ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

Dynamic Audit Services for Achieving Data Integrity in Clouds

Amala. U[1]

M.E, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, India[1]

ABSTRACT: **Cloud computing is a forthcoming revolution in information technology (IT) industry because of its performance, accessibility i.e., cloud storage enables users to access their data anywhere and at any time, pay per use service. Cloud computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructures, training new personnel or licensing new software. Cloud enables users to remotely store their data and enjoy on-demand high quality cloud applications without the burden of local storage and maintenance. Eventhough, the benefits are higher while storing data in cloud there may be chances for security risks such as missing or corruption of data. To ensure the data integrity and availability of the data we are using the auditing scheme. Thus, enabling public auditability for cloud storage is of critical importance so that the users can resort to the third party auditor (TPA) to check the integrity of outsourced data and can be worry-free. For this in our proposed scheme we are using the provable data possession (PDP) which is the cryptographic technique for verifying the integrity of data without retrieving it at an untrusted server which achieve zero knowledge property and the communication cost is also reduced here.**

Keywords: cloud computing, data integrity, cloud storage, provable data possession, audit service.

I. INTRODUCTION

Cloud computing has been envisioned as the next-generation architecture of IT enterprise, due to its long list of advantages in the IT history: on-demand service, location independent resource pooling, rapid resource elasticity and usage-based pricing. From users' perspective, including both individuals and enterprises, storing data remotely into the cloud in a flexible on-demand manner brings appealing benefits: relief of the burden of storage management, universal data access with dependent geographical locations and avoidance of capital expenditure on hardware, software and personnel maintenances, etc. While Cloud Computing makes these advantages more appealing than ever, it also brings new challenging security threats towards users' outsourced data. Since cloud service providers(CSP) are separate administrative entities, data outsourcing is actually relinquishing user's ultimate control over the fate of their data. As a result, the correctness of the data in the cloud is being put at risk due to the following reasons. First of all, although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity[4]. Examples of outages and security breaches of noteworthy cloud services appear from time to time [5], [7].

Secondly, there do exist various motivations for CSP to behave unfaithfully towards the cloud users regarding their outsourced data status. For examples, CSP might reclaim storage for monetary reasons by discarding data that has not been or is rarely accessed, or even hide data loss incidents to maintain a reputation [8], [9], [10]. In short, although outsourcing data into the cloud is economically attractive for the cost and complexity of long-term large-scale data storage, it does not offer any guarantee on data integrity and availability. This problem, if not properly addressed, may lead to unsuccessful storage in cloud.

As users no longer physically possess the storage of their data, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted because the data integrity and availability of data cannot be done without the local copy of data. In addition it is not a practical solution for data validation by downloading them due to the expensive transaction, especially for large files. More over the solutions to audit the correctness of the data in a cloud environment can be formidable and expensive for the cloud users. Thus, how to efficiently verify the correctness of outsourced cloud data without the local copy of data files becomes a big challenge for data storage in Cloud Computing. Note that downloading the data for its integrity verification is not practical solution due to the expensiveness in I/O cost and
transmitting the file across the network. Besides it is often insufficient to detect the data corruption only when accessing the data, as it does not give users correctness assurance for those unaccessed data and might be too late to recover the data loss or damage.

Considering the large size of the outsourced data and the user's constrained resource capability, the ability to audit the correctness of the data in a cloud environment can be formidable and expensive for the cloud users. Therefore, to fully ensure the data security and save the cloud users' computation resources, it is of critical importance to enable

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

audit scheme. Based on the audit result, TPA could release an audit report, which would help users to evaluate the risk of their data. But the introduction of the trusted third person should not lead to any new vulnerabilities such as leakage of information. The solution to all these problems will be provided in our proposed work. The paper is organized as follows. Section 2 describes the key technologies of cloud computing and Section 3 about audit schemes. In Section 4 we present our proposed work. Finally, we conclude the paper with Section 5.

II. KEY TECHNOLOGIES

A. Essential elements of Cloud Computing

The essential elements of cloud computing are as follows

On-demand self-service: A consumer with aninstantaneous need at a particular timeslot can avail computing resources in an automatic rashion without resorting to human interactions with providers of these resources.

Broad network access: These computing resources are delivered over the network and used by various client applications with heterogeneous platforms situated ata a consumer's site.

Resource pooling: A cloud service provider's computing resources are pooled together in an effort to serve multiple consumers using either the multi-tenancy or the virtualization model, with different physical and virtual resources dynamically assigned and reassigned
according to consumer demand. For example, consumers are not able to tell where their data is going to be stored in the cloud.

Rapid elasticity: For consumers, computing resources become immediate rather than persistent, there are no up-front commitment and contract as they can use them to scale up whenever they want, and release them once they finish to scale down. Moreover, resources provisioning appears to be infinite to them, the consumption can rapidly rise in order to meet peak requirement at any time.

B. Deployment models

There are three deployment models for cloud computing: public, private and hybrid cloud.

Public cloud: The public cloud is used by the general public cloud consumers and the cloud service providers has the full ownership of the public cloud with is own policy, value, and profit, costing and charging model. The services are made available to the users through the Internet. Eg: Amazon EC2, S3, Google AppEngine

Private cloud: The cloud infrastructure is operated solely within a single organization and managed by the organization or a third party regardless whether it is located premise or off premise. The motivation to setup a private cloud within organization has several aspects. First, to maximize and optimize the utilization of existing resources. Second, security concerns including data privacy and trust also make private cloud an option for many firms. Third, data transfer cost from local IT infrastructure to a Public cloud still rather considerable. Fourth, organizations always require full control over mission-critical activities that reside behind their firewalls.

Community cloud: Several organizations jointly construct and share the same cloud infrastructure as well as policies, requirements, values and concerns. The cloud community forms into a degree of economic scalability and democratic equilibrium. The cloud infrastructure could be hosted by a third party vendor or within one of the organizations in the community.

Hybrid cloud: It is a combination of both private and public cloud which makes data and application portability.

C. Service models

There are various kinds of service models available in cloud namely

1. Software as a Service(SaaS): SaaS provides various applications as service, that can be accessed through the network by users and deployed by people those who are using cloud.

   eg: Google mail, Google docs, SalesForce.com

2. Platform as a Service(PaaS): PaaS provides the platform to develop the application. The major difference between the Paas and SaaS is that SaaS only host the completed application but PaaS offers a development platform.
   eg: Google AppEngine

3. Infrastructure as a Service(IaaS): In IaaS, the cloud consumers can directly use IT infrastructures such as processing, storage, network and other fundamental computing resources[17].

   eg: Amazon EC2

D. Cloud DBMS Features

Efficiency: Given that the cloud computing pricing is structured in a way so that you pay for only what you use, the price increases linearly with the requisite storage, network bandwidth and compute power. Hence, if data analysis software product A requires an order of magnitude more compute units than data analysis software product B to perform the same task, then product A will cost an order of

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

magnitude more than B. Efficient software has a direct effect if the bottom line.

Fault tolerance: Fault tolerance in the context of analytical data workloads is measured differently than fault tolerance in the context of transactional workloads. For transactional workloads, a fault tolerant DBMS can recover from a failure without losing any data or updates from recently committed transactions and in the context of distributed databases, can successfully committed transactions and in the context of distributed databases can successfully commit transactions and make progress on a workload even in the face of worker node failure. For read-only queries in analytical workloads, there are no right transactions to commit, nor updates to lose upon node failure. Hence, a fault tolerant analytical DBMS is simple one that does not have to restart a query if one of the nodes involved in query processing.

Ability to run in heterogeneous environment: The performance of cloud compute nodes is often not consistent, with some nodes attaining orders of magnitude worse performance than other nodes. There are a variety of reasons why this could occur, ranging from hardware failure causing degraded performance on a node, to an instance being unable to access the second core on a dual core machine, to contention for non-virtualized resources. If the amount of work needed to execute a query is equally divided amongst the cloud compute nodes, then there is danger that the time to complete the query will be approximately equal to time for the slowest compute node to complete its assigned task. A node observing degraded performance would thus have a disproportionate affect of total query latency. A system designed to run in a heterogeneous environment would take appropriate measures to prevent this from occurring. A system designed to run in a heterogeneous environment would take appropriate measures to prevent this from occurring.

Ability to operate on encrypted data: Sensitive data may be encrypted before being uploaded to the cloud. In order to prevent unauthorized access to the sensitive data, any application running in the cloud should not have the ability to directly decrypt the data before accessing it. However, whipping entire tables or columns out of the cloud for decryption is bandwidth intensive. Hence, the ability of the data analysis system to operate directly on encrypted data so that a smaller amount of data needs to ultimately be shipped elsewhere to be decrypted could significantly improve performance.

E. Provable Data Possession(PDP)

We describe a framework for provable datapossession. This provides background for related work and for the specific description of our schemes. A PDP protocol checks that an outsourced storage site retains a file, which consists of a collection of n blocks. The client (data owner) pre-processes the file, generating a piece of metadata that is stored locally, transmits the file to the server, and may delete its local copy. The server stores the file and responds to challenges issued by the client. Storage at the server is in $\Omega(n)$ and storage at the client is in $O(1)$, conforming to our notion of an outsourced storage relationship. As part of pre-processing, the client may alter the file to be stored at the server. The client may expand the file or include additional metadata to be stored at the server. Before deleting its local copy of the file,the client may execute a data possession challenge to make sure the server has successfully stored the file. Clients may encrypt a file prior to out-sourcing the storage. For our purposes, encryption is an orthogonal issue; the "file" may consist of encrypted data and our metadata does not include encryption keys. The client requests that the server compute a function of the stored file, which it sends back to the client. Using its local metadata, the client verifies the response[6].

Threat model: The server S must answer challenges from the client C; failure to do so represents a data loss. However, the server is not trusted: Even though the file is totally or partially missing, the server may try to convince the client that it possesses the file. The server's motivation for misbehaviour can be diverse and includes reclaiming storage by discarding data that has not been or is rarely accessed (for monetary reasons), or hiding a data loss incident (due to management errors, hardware failure, compromise by outside or inside attacks etc). The goal of a

PDP scheme that achieves probabilistic proof of data possession is to detect server misbehaviour when the server has deleted a fraction of the file. Zero knowledge property prevents the leakage of verified data and the soundness property prevents the fraudulence of the prover.

III. AUDIT

Users choose cloud storage because of its convenient service provision. During the service process, user focus on the problem whether the data stored in the cloud is safe or not. But for the service provider, the main concern is the profits while providing convenient services. For both parties that focus on different aspects, the TPA operating as an independent and credible entity plays well in guaranteeing the trust relationship between the two parties. The TPA has professional authenticate knowledge and audit skill. To avoid or to overcome such security issues we are going for auditing scheme. There are two types of audit namely

Public audit: Here the trusted person or the TPA is used for the verification of data i.e., to ensure the data integrity. The TPA will be between user and Service Provider(SP). The user need not to be worried about their
data security.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

Private audit: In the case of private auditing the user is fully reponsible for verifying the integrity of their data. Here there is no involvement of the TPA.

In cloud computing there are two audit mechanism

1) **Internal audit** scheme inspects the internal behavior and processing service providers and try to avoid violation of SLA of understanding the service providers. When the internal audit starts, auditor and comprehensive understanding of risk in storage service and good measures about dealing with them in industry.

2) **External audit** provides end-to-end service quality metrics using SLA. The main purpose is to ensure data integrity in storage services. Through the APIs offered by the service providers, external audit can examine the data stored in the service providers by sampling and ensure their integrity, for example using the APIs provided by Amazon S3 to realize data access.

Phases of third party auditing:

There are three phases of third party auditing namely

1) Audit planning phase: First need to make sure the audit content, the audit details and so on and then determine the audit schedule. At the same time, it will also need to provide auditors the main purpose of the auditing and makes the work focused more clearly. Sufficient planning phase is helpful to maximize the efficiency of audit work.

2) Execute audit phase: It evaluates the superiority and insufficiency of current safety strategy during this phase. Auditors evaluate the matching degree between the methods used to solve the existing security threats and the internal and external security standard. In the audit report, it gives the suggestions for improving, with which the objects of audit can make suitable improvement. Verification process is mainly for the data integrity and the consistency of safety strategy and realization. It tries to find out the items lacking in those strategies or standards and make supplement in time.

3) Post audit phase: In this phase special organizations will deal with the corresponding problems and make improve on them according to the audit report given by the auditors.

IV. PROPOSED WORK

In this section we introduce an audit system architecture for outsourced data in clouds in Figure1. It consists of four main entities:

1) *Data owner(DO): who has data files to be stored in the cloud and relies on the cloud for data maintenance, can be an individual customer or an organization.*
2) *Cloud Storage Service Provider(CSP): who provides data storage service and has enough storage space to maintain clients data.*
3) *Third Party Auditor(TPA): a trusted perso who manage or monitor outsourced data under request of the data owner.*
4) *Authorized Application(AA): who have the right to access and manipulate stored data.*

The block diagram of audit system is explained as follows:

The data which the data owner wants to store in cloud first reaches the authorized application which will create digital signature and sends the data to the cloud storage.

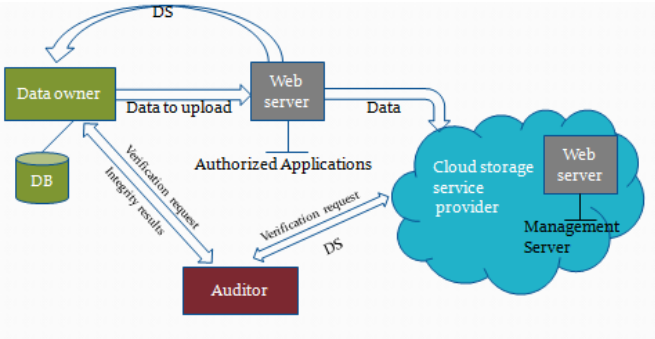

Fig. 1 Audit System Architecture

If the user needs to verify data means the verification request should be send to third party auditor(TPA), the TPA will retrieve the digital signature from the database and will send the verification request to the management server. The management server in turn will generate the digital signature for the data stored in the cloud and it will send only that digital signature instead of the whole data to the TPA. The TPA will decrypt the digital signature and compares the message digest for verifying correctness of data.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

The main aim of this scheme is to ensure the data integrity in cloud storage. Data integrity in the sense, the data stored by the data owner should be same without any lose or modification of the data in the receiver side. For auditing we are using provable data possession(PDP) allows a client that has stored data at an untrusted server to verify that server possess the original data without retrieving it. To enhance the data security we are using the digital signatures. Before the user store their data in the cloud the fixed length message digest will be created using MD5 algorithm regardless of the size of the data. Then the message digest is encrypted using RSA algorithm which we call it as digital signature. The digital signature is stored in the data base and data is send to cloud storage. At later time, if the user needs to verify the data means the verification request should be send to third party auditor(TPA), the TPA will retrieve the digital signature from the database and will send the verification request to the management server. The management server in turn will generate the digital signature for the data stored in the cloud stored in the cloud and it will send only that digital signature and compares the message digest. If both message digests are same then we can say that data is safe without any lose modification. If they are not same in the sense the TPA will send an alert mail to the data owner.

In the case of implementation our work consists of three important modules namely admin module, user module and auditor module. The admin module is responsible to maintain the whole security system. It maintains the users list which consists of the userid, name, gender, address and mailed. The admin is capable to add, edit or delete the list. Same like that the auditors list is also maintained by the admin. Next in the transaction details which will contain the logged in date and time and if profile of user or auditor is modified that details and the uploaded and downloaded file name, time and date are maintained. In the case of the user module the user details can be viewed. For that the user can click into add then choose a file by clicking browse from the system and then submit, the file will be uploaded successfully. Before the uploading process the authorized application will generate the message digest and encrypt it and then send it to the cloud storage. During the downloading process the user can select the file from the list what they have uploaded and then click download such that the file will be downloaded and can be saved in any of the places where the user wants. Here the user is also able to maintain the transaction details.

Next module is the auditor module which is used to do the integrity check of the data. During the verification the management server is responsible to generate message digest of the currently stored data in cloud and send that message digest to the auditor. Then the auditor will get the digital signature from the authorized application and decrypt it and compares the message digest. If both are same then the file is safe without any loss or modification or else there is some modification or loss in data what the user has stored and an alert mail will be send to the users mail account.

V. CONCLUSION

There are various existing auditing schemes available to ensure integrity. But in most of the scheme there is a leakage of information from the verifier side because verification is done with local copy of data, or it lead to extra burden to the data owner in the case of private auditing. In our proposed scheme the zero knowledge property is achieved such that the third party auditor who is responsible to verify the users' data will not be having any knowledge about data. So, our proposed scheme can be more secured when compared to the existing conventional schemes. Next, for verification only signatures are send instead of the whole data such that communication cost can also be reduced in our scheme. We believe all these advantages of the proposed schemes will shed light on economies of scale of Cloud Computing.

Acknowledgment

I take this opportunity to extend my heartiest thanks to our HOD, Dr. D. Thilagavathy, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu for her constant support during this work. I extend my sincere gratitude to my guide Professor E. Saravanakumar, Adhiyamaan College of Engineering for his valuable suggestions and exquisite guidance with encouragement ever since the commencement of the work. Finally, I would like to thank all the department faculty members, my parents and my friends for their forbearance and understanding that helped me to complete the work.

REFERENCES

[1] Yan Zhu, Hongxin Hue, Gail-Joon Ahn, Stephen S. Yau, "Efficient audit service outsourcing for data integrity in clouds", Elsevier Journal Of Systems and Software, vol.85, pp.1083-1095, 2012.

[2] Bhagyaraj Gowrigolla, Sathyalakshmi Sivaji, M. Roberts Masillamani "Design and Auditing of Cloud Computing Security", IEEE International Conference on Information and Automation for Sustainability, 2010.

[3] Cong Wang and Kui Ren, Wenjing Lou, Jin Li "Toward Publicly Auditable Secure Cloud Data Storage Services", IEEE International Journal On Networks, 2010.

[4] Cloud Security Alliance, "Top Threats to cloud computing", http://www.cloud securityalliance.org., 2010.

[5] Q. Wang, C.Wang, K. Ren, W.Lou and J.Li, "Enabling public auditability and data dynamics for storage security in cloud computing", IEEE Transactions on Parallel and Distributed Systems, vol. 22, no.5, pp.847-459, 2011.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 1, Issue 10, December 2012

[6] G. Ateniese, R. Burns, R. Curtola, J. Herring, L.Kisser and D. Song, "Provable data possession at untrusted stores", in Proc. Of CCS'07, pp.598-609.
[7] Daniel J. Abadi "Data Management in the Cloud: Limitations and Opportunities**",** IEEE International Conference on Data Engineering, 2009.
[8] Jianfeng Yang & Zhibin Chen, "Cloud Computing Research and Security Issues", IEEE International Conference on Computational Intelligence & Software Engineering, 2010.
[9] Ling Li, Lin Xu, Jing Li and Changchun Zhang **"**Study on the Third-party Audit in Cloud Storage Service**"** IEEE International Conference on Cloud & Service Computing, 2011.
[10] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson "Provable Data Possession at untrusted stores", in the ACM, 2007
[11] B. Priyadharshini, P. Parvathi "Data Integrity in Cloud Storage", IEEE International Conference On Advances In Engineering, Science And Management, 2012.
[12] Richard Chow, Philippe Golle, Markus Jakobsson, Ryusuke Masuoka, Jesus Molina "Controlling Data in the Cloud: Outsourcing Computation without outsourcing control", in ACM International Journal, 2009.
[13] Shuai Han, Jianchuan Xing "Ensuring data storage security through a novel third party auditor scheme in cloud computing", in the International Conference on Cloud Computing & Intelligence Systems, 2011.
[14] C. Wang, A. Wang, K. Ren and W. Lou, "Towards secure and dependable storage services in cloud computing", IEEE Transactions of Service Computing, 2011.
[15] Siani Pearson and Azzedine Benameur "Privacy, Security and Trust Issues Arising from Cloud Computing", in the 2nd IEEE International Conference on Cloud Computing Technology and Science
[16] Sravan Kumar R, Ashutosh Saxena "Data Integrity Proofs in Cloud Storage", IEEE International Conference on Communication Systems & Networks, 2011.
[17] Tharam Dillon, Chen Wu and Elizabeth Chang, "Cloud Computing: Issues and Challenges", IEEE International Conference on Advanced Information Networking and Applications, 2010.
[18] M.R.Tribhuwan, V.A.Bhuyar and Shabana Pirzade "Ensuring Data Storage Security in Cloud computing through Two-way Handshake based on Token Management", in the International Conference on Advances in Recent Technologies in Communication and Computing, 2010.
[19] Yashaswi Singh, Farah Kandah, Weiyi Zhang "A Secured Cost-effective Multi-Cloud Storage in Cloud Computing", IEEE International Conference on Computer Communication, 2011.

Biography

Miss. Amala. U has completed B.E in Computer Science and Engineering from ARJ College of Engineering and Technology, affiliated to Anna University, in the year 2011. Presently she is pursuing her M.E degree from Adhiyamaan College of Engineering (Autonomous), affiliated to Anna University, Chennai, Tamilnadu, India. Her area of interest includes cloud computing, network security.