DATA SECURITY MODEL FOR CLOUD COMPUTING

Pooja Dhawan

Assistant Professor, Deptt of Computer Application and Science Hindu Girls College, Jagadhri 135 001

ABSTRACT

Cloud Computing may be a way of making the businesses more efficient. It enables to lower the cost of I.T. functions of the business and enables the business managers to focus on the core business. What it does is to shift the information technology function of the business to a remote service provider about whom we may not have any information. The concept of shifting computing to a shared service provider is not new. What is new is that the cost of cloud computing is falling so dramatically, that considering outsourcing to the cloud is no longer rare. In an era of tight budgets, the opportunity to make financial savings means, that the cloud computing looks especially attractive. However data security becomes more and more important in cloud computing as the data is hosted on an unknown remote computer and all the transactions have to be made to this computer. This paper analyses the basic problem of data security in cloud computing. Data security requirements of cloud computing were identified and a data security model is proposed through the analysis of HDFS architecture under the framework of cloud computing.

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

INTRODUCTION

Cloud computing appeared in 2006, when Amazon's Elastic Computing Cloud (EC2) fired the world. Following this many information enterprises developed their platform for cloud computing. In 2007, Dell releases its solution of cloud computing, and at the same time IBM's Blue Cloud, Google's Map reduce, and Microsoft's Windows Azure appear. According to some estimates, that by 2012, the Cloud computing market would reach a staggering figure of $420 billion. All this shows the emergence of cloud computing as a practical and viable alternative technology. The emergence of cloud system has simplified the deployment of large-scale distributed systems by the business houses. The cloud system provides a simple and unified interface between the software service provider vendor and the business houses that use the services, allowing business houses to focus more on the functionality and business aspects rather than the underlying framework and implementation features. Applications on the Cloud include software as a Service system, transaction processing and Multi-tenant databases. The Cloud system dynamically allocates computational resources to the customer/ user in response to customers' resource reservation requests and also in accordance with customers' quality of service requirements. Risks come with opportunities and the problem of data security becomes bottleneck in Cloud computing.

In this paper the author presents a security model for cloud computing.

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 to be the fundamental component of the cloud computing platform. The basic difference between Blue Cloud and Windows Azure is that in one case virtual machine runs on Linux operating system and in the other it runs on Microsoft's Windows operating system. Virtual Machine technology brings obvious advantages by allowing the operation of the server that no longer depends on the physical device and the operating system, but on the virtual servers. In virtual machine, a physical change of the server or migration to a different server does not affect the services provided by the service provider. If user needs more services, the provider would meet the user's needs without any consideration of the existing physical hardware. However, the virtual server, from the logical server group brings a lot of security problems. The traditional data center's security measures on the edge of the hardware platform are no longer available in cloud computing. Cloud computing may be hosting a server on a number of virtual servers, and the virtual servers may belong to different physical server groups. Therefore there is the possibility of attacking one of them through the other, which brings virtual servers a lot of security threats. Virtual machine extending the edge of cloud makes the network boundary to disappear, thereby affecting almost all aspects of security, in the absence of the traditional physical isolation and hardware-based security infrastructure. Thus cloud computer environment cannot stop mutual attacks between the virtual machines.

B. The Existence of Super-user

The enterprise that provides the cloud computing service, would have to carry out the management and maintenance of data as a super-user, and this greatly simplifies the data management function. However it poses a serious threat to the users' privacy. Super-users' power is a double edged sword, and while it brings convenience to the users, at the same time it poses a threat to users. In an era of personal privacy, personal data should be really protected. Cloud computing platform cannot provide services with the desired level of confidentiality for the personal privacy due to the above limitation. Not only individual users but also the organizations would have to face similar potential threats, such as corporate users information and the trade secrets of the organizations stored on the cloud computing platform may be accessed by others and thus stolen. Therefore the use of super- user rights must be controlled in the cloud.

C. Consistency of Data

Cloud environment is a dynamic environment, where the user's data is transmitted from the data center to the user's client and vice versa. Most of the operations are read and write client's data. The traditional model of access control that is built in the edge of computers is based on authentication of the client. This authentication scheme of traditional model is a weak control for reading and writing

data among distributed computers that are shared by a number of users as is the case in cloud computing. In a virtual machine, different users' data exists on the same machine and must be strictly managed. It is clear that traditional access control is not suitable for cloud computing environments as it has serious shortcomings.

D.New Technology

The concept of cloud computing is built on new architecture. The new architecture comprises of a variety of new technologies, such as Hadoop, Hbase, which enhance the performance of cloud systems but bring in risks at the same time. In the cloud environment, users create many dynamic virtual organizations, and these would be set up in co-operation, which usually occurs in a relationship of trust between organizations rather than at individual level. So those users who accepted the restrictions on the basis of proof strategy, would often find it difficult to follow them; and this frequently occurs in many of the interactive nodes between the virtual machines. This noncompliance is dynamic and unpredictable. Cloud computing environment provides a user who "buys" has full access to resources with all the increased security risks.

REQUIREMENT OF SECURITY

We will analyse and get the data security needs of cloud computing for the widely used cloud computing technology—HDFS (Hadoop Distributed File System), HDFS is used in large-scale cloud computing environment. The distributed file system architecture, running on commercial hardware, has been selected as the basis for the cloud facilities due to the support provided to it by Google, and the advantages of open source. HDFS is very similar to the existing distributed file system of Google, the GFS (Google File System). They both 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 are called Datanodes. Datanodes control the access to their clients. In this storage system, a file is cut into small blocks and the Namenode maps the file blocks to the respective Datanodes. While HDFS does not have the POSIX compatibility, the file system still supports the operations create, delete, open, close, read, write and others on the files.

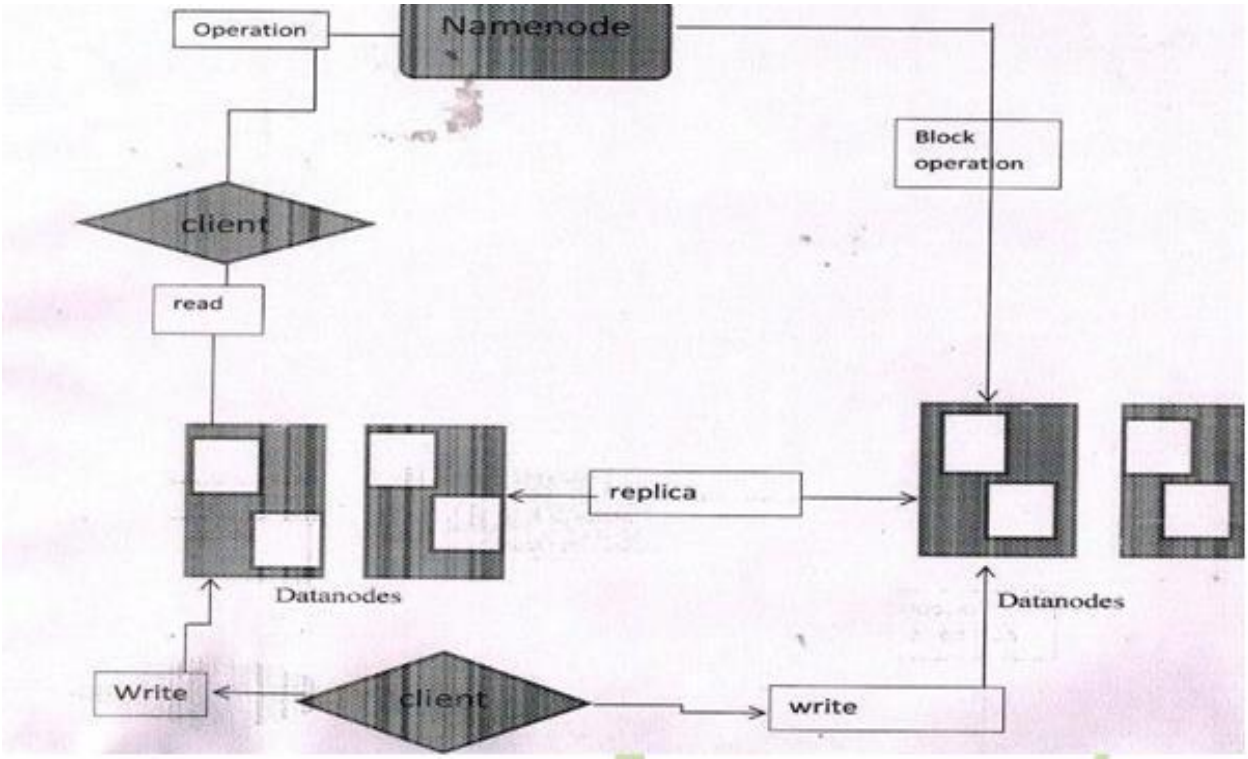

Figure 1 HDFS Architecture

The data security needs of HDFS based cloud computing may be considered under the following points:

- The client authentication requirements at login: The vast majority of cloud computing would be through a browser client, such as IE or Firefox The user's identity is to be authenticated for accessing primary needs of cloud computing applications.

- The existence of a single point of failure in Namenode: Successful attack on and failure of the namenode will have disastrous consequences on the system. So the effectiveness and the efficiency of protecting the Namenode is the key to the success of data protection in cloud computing and so to enhance the Namenode's security is a primary goal in coud computing.

- Rapid recovery of data blocks and r/w right controls: Datanode is a data storage nod, and its failure would mean the non-availability of data. This problem is often addressed through triple redundancy. Currently each data storage block in HDFS has at least 3 replicas and this is HDFS's backup strategy.

HDFS does not provide any detailed procedure for secure reading and writing of data. There is also the need to ensure rapid recovery from glitches and crashes and make the operations like reading and writing etc. fully controllable. In addition to access control, file encryption, scalable demand for cloud computing, and data security model, must be taken into account.

DATA SECURITY MODEL

All data security techniques are built on three basic principles namely: confidentiality, integrity, and availability of data whenever needed. Confidentiality refers to the process of hiding the data or information from those who do not have access rights to the data. In the cloud computing environment, the data is stored in "data center", and the confidentiality of data is more important and takes more stringent requirements.

The data integrity includes prevention 0f unauthorized deletion, addition, modification or damage of the data. The availability of data means that the users could access the data when needed. The requirement may include recovery of data from the corrupted or crashed system.

Figure2 Cloud Computing Data Security Model

A suggested data security model is presented in fig. 2. The model uses three-layered security system architecture, in which each layer performs its own function to achieve the desired data security of the cloud.

The first layer is responsible for the user authentication by verifying the user digital certificates issued by appropriate recognized body and then manage as per the user's access rights. This layer ensures only authorized persons can access data and perform permitted operations on the data. This will ensure that the data is not accessed and tampered by the unauthorized persons through additions, deletions etc.

The second layer is responsible for user's data encryption, and protects the privacy of users while the data is being transferred. This will also protect the data against those illegal users who get through the authentication system of layer1.

The third layer is responsible for fast recovery of the system and would ensure the user has data all the time, even when the system is damaged either intentionally or unintentionally.

CONCLUSION

The development of cloud computing has brought into focus a number of security issues. This paper discusses the cloud computing environment with the safety issues through an analysis of cloud computing framework - HDFS. Finally we propose a data security model cloud computing that addresses the data security needs.

REFERENCES

[1] *Rajkumar Buyya Market-Oriented Cloud Computing: Vision, Hype and Reality for Delivering IT Services as Computing Utilities 25-27 Sept.2008,5-13*

[2] *Jean-Daniel Cryans, Criteria to Compare Cloud Computing with Current Database Technology 2008 doi>10.1007/978-3-540-89403-2_11,114-126*

[3] *Christopher Moretti, All-Pairs: An Abstraction for Data-Intensive Cloud Computing IEEE 2008[4] Huan Liu, Dan Orban14-18 April 2008,1-11*

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

[5] *Department of Defense Information Management and Information Technology Strategic Plan2008-2009*

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

[7] *Map Reduce: Simplied Data Processing on Large Clusters Google, Inc2004.*