Bridging the Missing Link of Cloud Data Storage Security in AWS

[†]Jun Feng, [†]Yu Chen*, [‡]Pu Liu

[†]Dept. of Electrical & Computer Engineering, SUNY - Binghamton, Binghamton, NY 13902
[‡] Systems and Technology Group, IBM Endicott, Endicott, NY 13760

***Abstract* - The data that is stored and/or transmitted on the Internet has been called "the blood of the IT". Along with the infrastructure and network based applications, data storage has been recognized as one of the major dimensions of information technology. The prosperity of Cloud Computing requires the moving from server-attached storage to distributed storage. Along with variant advantages, the distributed storage also poses new challenges in creating a secure and reliable data storage and access facility over insecure or unreliable service providers. The security of data stored in the cloud is one of the challenges to be addressed before the novel pay-as-you-go business model is applied widely. In this research, we revealed the vulnerability in the Amazon's AWS cloud and discussed technical approaches towards potential effective solutions.**

Keywords: Cloud Computing, Data Storage, Information Security.

1. Introduction

Cloud Computing has gained great attention from both industry and academics since 2007. As a further step on top of Grid Computing, Cloud Computing aims to provide users more flexible services in a transparent manner – all services are allocated in a "cloud" that actually is a collect of devices and resources connected through the Internet. However, there are a lot of challenges to be addressed before the beautiful blueprint is accepted widely. One of the most impending tasks is the security of data storage in the cloud.

As shown in Figure 1, cloud computing services can fundamentally provide services based on the Internet, through which to access large amounts of data and computing resources. Although the definition of Cloud Computing is not clear yet, several pioneer commercial implementations have been constructed and open to public, such as Amazon's Computer Cloud AWS (Amazon Web Service) [1] and Microsoft Azure Service Platform [5].

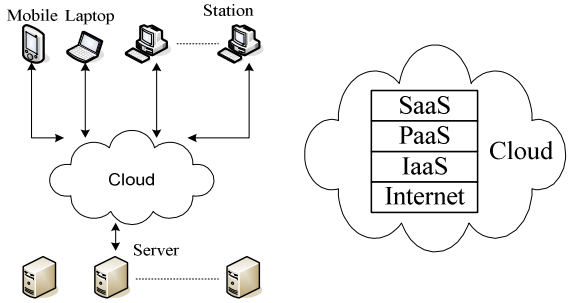

Figure 1. Illustration of Cloud Computing principle.

While cloud computing leads to the new pay-as-you-go business model, it also brings new security issues. The IDC survey in Aug. 2008 has shown that security is the most serious concerns customers have ascribed to the Cloud computing [3]. Meanwhile, the research in cloud computing security is far from mature [4].

First of all, the uniqueness of the cloud computing security is not recognized. Some researchers think that cloud computing security is not much different from existing security practices and the security aspects can be well managed with the existing techniques such as digital signature, encryption, firewalls, and/or the isolation of virtual environments, etc. [4]. Secondly, the specific security requirements in cloud computing are still cloudy to the community. Nevertheless, the cloud security is an ongoing important area of research. Many consultants and security agencies have issued warnings on the security threats in the cloud computing model [2]. And cloud consumers still wonder whether the cloud is secure. It requires more than conventional security mechanisms.

In this paper, we reveal the vulnerability in data integrity in the cloud environment using Amazon's AWS [1]. Then we propose to solve this problem with new signature proxy based solutions on top of the current cloud storage model.

2. Integrity Vulnerability in AWS

In normal network based applications, the user authentication, data confidentiality and integrity can be solved through IPSec proxy using encryption and digital signature, and the key exchange can be solved SSL proxy. Such methods have been applied to today's cloud computing to secure the data and communication and the service providers claim that their services are secure.

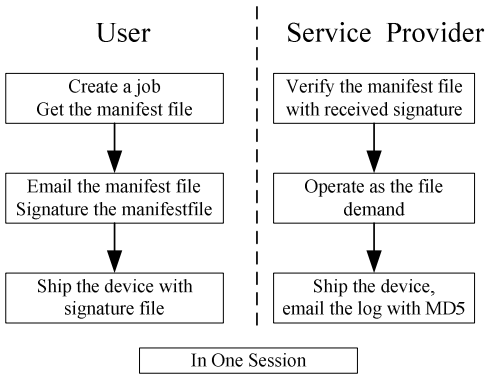

Figure 2. AWS data processing steps.

Figure 2 presents the flowchart of data processing adopted in Amazon's AWS [1], which shows how users can load their data into the cloud. To uniquely identify and authenticate the user request, AWS uses the signature file describing the method used to encrypt the value of data and

the signature value itself. Service Provider will verify the signature with the manifest file it received. The AWS adopts two paths (email and ship) to make the more secure when uploading or downloading the file.

The procedure is secure in each individual session. The integrity of the data in the transmission can be guaranteed by the IPSec or SSL proxy applied. However, for the cloud storage service, the same data integrity depends on the security in both of the two sessions, upload and download, and the operations between them.

Unfortunately, the current procedure of AWS cannot guarantee the integrity. The uploading phase can only ensure the data received by the cloud storage is the data user uploaded; the downloading phase can guarantee the data user retrieved is the data cloud storage recorded. However, one link is missing: there is no mechanism for the user/provider to check whether the record is modified in the cloud storage. This vulnerability leads to two following questions that need to be answered:

- **Upload-to-Download Integrity:** since the integrity in uploading and downloading phase are handled separately, how can the user/provider know the data retrieved from the cloud is the same data that the user uploaded previously? In another words, how to guarantee the data integrity between different data transfer sessions?
- **Repudiation between users and service providers:** When data error happened without the transmission error in the downloading phase, how can the user/provider prove his innocence? Or, if the user/provider modify the data and say the data is modified by the other site, how can the other site prove his innocence?

3. Bridging the Missing Link

This section presents couples of solutions based on the digital signature and the authenticate code. According to whether there is a third authorities certified (TAC) by the user and provider, whether using the secret key sharing technique (SKS), there are four solutions for the problem.

3.1 With both TAC and SKS

In the Uploading phase, at first both the user and the provider agree on the integrity of the uploaded data, then they get the data authenticated code by MAC or HASH, add their digital signature, encrypt it and upload the result to the third authorities, then the user and the provider share the key by the SKS.

In the Downloading phase, both user and provider recover the key by the SKS, and then provider checks the data before he sends it by the data authenticated code. The user checks the data after he receives it by the data authenticated code.

3.2 With TAC but without SKS

In the Uploading phase, firstly both the user and the provider agree on the integrity of the uploaded data, then they get the data authenticated code by MAC or HASH, add their digital signature, and upload the result to the third authorities.

In the Downloading phase, both user and provider download the data authenticated code, and then provider checks the data before he sends it by the data authenticated code. The user checks the data after he receives it by the data authenticated code.

3.3 With SKS but without TAC

In the Uploading phase, firstly both the user and the provider agree on the integrity of the uploaded data, then they get the data authenticated code by MAC or HASH, add their digital signature, and share the result by SKS

In the Downloading phase, both user and provider recover the data authenticated code by SKS, then provider checks the data before he sends it by the data authenticated code. The user checks the data after he receives it by the data authenticated code.

3.4 With Neither TAC nor SKS

In the Uploading phase, firstly both the user and the provider agree on the integrity of the uploaded data, then they get the data authenticated code by MAC or HASH, add each digital signature and send to each other.

In the Downloading phase, the provider checks the data by the data authenticated code before sending. The user checks the data on receiving by the data authenticated code.

4. Summary

Cloud Computing has been a hot topic since 2007. It has been considered as the future of IT that leads to novel computing models. For instance, the users can purchase computing/processing utility as we purchase water and electricity today. However, security has been one the major concerns that prevents commercial applications from being accepted widely.

In this paper, we revealed the existing vulnerability in the Amazon's AWS cloud in data integrity due to the missing of connection between the uploading and downloading phases. Based on the digital signature and the authenticate code, we proposed four solutions to make up the missing link. At present time we cannot tell which solution is the most suitable to be implemented in practice. In our ongoing effort, we are implementing them on our network security testbed. The correctness and performance of these schemes will be tested through intensive experimental study.

References

[1] Amazon Import/Export Developer Guide Version 1.2, http://aws.amazon.com/documentation/, August 2009.

[2] J. Heiser and M. Nicolett, "Assessing the Security Risks of Cloud Computing," *Gartner* Inc., June 2, 2008.

[3] F. Gens, "IDC on 'the Cloud': Get Ready for Expanded Research", http://blogs.idc.com/ie/?p=189, Sept. 23, 2008.

[4] K. M. Khan, "Security Dynamics of Cloud Computing," *CUTTER IT JOURNAL*, June/July 2009, p38-43

[5] Microsoft Azure Services Platform, http://www.microsoft.com/azure/default.mspx, 2009.