Nov 2nd, 10:30 AM - 10:45 AM

Data Integrity Check using Hash Functions in Cloud environment

Selman Haxhijaha
University for Business and Technology, selman.haxhijaha@ubt-uni.net

Gazmend Bajrami
University for Business and Technology, gazmend.bajrami@ubt-uni.net

Fisnik Prekazi
University for Business and Technology, fisnik.prekazi@ubt-uni.net

Data Integrity Check using Hash Functions in Cloud environment

Selman Haxhijaha [1], Gazmend Bajrami [2], Fisnik Prekazi [3]
[123] Faculty of Computer Science and Engineering, University for Business and Tecnology –
{selman.haxhijaha, gazmend.bajrami, fisnik.prekazi}@ubt-uni.net

Abstract. The concept of cloud computing is currently being widely adopted by many business and organizations. Cloud Computing offers immense amount of resources, available for end users by employing various flexible paying methods. The opportunity to choose between several cloud providers is referred by complexity of integrated cloud computing solution. Cloud services offer many benefits to the data owner and users, but to take advantage of the benefits of cloud computing and to make the cloud viable as a computing platform, the data and the service hosted in the cloud must be fully secured. This research paper points out how third party auditors can be avoided and proposes a specific solution which involves the customer safeguarding the data integrity by himself in a very simple and efficient way by utilizing existing hash generating algorithm.

Keywords: cloud, hash, data integrity, SaaS, PaaS, IaaS,

1 Introduction

Cloud computing is a new computing paradigm [12], where various computing resources are provided as a service to the end user. Cloud computing as a technology [1] is becoming more popular, both in the academic world and in the IT industry. One of the paramount advantages of Cloud Computing is on-demand self-service, providing dynamic scalability of the infrastructure and services. Cloud computing is a widely used computing technique which provides three main models SaaS (Software / Storage as a service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service)[6]. These models work on a pay-per-use system and provide users the ability to access their database and applications remotely. Very common examples are, Amazon's EC2, Microsoft Azure and Salesforce CRM. The typical services provided by cloud providers include resources like data storage or software and hardware services. The data owners can reduce the operational cost of installing and maintaining their own new software or hardware by moving their business applications and services into the clouds.There are different ways to deliver cloud services (Figure 1) as described in the following[2][3]. At the lowest level there is the possibility to run virtual machines on the infrastructure of a Cloud Service Provider (CSP). This is called *Infrastructure as a Service* (IaaS). One level higher there is the possibility to develop and deploy applications on the infrastructure of a CSP. This is called *Platform as a Service* (PaaS). On the highest level there are standardized applications which are delivered as a service. This is called *Software as a Service* (SaaS).

Software/Storage as a service (SaaS) - allows users to access software and applications hosted as a service. These services are deployed on clouds by cloud providers and can be accessed remotely across the internet. These services are available to use on pay per month method or pay per use method.

Platform as a service (PaaS) - allows user to use the infrastructure required for an application. Platform as a Service (PaaS) is a cloud delivery model which tries to make the development and deployment of applications a less complex and expensive task [2][4]. Platform as a Service (PaaS) is mainly designed for the developers to develop their own applications and to deploy it on PaaS environment.

Infrastructure as a service (IaaS) - allows users to use leased infrastructures and use it as if they are using their own hardware and software. Sometimes organizations have to buy infrastructure which they do not use frequently but they still have to have them. Infrastructure as a Service (IaaS) providers give them a solution and save them money by letting them rent their infrastructures. IaaS providers offer users their own separate instance of server, which is also called virtualization. A client can deploy own or vendor supplied virtual machines to run software in the cloud, and pays for the resources (CPU time, memory, storage and network usage) it consumes.

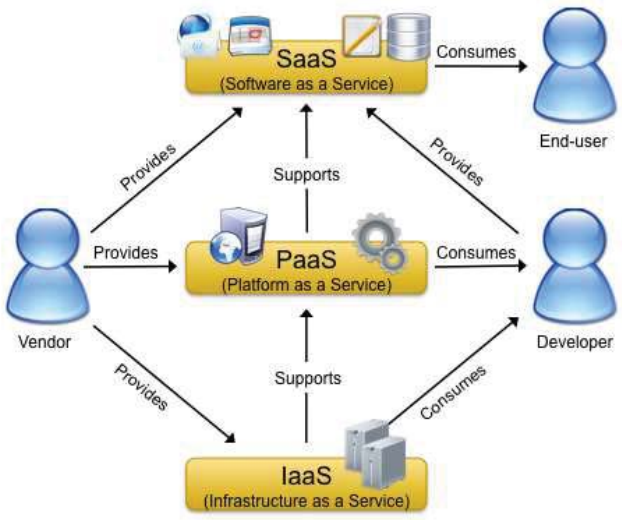

Fig.1. Service based models of cloud computing [5]

When using cloud services, businesses and organizations will have to trust a third party or technology for safekeeping their data. But the main problem lies in the fact that the data owner cannot always trust the cloud service provider due to the fact that cloud is located outside of the data owner's trusted domain. This fact alone creates several potential problems like sensitive data leakage or compromising integrity of data located in the clouds. The data stored in clouds are considered to be very confidential and sensitive by businesses and organizations and must not be disclosed to unauthorized third party. There are several methods that could be used to protect the sensitive data from unauthorized access. Usually, cloud service providers encrypt customer data before it is stored in the cloud[9]. But even the best encryption techniques can be compromised. Therefore, there is a need for a process that will check data integrity in order to make sure that no data modification has occurred. This will raise the confidence of cloud customers that their data is in safe and protected environment. In order to establish trust between the customer and cloud service provider, a third party auditors should be involved in order to check integrity of customer's data. It is common that third party auditors maintains a challenge/response pair for the data stored in the cloud. From time to time, it checks the data with the challenge/response method for the integrity of data. But in this scenario, it is assumed that the customer should trust third party auditor as well. Therefore, from the customer point of view third party auditors are similar to a cloud provider who has all the access to the customer's private data. Instead of helping the customer to maintain trust with the cloud provider, third party auditor itself could be the weakest link in this security chain and also could be the source of data integrity loss. It would be better for security reasons if this intermediate link (e.g., third party auditor) for integrity check can be avoided and the integrity check can be done by the customers themselves.
The solution provided in this paper, uses a hash algorithm utility which is implemented on the customer's side. The customer pre-computes the hash of the data content, stores the hash at the local secure hash repository and then sends the file to the cloud server for storage. When the customer wishes to verify the integrity of a data file, the customer takes that data contents and computes the hash and matches it with the pre-computed hash value. If there is any change in the data content, the two hash values will not match and the customer will know that the data integrity has been compromised. This method is less complicated than the third party auditor scheme and provides a simple way to check data integrity.

2 Related work

2.1 Recent attempts to address cloud security

Cloud computing is still considered as new and immature technology by both academic and IT world. Many researches have been done so far that have covered the aspects of security in cloud computing and many researches are still ongoing. Data integrity check by a third party auditing services is one of the newest topic in cloud computing research area. A customer always has reservations in trusting third party cloud service provider. One of the most basic question that they can ask themselves is whether they can integrate a third party auditing service in the existing system to check the integrity of the stored data in a cloud?

2.2 Public Auditing and data privacy preservation

In order to preserve integrity and privacy of the data stored in the cloud the scheme of the external audit service that checks the integrity of the data stored in a cloud can be employed [7]. In this method the public key based authenticator is employed and is positioned together with random masking technique to achieve their goal of efficient and privacy preserving auditing. This scheme guarantees that no data is stored locally for third party auditors and it does not create an extra overhead for the customer. Also, it is claimed that after integration, there will not be any further weaknesses in the existing security system. This makes the privacy preservation method a very efficient, secure and of high performance.

2.3 Digital Content Extraction and Privacy Preserving Audit

This scheme introduces a third party auditor which uses extraction protocol to ensure the integrity of customer data [8]. The technique mentioned here does not require from the customer to encrypt the data using some symmetric keys. This is because the keys can be lost over the time from the customer itself and the data is prone to get leaked. One of the big advantages of this solution to privacy preserving of data is that there is no need for customer to generate any secret keys or hash the data or encrypt the data. The customer can just call for the data and retrieve it as and when required.

3 Proposed method

The idea of relying on third party auditors for monitoring integrity of the data stored in cloud does not eliminate the "trust problem." This is more like moving the trust problem from one party to another party instead of solving the problem. If the customer cannot trust the third party cloud how can he/she trust third party auditor?

In many ways, a third party auditor (TPA) cannot be considered reliable. In order to get the integrity checked by auditors, customers have to reveal the data and the key used to decrypt it to TPA. However, all these third party auditor schemes assure that the customer data is not revealed to the auditors. But audit providers are still not completely trustworthy because their ultimate goal is money and not the security of customer data.

Using third party auditors makes the whole system even more complicated. Now the customer has to deal with the cloud provider as well as the auditor. Every time when the audit request is sent, the auditor will process the data and send it back to the customer. Communication between customer-auditor and auditor-cloud service provider will require more bandwidth and hence creates overhead.

Additionally, customers have to hire a third party auditor and pay extra money for auditing. It would be mere waste of money if the auditors cannot be trusted. Auditors have all the information of the customer data and can disclose it to an outsider intentionally or by mistake. This is an extra concern for customers that now they have to worry about how to keep data hidden from auditors. The actual problem of "trust" remains the same. In order to avoid third party auditors in this chain, this paper proposes that the integrity check of data stored in cloud can be checked at customer's side. This integrity check can be done by using cryptographic hash functions. In the following is presented and

describe our proposed method for data integrity check using hash function.

3.1 Integrity Check using Hash Function

For the data integrity check, we have to think about a simple solution that is feasible and easy to implement for a common user. The trust problem between TPA and customer can be solved, if users can check the integrity of data themselves instead of renting an auditing service to do the same. This can be achieved by hashing the data on user's side and storing the hash values in the local secure hash repository. Figure 2 presents the overview of the scheme.
This idea is based on the three properties of a hash function which eliminates the clash between two hash values and makes it possible to check integrity of data using hash. Based on our proposal, the customer first pre-computes the hash value of file, then sends the file to the cloud and the computed hash value is stored in the local secure repository. Whenever the customer wants to check integrity of the data, they retrieve the file from cloud and computes hash value of the file again and matches it with the pre-computed hash values stored at local hash repository. Since the hash value of a message is considered as its digital fingerprint, any changes in the original message will reflect in the result of its hash value. If the re-computed hash value matches with the pre-computed hash value then the file is intact and if it does not then the file was tampered and its integrity compromised.

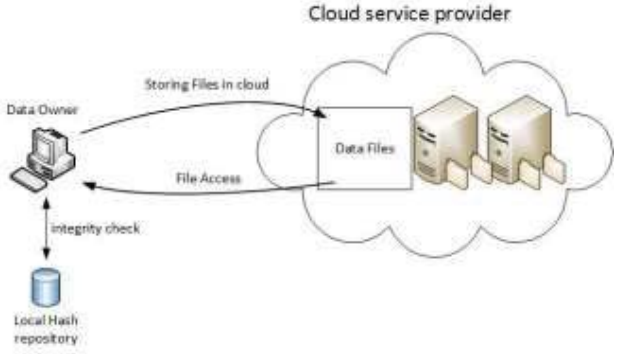

Fig.2. Data integrity check using hash functions

As this proposal is very cost effective, even a small scale company can implement this and avoid the trust problem they would have with a third party auditor. Using this integrity check method, once the trust is established with the cloud provider, the companies can stop integrity check, if not required.Customers can have an application (utility) which calculates these hash paths automatically and saves the files to their corresponding locations. Only customers can have access to this application.

Fig.3.Calculating hash of the file
In the Figure 3 we have presented a simple implementation of a utility that generates hash of the file. The tool used for this implementation is Netbeans IDE and the
programming language used is Java. The hash function used to generate hash value in this application is MD5. Algorithm with higher security like SHA can be used if needed. In case when unauthorized person makes the changes to the content of the file it inevitably changes also the integrity and the hash value of the file as well. In this case when we compare the hash values of the file we can see that hash has changed as presented in the figure 4.

Fig.4. Changed hash value of the file

4 Conclusion

Security of data and trust problem has always been a primary and challenging issue in cloud computing. This paper attempts to point out advantages and security concerns of cloud computing and focuses on avoiding third party auditors for data integrity check. Implementation of proposed utility, which computes hash values of files at the customer's side, can eliminate the need of third party auditors. The resultant hash values from this utility is stored at secure local hash repository. The data file can be retrieved back whenever needed and checked for any arguments among parties involved by re-computing and matching the hash result with the pre-computed hash value. This idea can be very effective on a small scale where customers initially want to test the cloud provider and want to establish trust and supplement already existing SLA[10][11]. This idea offers efficiency by minimizing human factor in data integrity checks and replacing them with technological solution which in terms will save money. In aspects of security, a lot more needs to be done to make cloud computing a prominent and reliable platform. The effort made in this paper is very basic and easy to use by customer

References

1. Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. Above the Clouds: A Berkeley View of Cloud Computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, Feb 2009.
2. NIST, "SP800-145: The NIST Definition of Cloud Computing."[Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. [Accessed: 21-Jan-2013].
3. L. M. Vaquero, L. Rodero-merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition," Computer Communication Review, vol. 39, no. 1, pp. 50–55, 2009.
4. Cloud Deployment Models – Private, Community, Public, Hybrid with Examples - By Basant Narayan Singh, October 1, 2011. http://www.techno- pulse.com/2011/10/cloud-deployment-private-public-example.html accessed 20 December 2012.
5. New Innovations Guide, www.newinnovationsguide.com/Cloud.html date accessed: 16/09/2013
6. Cloud Computing Security: How Secure is the Cloud? http://www.www.researchomatic.com/essay/Cloud-Computing-Security-How- Secure-Is-The-Cloud-92640.aspx accessed: 19/10/2013
7. Mehul A. Shah, Ram Swaminathan, Mary Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," IACR, 2008, http://eprint.iacr.org/2008/186.pdf
8. CATTEDDU, D. & HOGBEN, G. (2009): Cloud Computing: Benefits, risks and recommendations for information security; European Network and Information Security Agency (ENISA);
9. BSI (2008) BS ISO/IEC 27005:2008: Information Technology. Security Techniques. Information Security Risk Management. British Standards Institution.
10. Cecinio Silva Lacerda, "Service-level agreement (SLA), SearchiTChannel,

http://searchitchannel.techtarget.com/definition/service-level-agreement accessed: 24/08/2013
11. Service Level Agreement in the Data Center, Sun, 2011, http ://www.sun.com/blueprints/0402/sla.pdf
12. CHELLAPA, R. (1997) Intermediaries in Cloud-Computing: A new Computing Paradigm. Cluster: Electronic Commerce