ISBN 978-952-5726-06-0
Proceedings of the 2009 International Workshop on Information Security and Application (IWISA 2009)
Qingdao, China, November 21-22, 2009

Data Security Model for Cloud Computing

Dai Yuefa, Wu Bo, Gu Yaqiang, Zhang Quan, Tang Chaojing
Department of Electronic Science and Engineering, National University of Defense Technology,ChangSha,China
Email: darner248@sina.com

Abstract-**With the development of cloud computing, Data security becomes more and more important in cloud computing. This paper analyses the basic problem of cloud computing data security. With the analysis of HDFS architecture, we get the data security requirement of cloud computing and set up a mathematical data model for cloud computing. Finally we build a data security model for cloud computing.**

Index Terms—**Cloud Computing, HDFS, Data Security Model**

I.INTRODUCTION

Cloud computing appeared in 2006, when Amazon's Elastic Computing Cloud (EC2) fires the world. Many information Enterprises develop their platform for cloud computing. In 2007, Dell releases his solution of cloud computing, at the same time IBM's Blue Cloud comes in. Such as Google's Mapreduce, Microsoftware's Windows Azure .According to an estimation , by 2012, the Cloud computing market should reach $420 billion.All this have show the coming of the epoch time of cloud computing.

The emergence of the Cloud system has simplified the deployment of large-scale distributed systems for software vendors. The Cloud system provides a simple and unified interface between vendor and user, allowing vendors to focus more on the software itself rather than the underlying framework. Applications on the Cloud include Software as a Service system and Multi-tenant databases . The Cloud system dynamically allocates computational resources in response to customers' resource reservation requests and in accordance with customers' predesigned quality of service.

Risk coming with opportunity, the problem of data security in Cloud computing become bottleneck of cloud computing. In this paper we want to set up a security model for cloud computing , the rest of the paper is organized as follows: We present the data security problem of cloud computing in the next section and then discuss the details of requirement of security in Section 3. In Section 4,we focus on the Data security model .Finally, we conclude the paper in Section 5.

II.DATA SECURITY PROBLEM OF CLOUD COMPUTING

A.Security Problem Drive from VM

Whether the IBM's Blue Cloud or the Microsoft's Windows Azure, the virtual machine technology is considered as a cloud computing platform of the fundamental component, the differences between Blue Cloud and Windows Azure is that virtual machine running on Linux operating system or Microsoft Windows operating system. Virtual Machine technology bring obvious advantages, it allows the operation of the server which is no longer dependent on the physical device, but on the virtual servers. In virtual machine, a physical change or migration does not affect the services provided by the service provider. if user need more services, the provider can meet user's needs without having to concern the physical hardware.

However, the virtual server from the logical server group brings a lot of security problems. The traditional data center security measures on the edge of the hardware platform, while cloud computing may be a server in a number of virtual servers, the virtual server may belong to different logical server group, virtual server, therefore there is the possibility of attacking each other ,which brings virtual servers a lot of security threats. Virtual machine extending the edge of clouds makes the disappearance of the network boundary, thereby affecting almost all aspects of security, the traditional physical isolation and hardware-based security infrastructure can not stop the clouds computer environment of mutual attacks between the virtual machine.

B.The Existence of Super-user

For the enterprise providing cloud computing services, they have the right to carry out the management and maintenance of data, the existence of super-users to greatly simplify the data management function, but it is a serious threat to user privacy. Super-powers is a double-edged sword, it brings convenience to users and at the same time poses a threat to users. In an era of personal privacy, personal data should be really protected, and the fact that cloud computing platform to provide personal services in the confidentiality of personal privacy on the existence of defects. Not only individual users but also the organizations have similar potential threats, e.g. corporate users and trade secrets stored in the cloud computing platform may be stolen. Therefore the use of super user rights must be controlled in the cloud.

corresponding author:Dai Yuefa
Email:darner248@sina.com

C.Consistency of Data

Cloud environment is a dynamic environment, where the user's data transmits from the data centre to the user's client. For the system, the user's data is changing all the time. Read and write data relating to the identity of the user authentication and permission issues. In a virtual machine, there may be different users' data which must be strict managed.

The traditional model of access control is built in the edge of computers, so it is weak to control reading and writing among distributed computers. It is clear that traditional access control is obviously not suitable for cloud computing environments. In the cloud computing environment, the traditional access control mechanism has serious shortcomings.

D.New Technology

The concept of cloud computing is built on new architecture. The new architecture comprised of a variety of new technologies, such as Hadoop, Hbase, which enhances the performance of cloud systems but brings in risks at the same time. In the cloud environment, users create many dynamic virtual organizations, first set up in co-operation usually occurs in a relationship of trust between organizations rather than individual level. So those users based on the expression of restrictions on the basis of proof strategy is often difficult to follow; which frequently occurs in many of the interactive nodes between the virtual machine, and is dynamic, unpredictable. Cloud computing environment provides a user of the "buy" the full access to resources which has also increased security risks.

III.REQUIREMENT OF SECURITY

Following, with the analysis of the widely used cloud computing technology--HDFS (Hadoop Distributed File System), we will get the data security needs of cloud computing. HDFS is used in large-scale cloud computing in a typical distributed file system architecture, its design goal is to run on commercial hardware, due to the support of Google, and the advantages of open source, it has been applied in the basis of cloud facilities. HDFS is very similar to the existing distributed file system, such as GFS (Google File System); they have the same objectives, performance, availability and stability. HDFS initially used in the Apache Nutch web search engine and become the core of Apache Hadoop project.

HDFS used the master/slave backup mode. As shown in Figure 1. The master is called Namenode, which manages the file system name space and controls access to the client. Other slave nodes is called Datanode, Datanode controls access to his client. In this storage system, a file is cut into small pieces of paper, Namenode maps the file blocks to Datanodes above. While HDFS does not have the POSIX compatibility, the file system still support the creation, delete, open, close, read, write and other operations on files.

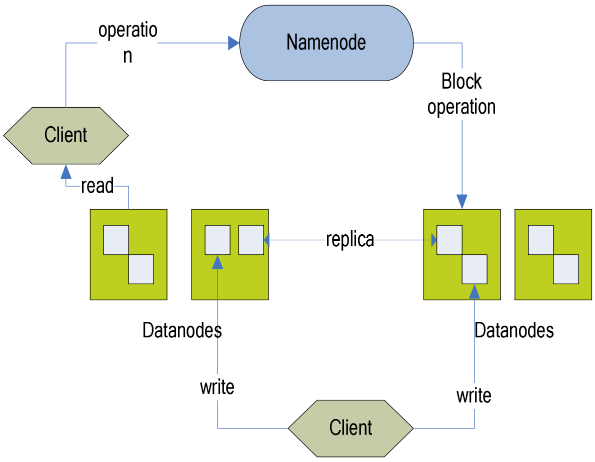

Figure 1 HDFS Architecture

By analyzing of HDFS, data security needs of cloud computing can be divided into the following points:

The client authentication requirements in login: The vast majority of cloud computing through a browser client, such as IE, and the user's identity as a cloud computing applications demand for the primary needs.

The existence of a single point of failure in Namenode: if namenode is attacked or failure, there will be disastrous consequences on the system. So the effectiveness of Namenode in cloud computing and its efficiency is key to the success of data protection, so to enhance Namenode's security is very important.

The rapid recovery of data blocks and r/w rights control: Datanode is a data storage node, there is the possibility of failure and can not guarantee the availability of data. Currently each data storage block in HDFS has at least 3 replicas, which is HDFS's backup strategy. When comes to how to ensure the safety of reading and writing data, HDFS has not made any detailed explanation, so the needs to ensure rapid recovery and to make reading and writing data operation fully controllable can not be ignored.

In addition to the above three requirements, the other, such as access control, file encryption, such as demand for cloud computing model for data security issues must be taken into account.

IV.DATA SECURITY MODEL

A.Principle of Data Security

All the data security technic is built on confidentiality, integrity and availability of these three basic principles. Confidentiality refers to the so-called hidden the actual data or information, especially in the military and other sensitive areas, the confidentiality of data on the more stringent requirements. For cloud computing, the data are stored in "data center", the security and confidentiality of user data is even more important. The so-called integrity of data in any state is not subject to the need to guarantee unauthorized deletion, modification or damage. The availability of data means that users can have the expectations of the use of data by the use of capacity.

B.Data Security Model

Data model of cloud computing can be described in math as follows:

$$D_f = C(NameNode)\text{；} \quad （1）$$

$$K_f = f * D_f\text{；} \quad （2）$$

C(.):the visit of nodes；

D_f :the distributed matrix of file f ；

K_f :the state of data distribution in datanodes；

f ： file，file f can be described as ：

f ={F(1),F(2),…..F(n)}，means f is the set of n file blocks。F(i) $\cap$ F(j)= ϕ ,i $\neq$ j;I,j $\in$ 1,2,3,...n ;

D_f is a Zero-One matrix，it is L*L，L is the number of datanode.

To enhance the data security of cloud computing, we provide a Cloud Computing Data Security Mode called C2DSM.It can be described as follows:

$$D_f^{'} = C_A \text{（namenode）} \quad (3)$$

$$D_f = \text{M. } D_f^{'} \quad (4)$$

$$K_f = E(f)\ D_f \quad (5)$$

C_A (.): authentic visit to namenode；

$D_f^{'}$:private protect model of file distributed matrix；

M：resolve private matrix；

$E(f)$：encrypted file f block by block，get the encrypted file vector；

This model can be show by Figure 2.

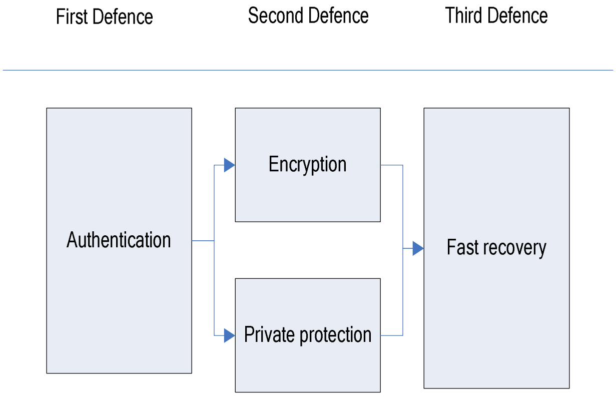

Figure2 Cloud Computing Data Security Mode

The model used three-level defense system structure, in which each floor performs its own duty to ensure that the data security of cloud layers.

The first layer: responsible for user authentication, the user of digital certificates issued by the appropriate, manage user permissions;

The second layer: responsible for user's data encryption, and protect the privacy of users through a certain way;

The third layer: The user data for fast recovery, system protection is the last layer of user data.

With three-level structure, user authentication is used to ensure that data is not tampered. The user authenticated can manage the data by operations: Add, modify, delete and so on. If the user authentication system is deceived by illegal means, and malign user enters the system, file encryption and privacy protection can provide this level of defense. In this layer user data is encrypted, even if the key was the illegally accessed, through privacy protection, malign user will still be not unable to obtain effective access to information, which is very important to protect business users' trade secrets in cloud computing environment. Finally, the rapid restoration of files layer, through fast recovery algorithm, makes user data be able to get the maximum recovery even in case of damage.

From the model there will be follow theorems:

Theory one: If D_f is not a full order, then the user lost his data.

Verify:

D_f if the file distribution matrix, so with the formula (5) , K_f is the L length vector.

If D_f is not full order, D_f can be convert to D_f^* , D_f^* is （L-i）*（L-i）matrix, i $\geq$ 1;

K_f become L-I length vector, that make confliction to the definition of the model.

Theory two: if $\sum_{i=1}^{n} K_f(i)$<n, then the data of the user is damaged. $K_f(i)$ means the value of position i of file vector K_f

Verify:

$\sum_{i=1}^{n} K_f(i)$ means the number of store data in datanode, with definition f ={F(1),F(2),…..F(n)},if F(i) not existence, i=1，2….n，then the file store failure. if $\sum_{i=1}^{n} K_f(i)$ <n, then there will be i=1，2….n, let $K_f(i)$ =0,F(i) not existence in f ,the file is damaged.

Theory three: if there existed matrix J,J ≠ M, but $D_f = J \,.\, D_f^{'}$,the private of user leak.

Verify:

M is the user's private matrix. With the matrix M we can get D_f .if J existed then illegal user may get D_f by J .there is existence of private leakence.

V.CONCLUSION

As the development of cloud computing, security issue has become a top priority. This paper discusses the cloud computing environment with the safety issues through analyzing a cloud computing framework--HDFS's security needs. Finally we conclude a cloud computing model for data security.

REFERENCES

[1] Rajkumar Buyya Market-Oriented Cloud Computing:Vision,Hype,and Reality for Delivering IT Services as Computing Utilities 2008

[2] Jean-Daniel Cryans,Criteria to Compare Cloud Computing with Current Database Technology 2008

[3] Christopher Moretti,All-Pairs: An Abstraction for Data-Intensive Cloud Computing IEEE 2008

[4] Huan Liu, Dan Orban， GridBatch: Cloud Computing for Large-Scale Data-Intensive Batch Applications IEEE DOI 10.1109/CCGRID.2008.30

[5] Mladen A. Vouk， Cloud Computing – Issues, Research and Implementations Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246

[6] Bob Gourley ， Cloud Computing and Net Centric Operations ， Department of Defense Information Management and Information Technology Strategic Plan 2008-2009

[7] Cloud Computing Security:making Virtual Machines Cloud-Ready ,www.cloudreadysecurity.com 2008

[8] Greg Boss, Cloud Computing，IBM 2007.10

[9] Jeffrey Dean and Sanjay Ghemawat ， MapReduce: Simplied Data Processing on Large Clusters Google, Inc 2004.

[10] David Chappell ,Introducing the Azure Services Platform October 2008

[11] O'Reilly，Programming Amazon Web Services， March 15, 2008