Towards Secure Cloud Computing Phase: Information Possession

Jigisha Pandya

Abstract: **In contrast to traditional solutions, where the IT services are under proper physical, logical and personnel controls, Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services may not trustworthy. This unique attribute, however, poses many new security challenges which have not been well understood. In this paper, I characterize the problems and their impact on adoption. In addition, and equally importantly, I describe how the combination of existing research thrusts has the potential to alleviate many of the concerns impeding adoption. I argue that with continued research advances in trusted computing and computation-supporting encryption, life in the cloud can be advantageous from a business intelligence standpoint over the isolated alternative that is more common today. This Paper also analyses the basic problem of cloud computing data security. With the analysis of HDFS architecture.**

Keywords: **Cloud Computing, Data Security, Information integrity, Security requirements, Network architecture**

1. INTRODUCTION

Several trends are opening up the era of Cloud computing, which is an Internet-based development and use of computer technology. The ever cheaper and more powerful processors, together with the software as a service (SaaS) computing architecture, are transforming data centers into pools of computing service on a huge scale. The increasing network bandwidth and reliable yet flexible network connections make it even possible that users can now subscribe high quality services from data and software now subscribe high quality services from data and software that reside solely on remote data centers**.** Moving data into the cloud offers great convenience to users since they don't have to care about the complexities of direct hardware management. The pioneer of Cloud Computing vendors, Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) are both well known examples. While these internet-based online services do provide huge amounts of storage space and customizable computing resources, this computing platform shift, however, is eliminating the responsibility of local machines for data maintenance at the same time. As a result, users are at the mercy of their cloud service providers for the availability and integrity of their data. Recent downtime of Amazon's S3 is such an example**.**

From the perspective of data security, which has always been an important aspect of quality of service, Cloud Computing inevitably poses new challenging security threats for number of reasons. Firstly, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted due to the users' loss control of data under Cloud Computing. Therefore, verification of correct data storage in the cloud must be conducted without explicit knowledge of the whole data.

A representative network architecture for cloud storage service architecture is illustrated in Figure 1. Three different network entities can be identified as follows:

User: an entity, who has data to be stored in the cloud and relies on the cloud for data storage and computation, can be either enterprise or individual customers. Cloud Server (CS): an entity, which is managed by *cloud service provider* (CSP) to provide data storage service and has significant storage space and computation resources (I will not differentiate CS and CSP hereafter.). Third Party Auditor (TPA): an optional TPA, who has expertise and capabilities that users may not have, is trusted to assess and expose risk of cloud storage services on behalf of the users upon request.

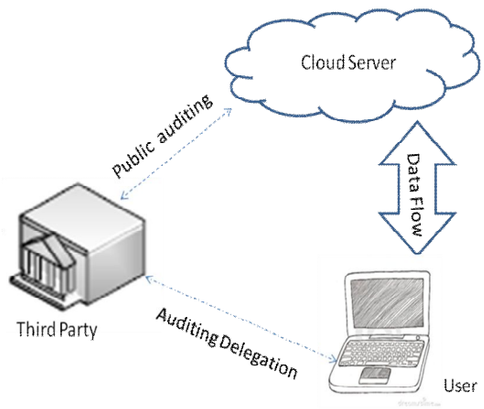
In cloud data storage, a user stores his data through a CSP into a set of cloud servers, which are running in a simultaneous, cooperated and distributed manner. Data redundancy can be employed with technique of erasure correcting code to further tolerate faults or server crash as user's data grows in size and importance. Thereafter, for application purposes, the user interacts with the cloud servers via CSP to access or retrieve his data.

In some cases, the user may need to perform block level operations on his data. The most general forms of

these operations I am considering is block update, delete, insert and append. Note that in this paper, I put more focus on the support of file-oriented cloud applications other than non-file application data, such as social networking data. In other words, the cloud data I am considering is not expected to be rapidly changing in a relative short period. As users no longer possess their data locally, it is of critical importance to ensure users that their data are being correctly stored and maintained. That is,

Department of MCA, G. H. Raisoni College, pune-07.
Jigisha.pandya@raisoni.net

users should be equipped with security means so that they can make continuous correctness assurance (to enforce cloud storage service-level agreement) of their stored data even without the existence of local copies. In case that users do not necessarily have the time, feasibility or resources to monitor their data online, they can delegate the data auditing tasks to an optional trusted TPA of their respective choices. However, to securely introduce such a TPA, any possible leakage of user's outsourced data towards TPA through the auditing protocol should be prohibited. In our model, I assume that the point-to-point communication channels between each cloud server and the user is authenticated and reliable, which can be achieved in practice with little overhead. These authentications handshakes are omitted in the following presentation.



In this paper I want to set up a security model for cloud computing, the rest of the paper is organized as follows: I 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. I focus on the Data security model. Finally; I conclude the paper in Section 5.

2. SECURITY PROBLEM OF CLOUD COMPUTING

A. 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.

B. Administrative Access To Servers And Applications:

One of the most important characteristics of cloud computing is that it offers "self-service" access to computing power, most likely via the Internet. In traditional datacenters, administrative access to servers is controlled and restricted to direct or on-premise connections. In cloud computing, this administrative access must now be conducted via the Internet, increasing exposure and risk. It is extremely important to restrict administrative access and monitor this access to maintain visibility of changes in system control.

Information Security Magazine asks: "How do you perform an on-site audit when you have a distributed and dynamic multi-tenant computing environment spread all over the globe? It may be very difficult to satisfy auditors that your data is properly isolated and cannot be viewed by other customers."

3. REQUIREMENT OF SECURITY

In later sections, with the analysis of the widely used cloud computing technology--HDFS (Hadoop Distributed File System), I will get the data security requirements of cloud computing. HDFS is applied in enterprise-scale cloud computing in typical distributed file system architecture, its design aim is to run on commercial hardware, due to the support of Google, and the advantages of open source technology, it has been used 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 derivatives, performance, scalability and stability. HDFS initially implemented in the Apache Nutch web search engine and become the core of Apache Hadoop project. HDFS applied the master/slave backup strategy. As shown in Figure 1. The master is identified as Namenode, which administrates the file system name space and controls access to the client. Other slave nodes is named as Datanode, Datanode defines access to his client.

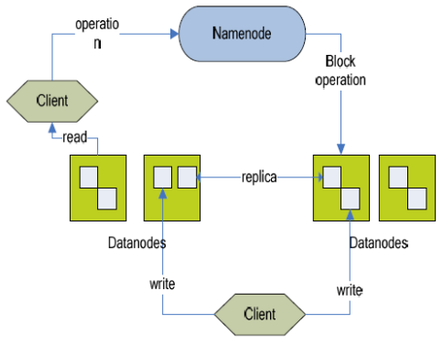


Figure 1: HDFS Architecture

4. SOLUTION FOR SECURITY PROBLEM

I now describe primary elements of our strategy. The core concern is that, with the invention of the cloud computing, the cloud provider also has granular access control of the cloud users' data\information. My aim to supply tools supporting the existing capabilities of the cloud while limiting cloud provider control of data and enabling all cloud users to benefit from cloud data via enhanced enterprise intelligence.

A. Information-centric security:

In order for business to extend control to data\information in the cloud, I propose moving from protecting data from the outside (system and applications which use the data) to protecting data from within. I call this approach of data and information protecting itself information-centric. This self-protection demands intelligence to be put in the data itself. Data requires to be self-describing and defending, regardless of its functioning environment. Data needs to be encrypted and packaged with a usage strategy. When used, data should consult its strategy and attempt to re-create a secure environment using virtualization and provide access to itself only if the environment is verified as authentic (using Trusted Computing). Information-centric security is a natural extension of the trend toward granular, persistent, and more usable data security.

B. High-Assurance Remote Server Attestation:

I have noted that lack of transparency is discouraging enterprises from shifting their information to the cloud.

Data owners are curious to audit how their data is being handled at the cloud, and in specifically, to ensure that their data is not being abused or leaked, or at least have an unalterable audit trail when it does happen. Currently customers must be satisfied with cloud providers using manual auditing procedures like SAS-70.

Another direction to address this issue is based on Trusted Computing. Imagine a authenticated monitor available at the cloud data server that can watch or monitor the operations of the cloud data server. The authentic monitor can provide "proofs of compliance" to the data owner, specifying that certain access policies have not been compromised. To guarantee integrity of the monitor, Trusted Computing also grants secure bootstrapping of this monitor to execute beside (and securely isolated from) the underlying operating system and other programs. The monitor can entitle access control policies and perform monitoring/auditing tasks. To generate a "evidence of compliance", the code of the monitor is authenticated, as well as a "sheet of compliance" produced by the monitor. When the data owner gets this proof of compliance, it can show its compliance that the required monitor program has executed, and that the cloud data server is in check with access control policies.

C. Privacy-Embellished Business Intelligence

A secondary way to retaining control of information is to require the ciphering of all cloud data. The problem is that ciphering has its own restrictions for data use. In particular finding and indexing the information gets problematic. For example, if information is available in clear-text, user can effectively find a document by specifying a tag. This is not possible to do with regular, randomized cipher schemes. State-of-the-art crypto-analysis may offer unconventional techniques to mitigate these problems. Cryptography analysts have recently invented out-of-the-box encryption schemes that allow operation and computation on the cipher-text. For example, searchable cipher technique allows the data owner to calculate a capability from his secret key. A capability encodes a search query, and the cloud can use this capability to decide which documents match the search query, without specifying any supportive information. Other cryptographic primitives such as encryption and Private Information Retrieval (PIR) can do calculations on ciphered data without deciphering it. As these cryptographic strategies mature, they will open up new horizon for cloud computing data security.

4. DATA SECURITYMODEL

A. Principle of Data Security:

All the data security technique 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 theuse 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); (1)
K f= f * D f ; (2)
C(.):the visit of node;
D f :the distributed matrix of file f ;
K f :the state of data distribution in data nodes;

f :file, file f can be described as :
f = {F(1),F(2),…..F(n)},means f is the set of n file blocks $F(i)\cap F(j)=\phi$,$i \neq j;I,j \in 1,2,3,...n$;
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, I provide a Cloud Computing Data Security Mode called C2DSM.It can be described as follows:

D' f = C A (namenode) (3)
D f =M. D' f (4)
K f = E(f) D f (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.

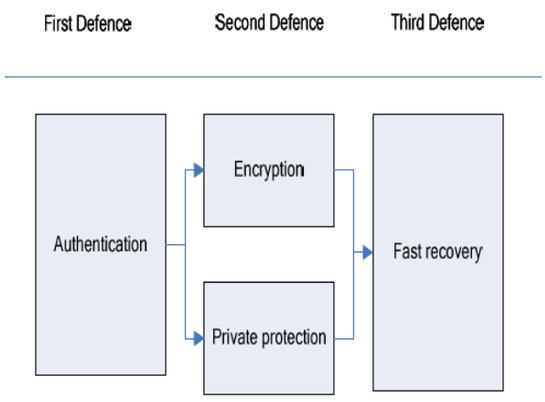


Figure 2 Cloud computing data security model

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) , f K 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.

6. 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 I conclude a cloud computing model for data security.

7. 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 DOI10.1109/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