 JIOT, vol.1, no.1, pp.33-39, 2019

Ensuring Readability of Electronic Records Based on Virtualization Technology in Cloud Storage

Qirun Wang[1, 2], Fujian Zhu[3, 4], Yan Leng[3, 4], Yongjun Ren[3, 4, *] and Jinyue Xia[5]

Abstract: With the rapid development of E-commerce and E-government, there are so many electronic records have been produced. The increasing number of electronic records brings about storage difficulties, the traditional electronic records center is difficult to cope with the current fast growth requirements of electronic records storage and management. Therefore, it is imperative to use cloud storage technology to build electronic record centers. However, electronic records also have weaknesses in the cloud storage environment, and one of them is that once electronic record owners or managers lose physical control of them, the electronic records are more likely to be tampered with and destroyed. So, the paper builds a reliable electronic records preservation system based on coding theory. It can effectively guarantee the reliability of record storage when the electronic record is damaged, and the original electronic record can be restored by redundant coding, thus ensuring the reliable storage of electronic records.

Keywords: Electronic record, cloud storage, preservation mechanism, coding theory.

1 Introduction

Along with the continuous development of information technology, office automation has popularized rapidly at all levels of various state organs, social organizations, enterprises and public institutions from all over the world. Under such a large environment, electronic records have sprung up everywhere in the world of various types of workplaces and people's daily lives. The standard of China "electronic record filing and management norms" defines electronic records as follows. Electronic records are the record of social activities generated in computers and networks, which is recorded in digital form on the carriers, such as tapes, disks, CDs, etc., depending on the computer system to read and process. It is a kind of records which can be transmitted on the communication network. Other national electronic record management departments and

[1] School of engineering and technology, University of Hertfordshire, Hertford, UK.

[2] School of Computer Information and Engineering, Changzhou Institute of Technology, Changzhou, China

[3] Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing, China.

[4] School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, China.

[5] International Business Machines Corporation (IBM), New York, USA.

[*] Corresponding Author: Yongjun Ren. Email: renyj100@126.com.

international archives academic institutions define many versions of concept of electronic records. They all revealed that the electronic records have two basic attributes: the property of the binary record and the attribute of traditional documents [Tu and Zhang (2016); Liu and Li (2017)].

The readability of electronic records is the basis of its existence and preservation. If the electronic records could not be read out successfully, then the effective information in it would become invalid. No matter how value it is originally, at the time it no longer has the meaning of preservation because it cannot be read. The readability of electronic records refers to regard protecting the readability of its contents as a target, and whether it can be read and understood by human being as a judgment standard. Electronic records can achieve the above objectives and standards through storage, transmission, compression, media conversion and migration and other treatment.

The rest of the paper is organized as follows. Section 2 introduces the related work. In Section 3, the problem of the preservation for network electronic record in urban construction is stated. And Section 4 introduces the technology of broadcast-storage network. Section 5 gives our preservation scheme based on broadcast-storage network. Finally, Section 6 concludes this paper.

2 Related work

With the continuous development of information technology, office automation is rapidly popularized in various state organs, social organizations, enterprises and small business units at all levels in the world [Ren, Shen, Liu et al. (2016)]. The increasingly obvious convergence of Internet in daily work and life, the maturity of computer aided design (CAD) and computer aided manufacturing (CAM) in the field of scientific research and industry. Electronic data exchange (EDI) and the application of electronic commerce (EC) in domestic and international trade make the electronic work environment more and more common. In this environment, electronic records have sprung up in all kinds of workplaces around the world and in people’s daily lives. The interpretation of electronic records in the national standard of China “Standard for filing and Management of Electronic records” (GB/T18894-2002) is in accordance with the numerous version definitions of the concept of electronic records by national electronic records management authorities and international archival academic institutions, materially. In them, the two basic attributes of electronic records are revealed: the attributes of binary records and the attributes of the traditional records [Xiao and Wu (2015); Wang (2013)].

When understanding the record from the angle of the electronic records, besides the function of the record, it should also make clear the form of the record. The electronic records committee of the International Archives Council (ICA) set up the guideline for the management of electronic records in 1996, pointing out that the three elements of the records are content, background and structure. The readability of electronic records refers to the basis for the existence and preservation of electronic records. If the electronic records cannot be read out smoothly, the effective information in the electronic record will become invalid information, and what is worth at that time will not be saved because it cannot read at the moment [Zhang and Zang (2013); Han, Li and Bi (2012)]. The readability of electronic records refers to the purpose of keeping the readability of its content as its target. The discriminant standard is human readable and understandable output. The management of electronic records general the following processing: storage, transmission, compression, encryption, media conversion and migration [Shen, Zhou, He et al. (2017)].

Safeguards for the readability of electronic records must always run through all aspects of electronic records management. When filing electronic records, we must specify the storage format adopted in the electronic documents according to their own types and characteristics, the hardware and software environment depending on the production of electronic records, related data and explanatory materials, parameter manuals, etc. To ensure the readability of electronic records, the electronic record information can be effectively identified. The storage media is safe and reliable. And the storage system compatibility and other parties are no loss, no damage, ineffective and trouble-free. The ultimate goal of the electronic record readability guarantee work is to make the electronic records after archiving and saving in the long-term software and hardware environment can be realized smoothly in a way that people can read and understand. And maintain the readability of its content.

3 Problem description

There are many reasons about the un-readability of electronic records. The main factors that affect the readability of electronic records are as follows.

(1) Since electronic records are generated and processed until archived, all administrative activities need to rely on the computer system for their operation. If the electronic records generated by incompatible computer systems and applications of different versions need to be exchanged for use, handlers are bound to encounter obstacle to use. Due to the system dependencies of electronic records, ones have been saved can only be effectively read in a particular system environment. However, the electronic records cannot usually be effectively read because of system incompatibility. The operating system that runs the software and hardware to generate electronic record, are not compatible with the original system. Thus, we need to save the old system or perform a series of transitions and migrations to accommodate the new system in order to ensure readability and manageability of electronic records.

(2) The reasons that affect the readability of electronic records may also be related to the generation, management and final storage of them. Computer software and hardware have been upgraded and updated. The new electronic records management system and computer access environment cannot effectively read the electronic records generated under the original system environment because they cannot accept the record formats that have been eliminated. In addition, there are many reasons that lead to changes of the electronic record information, because the information in electronic records is fluid. When information can exist after separating from a particular carrier, the storage carrier has no bond. The relative independence of electronic record information in computer systems makes it easy for users to add, delete and modify electronic records. The dynamic data in electronic records are continuously updated and modified.

Suffering from virus infection, the storage carrier is damaged, the signal is attenuated due to interference, and the encrypted electronic records may affect the smooth reading of the electronic records, because there is no corresponding password or other reasons. In addition, because of the instability of the performance of the electronic record storage carrier, the information in the electronic records may also change unexpectedly. Therefore, the data on the original storage carrier should be regularly backed up so that the data of the stored electronic records will be kept updated from time to time to prevent the readability of the electronic records from being damaged.

4 Preservation of electronic records based on MBRC

4.1 Server virtualization

The service model of cloud computing contains the following components: Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS) [Xiong and Shi (2018)]. The cloud conceptual model is shown in Fig. 1. Server virtualization technology is the core technology of IaaS. It virtualizes a physical server into several independent virtual servers and gives full play to the server's hardware performance. Server virtualization technology transforms physical resources such as CPU, memory and I/O devices into logical resources that can be managed uniformly and provides each virtual server with abstract resources that can support its operation.

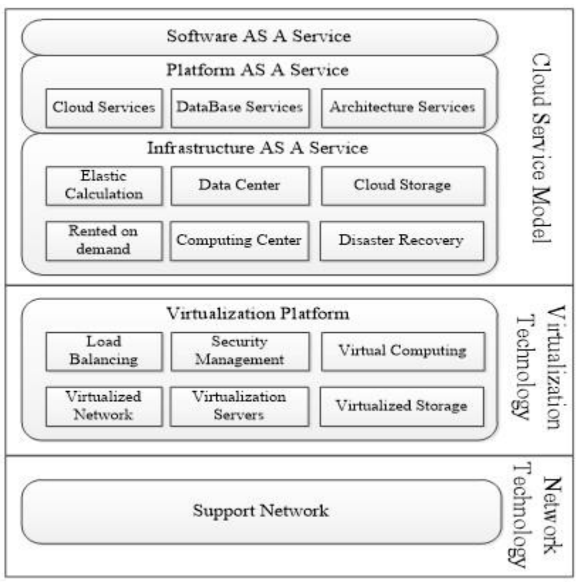

Figure 1: Cloud computing model

In the face of complex and volatile business environment, a fast and flexible application system solution is essential. This application system often includes many different systems, so building such a large application system on a physical server entails expensive physical server costs. On the other hand, these physical servers must deliver high-performance operational performance in the face of peak workloads in operational efforts. However, this can lead to server utilization down sets and waste of resources during other general business hours. In order to ensure the high availability of the application system, it is necessary to set up a backup system that allows the system to continue operating in the event of an error. However, building and maintaining this standby system, which is rarely used in traditional situations, also requires additional costs and time. The emergence of server virtualization technology can be a good solution to these problems. When building a large number of virtual machines on a small number of physical servers, the installation costs and time required are significantly reduced compared to establishing a large number of physical servers.

In addition, server virtualization technology can increase the utilization of physical servers from 5%～20％ to 85%～90%. In general, server virtualization reduces total cost of ownership and quickly changes system configuration as system usage changes allocate resources needed for the system in a timely fashion.

4.2 Storage virtualization

Storage virtualization is a very important part of the cloud storage system. It lies in the core layer of the cloud storage system, that is, in the basic management, to achieve a unified management of a large number of storage devices. In cloud storage, users are not concerned with the specific storage location of the data but are concerned with how to use the stored data, how the data will be presented to the cloud operating system terminal, whether the data storage is secure, and so on. The providers of cloud storage services need to fully consider how different data will be stored and managed uniformly.

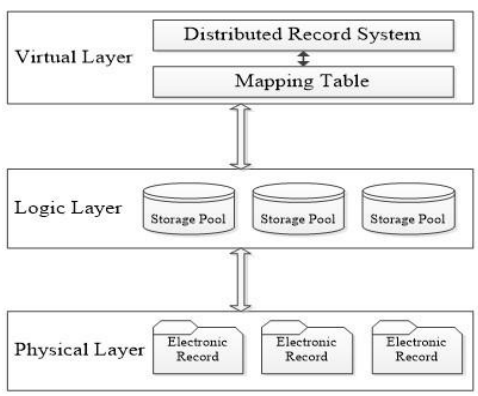

Figure 2: Architecture of storage virtualization

Storage capacity is one of the important indicators to measure cloud storage system. Compared with simply adding physical storage to expand storage capacity and increase storage utilization, increasing storage utilization is the best way to expand capacity based on reducing purchasing costs. Using storage virtualization technology in the cloud storage system, on the one hand, in order to eliminate the physical difference of storage devices produced by a large number of different manufacturers in the cloud storage system. On the other hand, cloud storage system makes the storage space scalability, which enables the dynamic expansion of storage capacity and the dynamic allocation of user storage space.

For each distributed record system, there is a mapping table between the virtual layer and the logical layer. The mapping table includes the information that maps the relationship between the virtual volume and the logical volume. Through the establishment and updating of the mapping table, virtual volume capacity will be expended dynamically to meet each dynamic demand of the storage capacity. For storage batteries; there is also a mapping table between the logical and physical layers. This mapping table describes the mapping between logical volumes and physical volumes to confirm the true address of the logical volume in the physical layer. In the meantime, in order to manage the mapping table better, a mapping table manager is set up between the virtual layer and the logical layer, between the logical layer and the physical layer so as to update the mapping table quickly, effectively and safely. This kind of storage virtualization structure can effectively realize the unified management and dynamic allocation of storage resources, which greatly improves the utilization of storage devices. The mapping manager mechanism offers more security and convenience to manage virtual storage space, physical storage space and logical storage space.

4.3 Application virtualization

In general, each application depends on the operating system in which it resides, such as CPU, memory allocations, device drivers and so on. Different applications running on the same operating system often contain a large amount of common system information that can lead to conflicts between applications. For example, one application needs a specific version of a dynamic link library, while another application needs the same dynamic signature library that is another version. When both applications are running at the same time, this will lead to a disaster. Enterprises usually avoid the problem by installing a large number of applications for testing and deploying available applications. Although it is effective, but it cannot focus on the application to update and maintain, and the price is huge, which greatly increases the difficulty of management.

Application of virtualization technology can be a good solution to the above problems, so that applications in the cloud can show great freedom and independence. Application virtualization is the foundation of SaaS, which provides a virtual layer, which is a virtualization platform on which all applications can run. It can provide all application related registry information, configuration records, etc. At the same time the application is relocated to a virtual location, packaged with the operating environment associated only with it, forming a single record. At run time, because applications rely on only a single record, they can run in different environments, and even incompatible applications can also run simultaneously as long as in the same environment. Packaged virtual applications are centrally managed in the data center. When a new application deployment (such as installation, update, maintenance, etc.) is required, there is no need to reinstall the application, but to download it in the data center to complete.

5 Conclusion

This paper aims to readability and security issues of electronic record more in-depth research, better summary judgment and maintenance of electronic record readable approach, especially from the technical means on readability ensures a problem to undertake relatively detailed reviews, and by the author in the electronic record readable security project example of the practice of specific problems were analyzed, so that the electronic record readable security practice has practical reference value.

Acknowledgement: This work is supported by the NSFC (61772280), and the PAPD fund from NUIST.

References

Gopinath, V.; Bhuvaneswaran, R. S. (2018): Design of ECC based secured cloud storage mechanism for transaction rich applications. *Computers, Materials & Continua*, vol. 57, no. 2, pp. 341-352.

Han, D.; Li, N.; Bi, K. (2012): Study of virtualization technology in cloud environment. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*. vol. 40, no. 12, pp. 262-265.

Liu, Y. N; Li, J. Y. (2017): Conceptual comparison and linkage between electronic data in law field and electronic records in archival field. *Archives Science Study*, no. 4, pp. 92-99.

Ren, Y. J.; Shen, J.; Liu, D. Z.; Wang, J.; Kim, J. (2016): Evidential quality preserving of electronic record in cloud storage. *Journal of Internet Technology*, vol. 17, no. 6, pp. 1125-1132.

Shen, J.; Zhou, T.; He, D.; Zhang, Y.; Sun, X. et al. (2017): Block design-based key agreement for group data sharing in cloud computing. *IEEE Transactions on Dependable and Secure Computing*, pp. 1.

Tu, **Y. M; Zhang, M. X.** (2016): The influences and strategies deal with the long-term preservation of electronic records from the view of life cycle. *Archives and Construction*, vol. 4, no. 11, pp. 8-13.

Wang, X. (2013): Australian《digital information essence of long-term storage》principle. *Chinese Archives*, no. 4, pp. 52-53.

Xiao, Q.; Wu, L. (2015): Study on digital continuity plan of national archives of Australia. *Information Resources Management Journal*, no. 4, pp. 19-23.

Zhang, Y.; Zang, W. (2013): Storage infrastructure management and optimization scheme in virtualized environment. *Journal of Computer Research and Development*, no. 10, pp. 226-228.