service agreements with cloud storage system and uses opening service interface standards to achieve the transformation of cloud storage to cloud-storage service. It's the core idea of the architecture proposed as the follow in this article.

2.3 The Solution of Isolated Distributed Architecture for Cloud-Storage Services Based on Structured Strategy

The isolated distributed cloud-storage services based on structured strategy have 3 aspects: structured, decentralized and isolated distributed storage. In essence, it breaks data into a couple of individually meaningless subsets of data basing on the scheme of data, and then use the isolated distributed cloud-storage services belonging to different cloud providers to store the subsets. For the cloud, it provides a secure available storage solution and can be trusted (Shown in Fig. 1).

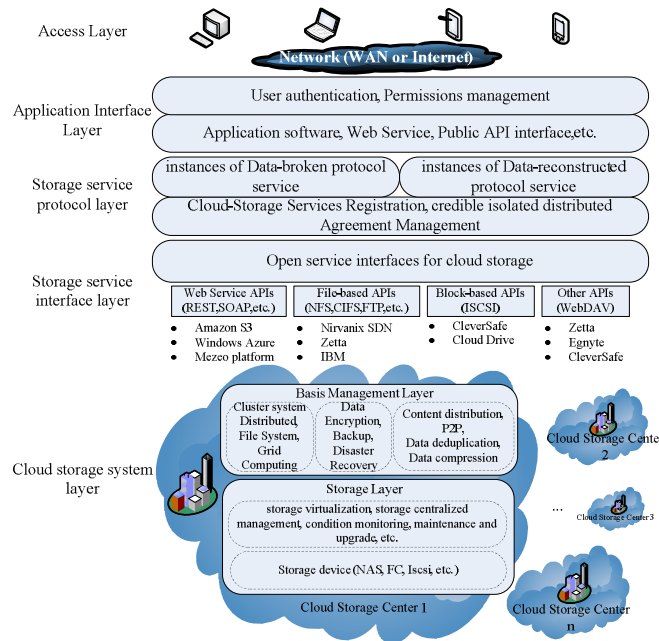


Fig. 1. Isolated distributed Services Architecture based on structured strategy

The architecture is a trust-mechanism. Data is stored in isolated distributed cloud-storage centers from different cloud-providers, in this way, users can recapture the control of data originally belonging to cloud providers, and then have a higher level of centralized control of the distributed control. Combined with tolerable intrusion approaching to data protection(data-broken based on data scheme),such separate meaningless data storage lets users no longer have to worry about the security of cloud storage. In addition, the method of isolated distributed storage disperses the overall risk, with the superposition of each cloud storage service provider's own security the system has a higher security, and the simple structured distribution makes it very efficient. In the architecture, cloud computing applications in accordance with the cloud-storage service agreement (data broken protocols, data reconstruction protocols and isolated distributed agreement) can create service instances in their own cloud computing environment to call the interfaces of cloud storage services for data storage, and also can directly use the third-party services which implement the structured isolated distributed interfaces above.

2.4 The Design of Structured Isolated Distributed Storage

The structured, isolated distributed storage-service layer involves three main service instances: data-broken, data-reconstruction and isolated distributed storage services.

The output of data-broken service instances is data segments based on the logical structure of data. Data submitted by cloud applications is divided into two categories: strong structured data and non-logical data. For the structured data, it's firstly segmented based on the structure of the data, secondly, each data segment continues to be broken and then forms several subsets which have the same scheme; For the second data, it will firstly be segmented based on a default rule-format.

After data subsets are formed, isolated distributed storage services will receive a notification. By its access to the user authentication and available cloud storage services provided by different cloud organizations, it forms the mapping of multi-subsets to cloud storage multi-services, then the mapping and all the subsets are stored in different isolated distributed cloud-storage centers. Finally, the stored path of mapping is stored in the cloud service registry and feedback results will be sent to users (Data write and modify operations both can get a result of feedback). (Fig. 2)

For the instance of data-reconstruction service, when the request of data-reading arrives, it gets the stored path of mapping from the cloud service registry, then requests required data pieces from the isolated distributed subsets according to the structured mapping and restore the data based on the structure. If the request is data-updating or data-deleting, the instance will proceed with corresponding operation according to the original mapping, and then update the mapping after deleting (it also mostly exists in the process of data-updating when the affected pieces count more) or updating the affected data pieces.

The isolated distributed storage-services run based on the open cloud-storage service interfaces in the lower layer. With the logical models for the Family Key-Value, Column, Document, the Graph and other data, this interface layer includes: REST, Thrift, and the Map Reduce, GET / PUT, language-specific API, the SQL subset.

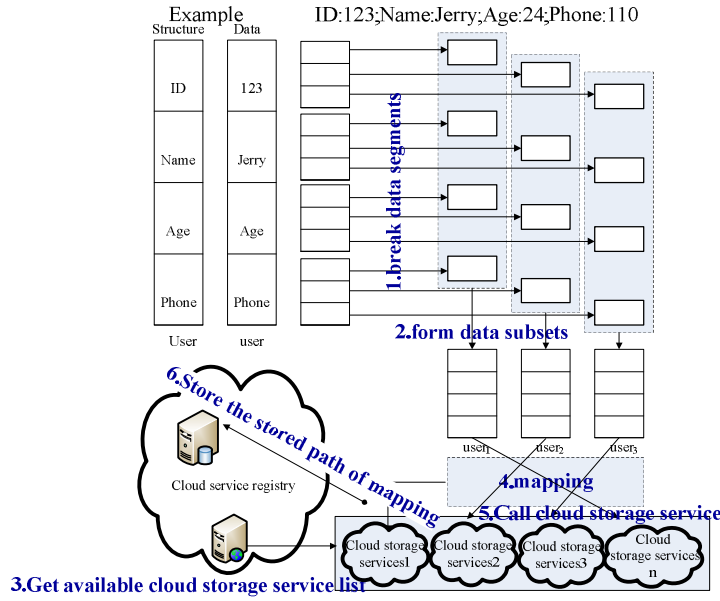


Fig. 2. The basic process for structured isolated distributed storage

2.5 The Method of Data Separation

As early as before the coming of cloud computing, Data separation and reconstruction in a distributed system have made great progress, currently, the most common and effective methods are rooted in information dispersal algorithms (IDA or RS), but in the cloud-storage services architecture proposed in this article, the separation and reconstruction of data are based on the information with its own schema. This simple and efficient data dispersion algorithm has a lower storage, computational and bandwidth cost than IDA and RS.

2.5.1 The Method of Data Separation

The basic idea of information dispersal algorithms based on redundant coding techniques is that a length L of the file F is split into n slices F_i ($1 \leq i \leq n$) in such a way that the file can be reconstructed by any m slices. In space, we can see that the sum of the n -slices F_i is as n/m times as original file, n/m is greater than or close to 1.

IDA equally divides the data volume D required to be distributed into fragments (size m). Assuming the size N of the volume is $|D|$, then the number of fragments N' is N/m . The data is encoded as n piece of data using stage independent $n*m$ vector matrix multiplication, and its size is N' . The only difference between RS and IDA is that RS uses stage independent $(n-m)*m$ Vander monde matrix to calculate the checksum data for each piece of data, then constitute the n fragments with the original data in accordance with the data left-justified way similar to RAID 6 parity. In the case of the same n , m and fault fragment data numbers, the time cost of IDA and RS is basically the same, but with the increase of n and m , the operating efficiency of the RS method will be slightly higher than the IDA [4].

2.5.2 The Information Dispersal Algorithm Based on Schema

The data dispersion algorithm based on the structure is rooted in the duality of information: form and content. The data of no schema does not exist, as long as the data exists in systems, whether it is stored or applied, it has its own schema. Data needs to be resolved and demonstrated its qualified sense, this is the content of data; how to construct or parse the data is the form of data that schema reflects. Thus, a complete and effective information object is a unity of form and content.

In the process of application, data all has its own form of organization (referred to here as a schema), the dispersion algorithm based on the structure of data is in fact decentralized using the schema that comes with the data, such as to a length L of information I , using the data field, table or file as the basic particle size, the form and content of the data are stripped out and data is dispersed into n data subsets S_i ($1 \leq i \leq n$) with the same schema, L_i corresponds to the length of the S_i ($1 \leq i \leq n$), where $\sum_{i=1}^n L_i = L$.

In space, the total length of data subsets in the algorithm based on schema equals to the original file's and the required storage space is less than the information dispersal algorithms' (IDA, RS); In time, because it is based on the data schema to simply segment and restructure with the integration of the structuring stage and the parsing stage about data, it is very efficient. It doesn't require coding with vector matrix multiplication, so the operating efficiency is better than the information dispersal algorithms. In addition, to avoid unnecessary duplication, disaster recovery and error correction processes are put on the cloud-storage services in the lower layer provided by various cloud providers, so it can highly reduce the consumption of time and space.

3 System Analysis and Verification

The architecture is different from the traditional encryption methods on ensuring information security; it uses the cryptography mechanisms in the organizational structure of information making the information content and form of separation, combined with the isolated distributed cloud-storage service solution to establish the Information Trust-System to ensure the reliability, availability and privacy of information.

Application data in the program after structured broken becomes separate meaningless subsets, and all the subsets are isolatedly distributedly stored in the cloud-storage mediums provided by different cloud providers with the pulled out schema and mapping file. Separate one or several data sub-bodies are not valid, only the schema mapping with all the required subsets can reconstruct the original message-body. Such content and form of isolated sharing, mutual checks and balances of the various cloud providers ensure the security of internal and external information, and users have no need to worry about the invasion of the cloud data from external hackers or internal staffs or users of cloud providers, theft and abuse etc. such a series of threats. To breakthrough the defense line of the architecture, at least need to pass the following checkpoints: user security authentication, the key of schema mapping

and cloud-storage service safety certification of various cloud providers. Therefore, this line of defense is very solid, and reliable cloud providers also protect the authority of the security of the system. The solution was through system simulation and verification in the LINUX environment, and the information dispersal algorithms were compared.

In contrasting schema-based dispersal algorithm with information dispersal algorithms, the inputs of the experiments are: data files with User structure (Fig. 2), the length is 1M, $n = [5, 15]$; the broken sub-unit of schema-based dispersal algorithm is file, and its basic granularity is table; $m=n-1$ in information dispersal algorithms. Experiments were repeated 10 times averaged to obtain the time-costs of dispersion and reconstruction (Fig. 3). Figures show that the efficiency of the schema-based dispersal algorithm is better than the information dispersal algorithms', and there is a linear relationship between its efficiency and the size of the original data file.

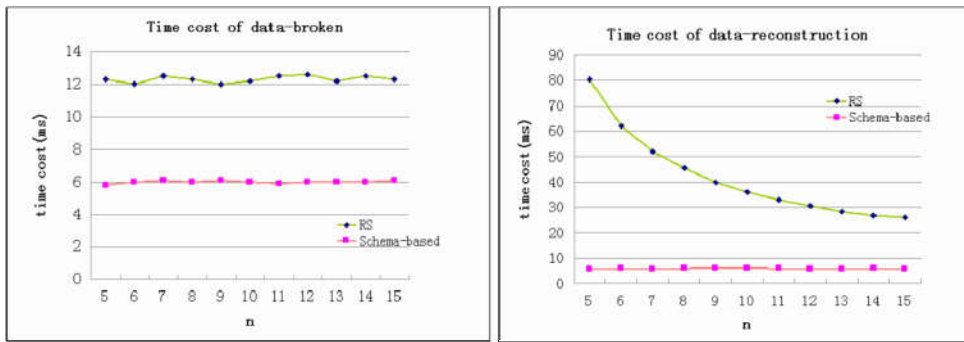


Fig. 3. Time cost of data-broken and data-reconstruction

4 Conclusion

Compared with distributed storage, this article creatively proposes a trust-control architecture based on the concept of cloud storage service. Users no longer have to worry about the security of cloud storage; it removes the biggest obstacle faced in the promotion of cloud, and has a great significance to the development of cloud.

Acknowledgement. The research is supported by:

- Science and Technology Planning Project of Guangdong Province (2009B010800047).
- Ministry of Education University-Industry-Research Project of Guangdong Province (2009B080702037).
- Ministry of Education University-Industry-Research Project of Guangdong Province (2010B090400535).
- The Fundamental Research Funds for the Central Universities (20112M0068).

References

- [1] Yi, Y., Chun, L.Y.: Security Issues on the cloud computing environment. *Computer Knowledge and Technology* (5) (2009)
- [2] From cloud computing to the cloud security. *Information Systems Engineering* (2009)
- [3] Yi, C.S.: On the cloud computing security issues. *Network Security Technology and Application* (10), 22–23 (2009)
- [4] Rabin, M.O.: Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance. *Journal of the Association for Computing Machinery* (36), 2–4 (1989)