AIS | American Institute of Science

Research on Storage Techniques in Cloud Computing

Dapeng Song*, Lei Peng

College of Information Engineering, Taishan Medical University, Taian, Shandong, China

Abstract

The cloud computing technology is one of the most important orientations for nowadays development of informatization, as well as one of the most significant applications in the field of informatization development for each and every industry. Therefore, the research into the cloud computing technology is of great significance. Starting from the definition of the cloud computing technology, this paper explores the current situation of the cloud computing technology and its related services, as well as the core techniques and constraints of cloud computing for the purpose of improving insiders' understanding of the cloud computing technology, accelerating the development of the core techniques of cloud computing, strengthening the unification of the cloud computing standard and applying the cloud computing platforms to all kinds of areas in this society.

Keywords

Cloud Computing Technology, Cloud Computing Services, Mining of Mass Data, Distributed Computing

1. The Definition of the Cloud Computing Technology

As for the definition of cloud computing, no unanimous conclusion has been drawn yet. In the narrow sense, cloud computing refers to that the manufactures build data centers or supercomputers via distributed computing and virtualization technology and provide technology developers or enterprises with the services including data storage, analysis, scientific computing, etc. free of charge or through renting as per their requirement. In a broader sense, cloud computing refers to that the manufactures provide a variety of services, such as Software as a Service (SaaS), hardware lease, data storage, computational analysis, etc. for different customers through building web server clusters. To be colloquially, cloud computing is to distribute computing tasks in a resource pool made up of a large number of computers in order to allow a variety of application systems to acquire computing power, storage space and various hardware and software services according to their needs [1].

Seen from different aspects, cloud computing will have different interpretations. For instance, in terms of resources, cloud can be seen as the infinite software and hardware resources that are expandable and accessible at all times and places. In the eyes of end-users, cloud is the software and hardware resources that they can use according to their needs by means of network transmission and various smart devices. And speaking of the services it provides, the cloud computing technology can be regarded as the public utility of the computing system, making the utilization of the software and hardware resources similar to that of water and electricity.

2. The Generation of the Cloud Computing Technology

The cloud computing technology is inspired by the "power plant" mode, namely, the power plant's mode of centralized power supply. We use electricity every day, but we don't have to buy a generator to generate electricity for ourselves. In a similar way, we drink water every day, but we don't have to

* Corresponding author

E-mail address: dpsong@tsmc.edu.cn (Dapeng Song)

dig a well on our own. Instead, water is supplied to us by water supply companies. Actually, the cloud computing technology is a concept in terms of public utilities, which puts all the main computing resources together into a public cloud computing center and complies with the public protocols. Similar to the voltage of electricity (220v/110v) and SS7 (Signaling System Number 7) of communications, both enterprises and individuals could use it easily and conveniently [2]. Because of the great advantages in the scale of such a mode, the operating cost could be dramatically decreased. Moreover, as it is mainly operated by large telecommunications companies in our own country, it is easy for it to fully win the trust of the end-users. The cloud computing technology has changed people's mindset of utilization from directly purchasing software and hardware to purchasing services as per their needs, which not only reduces the expenses of maintaining their software and hardware resources, but also makes it possible for them to customize their own cloud services based on their actual needs.

3. The Current Situation of the Cloud Computing Technology

The cloud computing technology has enjoyed great popularity among all IT tycoons since its advent. Amazon, Google, IBM and other enterprises have launched their own cloud plans and cloud products respectively. For instance, Amazon uses EC2 (Elastic Computing Cloud) and S3 (Simple Storage Service) to provide enterprises with computing and storage services. Google has built more than one million servers that provide powerful search engine services (SES). IBM has launched the "Blue Cloud" computing platform that brings customers the cloud computing platform that could be used directly once they pay for it. And Microsoft has also introduced the "Windows Azure" operating system.

4. The Establishment of the Data Storage System in Cloud Computing

In practice, the data storage systems in cloud computing mainly include data centers, cloud service interfaces, cloud users, service protocols, etc., among which data centers are primarily composed of storage management, distributed file systems, storage devices, etc. Meanwhile, they are also the foundation for realizing data storage in cloud computing. Generally, there is no special requirement for the types of cloud storage devices, which could be either a personal computer (PC) or a special storage device used in the development process of the enterprise. It could even be the combination of the two types mentioned above. The physical

storage resources of file system management in distributed file systems are connected through network nodes. And the design of this system is mainly based on the client and the server modes. [3,4]

4.1. Server Architecture

Before the establishment of the cloud storage system, we should firstly build the cloud storage server architecture. As an important part of IAAS (Infrastructure as a Service) of cloud computing, it could be achieved through various architectural techniques. Among them, NAS (Network Attached Storage) and SAN (Storage Area Network) are the most common ones. NAS is the file storage system of distributed architectural systems. It is a kind of loosely coupled cluster that could effectively meet the cloud storage-centered environmental requirements in practice. Moreover, all nodes in this cluster are a relatively independent entity. In this system, the smallest unit is the file. When a file is saved in the cluster, its data or information will be efficiently saved in a specific node in the cluster. Although there are many copies of the file which might be relocated to another node as well, thereby resulting in a redundancy, the file's single instance provided by this specific node is unique. Figure 1 shows the architecture of NAS storage as follows.
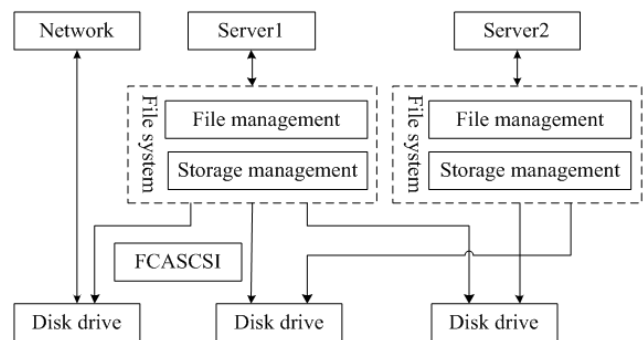

Figure 1. Architecture of NAS Storage.

It could be seen from the practice that SAN is a kind of block storage system of the cluster architecture. It is a tightly coupled cluster system. When a file request is generated, every node will access to a different data block of the file and respond to the user's request in time. The more requests from the users, the more nodes will be available for SAN to respond to the request, and the greater its performance will be. Practices show that this storage architecture is more suitable for cloud servers. Furthermore, as it has excellent expandability and high data transmission efficiency, the requirements for primary storage could be well met if works together with P2P and de-duplication technology.

4.2. Main Services

4.2.1. Cloud Services

Services of cloud computing mainly include SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service).

4.2.2. Infrastructure as a Service (IaaS)

IaaS provides hardware infrastructure services. It integrates memory, I/O devices, storage and computing power into a virtual pool of resources to provide hardware-related services. For instance, Amazon's virtual server AWS (Amazon Web Services) is exactly a kind of IaaS. The advantages of IaaS are that its users do not need to think about the cost or the maintenance of the hardware, which reduces the cost of using the hardware. All that the users need to do is to rent the hardware infrastructure provided by a third party as per their practical needs. [5]

4.2.3. Platform as a Service (PaaS)

The service providers of PaaS provide development platforms, server platforms and other resources for its users to allow them to develop application platforms according to their needs. PaaS, actually, provides its users with the resources on the Internet in the form of programmable interfaces, which completely subverts the concept of building, deploying and operating basic platforms by the users themselves. In this way, users could obtain more programmable resources, which will bring developers more convenience, increase development efficiency and save development costs.

4.2.4. Software as a Service (SaaS)

SaaS refers to the application software services or the custom-made application software services provided for the users by the service providers of SaaS. To use the software, users only need to access to the Internet. They don't have to download or install corresponding software, and they can use it at all times and places through a browser that will access to the remote server. For instance, Zoho office, Google Apps, etc. are SaaS. It reduces users' maintenance expenses in the software, especially the software that has a relatively high demand for hardware, and dispels users' misgivings about the hardware. Salesforce.com is currently the most famous company that provides SaaS services.

4.3. Core Techniques

Cloud computing is the outcome of the development and integration between the network technology and the traditional computer technology including grid computing, distributed computing, parallel computing, utility computing, network storage, virtualization, load balancing, etc. Its main techniques are as follows:

4.3.1. Mining of Mass Data's Storage

Cloud storage connects all kinds of storage devices on the Internet through application software for collaborative work to provide data storage and access services. A large number of data are stored on an assemblage made up of different servers and storage devices and managed in the way of distributed data storage. To be specifically, Google's BT (BigTable) data management technology and the open source data management module HBase developed by Hadoop [6] fall within this scope.

4.3.2. Distributed Computing

The cloud computing platform includes a large number of servers. In the server cluster, there is a master server and a lot of slave servers. The coordination between the master and the slave servers allows users to access to all the resources in the server cluster via logging on to only one server. Moreover, load balancing can also be achieved, making it possible for a large number of users to access to and use the resources on the cloud platforms at a high speed at all times and places.

4.3.3. Virtualization Technology

The virtualization technology that includes storage virtualization, computing virtualization and network virtualization could achieve the phase separation between underlying hardware resources and software [7]. The resources are managed through aggregating and splitting the virtual resources, so that end-users are able to access to the resources in a way similar to operating a local machine while logging on to the cloud technology platform. In addition, it could also ensure the high speed, real time and reliability of resources access.

5. Design of Secure Cloud Storage Systems

5.1. Solutions to Secure Cloud Storage Systems

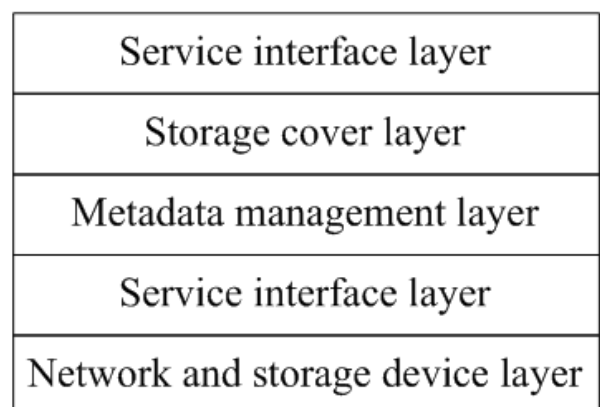| Service interface layer |
| --- |
| Storage cover layer |
| Metadata management layer |
| Service interface layer |
| Network and storage device layer |

Figure 2. A classical cloud storage model.

Composed of large-scale network clusters, distributed file systems and other storage middleware, cloud storage systems provide users with cloud storage services. Typical cloud storage architecture includes storage resource pools, distributed file systems, service level agreements, service interfaces, etc. A typical cloud storage model [8] is shown in Figure 2.

In regard to the solutions to cloud storage security, the first priority is to ensure the confidentiality of the data stored in the cloud. Hence, both the credibility and the incredibility of the cloud should be taken into consideration. In order to allow the system to have a better efficiency, the traditional block cipher cryptosystem and the new cryptosystem should be combined when designing the system. Additionally, for the communication between a trusted third party and users, a secure solution should also be considered and the cloud should protect the data integrity as well. Moreover, all these operations should be transparent to users and able to provide audit records for the users' reference. Therefore, the applications of block cipher-based encryption, attribute-based encryption (ABE), digital signatures, integrity verification [9] and other basic techniques must be focused on when designing the solutions for preserving the security of the cloud storage system. Through analyzing the features of these techniques and of the cloud storage technology, we design a complete set of rationally secure cloud storage system which could provide a more robust security and privacy protection of existing cloud storage systems without significantly reducing the storage efficiency.

5.2. Block Cipher-Based Encryption of Cloud Storage

For users, the data could be classified into general data and private data. At present, users tend to store and protect only their general data in the cloud, and the cloud protects users' data generally in the way of encrypted storage. In such a case, in consideration of the possibility that the encrypted data stored in the cloud might be stolen by attackers, the cloud should take proper block cipher algorithms.

In terms of protecting the general data, the system designed in this paper can be divided into two cases according to the willingness of the users to share the data or not. Only authenticated users can get the download address once the data are set to be prohibited from sharing. And meanwhile, the cloud will use users' authentication information to decrypt users' data first. Without the authentication information, those unauthenticated users could only get the ciphertext even they get the download address. If the data are set to be shareable, then all users could get decrypted data from the cloud, as shown in Figure 3. In cloud storage, it is inevitable to have a lot of repetitive data, as the amount of users' data is really

huge. Therefore, general block cipher often encrypts the same plaintext block data into the same ciphertext block data. In other words, if users' general data are encrypted using the general block cipher, attackers may attack the data through statistical analysis and linear analysis when the cloud is credible. Hence, firstly, in consideration of the influence on the system, the efficiency of the algorithm should be ensured not to be significantly lower than the existing general encryption algorithm. And secondly, in consideration of the possible means of attack, the algorithm should be ensured to be able to encrypt the same plaintext data into completely different ciphertext data. The algorithm put forward based on the two points mentioned above can meet the requirements of the secure cloud storage system.
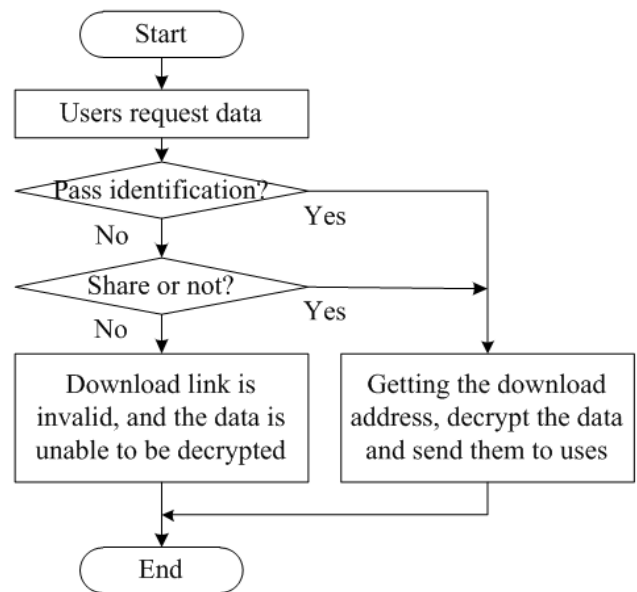

Figure 3. Download processes of users' general data.

5.3. Attribute-Based Encryption (ABE) of Cloud Storage

It is the users' hope that the cloud should ensure absolute security and confidentiality of their private data and that the cloud should provide a safer and more confidential way of protection. Users always think that the cloud is not credible. The data should be encrypted not only during the transmission, but also in the cloud by a secret key provided by a trusted third party. Taking into account all of these factors, it is safer to encrypt firstly on the client side using a secrete key generated by a third party, and then save the ciphertext into the cloud. In such a case, it is unnecessary to use the secure protocol that will consume the network and computing resources on the server side, but it is ensured that the users' ciphertext will not be decrypted by the cloud. In general, a trusted third-party could be invited to manage the general secret keys, i.e., a third party is responsible for generating and providing keys for each user. The advantages of doing in this way is that only the third

party and the user know the key, while the cloud and attackers know nothing about it. However, there might be such a situation that the file uploaded by the user are shared by a group instead of the user he himself.

Private files are generally encrypted by multilevel keys. In the beginning of the data is some basic information about the data, following that is the ciphertext encrypted by a trusted third party using an attribute-based secrete key, and the last part is the data encrypted by the secrete key. The secrete key is generated randomly. The encryption of the data adopts the traditional fast block cipher mode. Adopting this approach could not only decrease the influence on the efficiency caused by attribute-based encryption of the entire file, but also ensure that the files are protected by ABE. Moreover, it could ensure that the beginning of the data could be correspondingly updated when the attribute is changed, making the file be shared by multiple specified legitimate users without the need to re-upload.

In addition, we have briefly introduced ABE earlier in this paper. This method can, firstly, achieve one-to-many transmission, divide users based on the attribute and determine that the higher level could download and decrypt the files of the lower level through division the role of users at different levels. Secondly, each user's key is unique, and each user could apply for a new attribute-based secrete key from the third party according to the changes in the attributes to adapt to their new roles. In such cases, the former attribute key encrypted files could still be decrypted using the former rules while the updated attribute key encrypted files could only be decrypted by the updated rules.

6. Constraints on the Development of the Cloud Computing Technology

6.1. Protection of Data Security and Privacy

The greatest hidden danger that the cloud computing technology is confronted with while providing development platforms is the protection for the security and the privacy of the data. A large amount of public information resources are distributed to multiple servers, which increases the risks of cloud computing platforms. Therefore, the users and service providers of cloud computing have to face the ordeal which is also the key factors affecting the security of cloud computing platforms. To solve this problem, we need to adopt advanced information security technology to ensure the security of information sharing. In addition, our government should also establish and improve relevant laws and regulations, so that those people who disclose the privacy will be punished accordingly, and even face legal sanctions.

6.2. Inconsistent Standards of Cloud Computing Platforms

Cloud computing is still under development, and no unanimous or unified technology standard has been drawn yet within relative fields. Without a unified standard, it is difficult to realize the large-scale expansion and applications of cloud computing platforms. On the one hand, cloud computing doesn't establish or form a unanimous norm or standard, which is not conductive to its promotion or obtaining users' recognition [10]. On the other hand, no unified technology standard or data norm has been formed within relative fields. In the information management system within the industry, different software companies often set different standards and norms under different circumstances. It greatly restricts the expansion and improvement of cloud computing platforms and makes it impossible for the software developed by different companies to aggregate into a cloud computing platform. As a result, it is difficult to form urbanized cloud server clusters, which largely blocks the progress of cloud computing platforms in terms of standardization and industrialization.

References

[1]   Li Qiao and Zheng Xiao, *Research Survey of Cloud Computing*, [J]. Computer Science, 2011, 38(4):32-37;

[2]   Lu Gaohui, *Research on Personal Learning Environment Construction Based on Cloud Computing*, [D], Henan: Henan Normal University, 2013:1-57;

[3]   Chu C K, Chow S S M, Tzeng W G, et al. Key-aggregate cryptosystem for scalable data sharing in cloud storage [J]. Parallel and Distributed Systems, IEEE Transactions on, 2014, 25(2): 468-477.

[4]   Ni J, Yu Y, Mu Y, et al. On the security of an efficient dynamic auditing protocol in cloud storage [J]. Parallel and Distributed Systems, IEEE Transactions on, 2014, 25(10): 2760-2761.

[5]   Fu Y, Jiang H, Xiao N, et al. Application-aware local-global source deduplication for cloud backup services of personal storage [J]. Parallel and Distributed Systems, IEEE Transactions on, 2014, 25(5): 1155-1165.

[6]   Liu Zhengwei, Wen Zhongling and Zhang Haitao, *Cloud Computing and Cloud Data Management Technology*, [J]. Journal of Computer Research and Development, 2012, (49):26-31;

[7]   Chen Quan and Deng Qianni, *Cloud Computing and Its Key Techniques*, [J]. Journal of Computer Applications, 2009, 29(9):2562-2568;

[8]   W. Zeng, Y. Zhao, K. Ou, et al. Research on Cloud storage architecture and key technologies. ICIS, 2009. 1044~1048

[9]   K. D. Bowers, A. Juels, A. Oprea. HAIL: A High-Availability and Integrity Layer for Cloud Storage. Cryptology ePrint Archive Report, 2008. 187~198

[10] Zhang Feng, *On Modes of Cloud Computing Services*, [J].Information Technology and Informatization, 2012, (2):81-83;