Dynamic Data Storage Auditing Services in Cloud Computing

Rakhi Bhardwaj, Vikas Maral

Abstract— Cloud based data storage systems have many complexities regarding critical/confidential/sensitive data of client. The trust required on Cloud storage is so far had been limited by users. The role of the paper is to grow confidence in Users towards Cloud based data storage. The paper handles key questions of the User about how data is uploaded on Cloud, maintained on cloud so that there is no data loss; data is available to only authorized User(s) as per Client/User requirement and advanced concepts like data recovery on disaster is applied.

In this paper we look at the various current researches being done to solve these issues, the current trends in securing, ensuring privacy and availability of these data on cloud storage services.

General Terms

Cloud computing, Security and Reliability.

Keywords-Cloud Storage, Data Availability, data storage auditing, data owner auditing, Privacy, and Security.

I. INTRODUCTION

Cloud computing is a promising computing model that enables convenient and on-demand network access to a shared pool of computing resources. Cloud storage is an important service of cloud computing, which allows data owners to move data from their local computing systems to the Cloud. More and more data owners start choosing to host their data in the Cloud. [10]

The local management of such huge amount of data is problematic and costly due to the requirements of high storage capacity and qualified personnel. Therefore, Storage-as-a-Service offered by cloud service providers (CSPs) emerged as a solution to mitigate the burden of large local data storage and reduce the maintenance cost by means of outsourcing data storage.[1,3][10]

Cloud Storage Providers like Microsoft with Sky Drive, Google Documents, and Drop Box etc have successfully dropped rates of storage available on internet. They promise availability of the data from different systems/locations/networks. Basic security like User based authentication access of data and maintaining offline data to the client's machine is also supported.[10]

Given all the above features still User confidence on the Cloud storage still hampers the usage of the Cloud based Storage. The companies are investing heavily on the servers with massive storage devices divided geographically and interconnected with high bandwidth and speed networks. The utilization, if analyzed is still low in terms of Confidential/secure data hosted by clients.

The paper presents Dynamic Auditing method which can be provided to client for growing user's confidence on Cloud storage in terms of realizing the cost-effective benefits of cloud-based storage solutions.

Manuscript received April, 2013.

Rakhi Bhardwaj, Computer Engineering. K.J's College Of Engineering. And Management Research, Pune, India.

Vikas Maral, Computer Engineering, K.J's College Of Engineering And Management Research, Pune, India.

II. DYNAMIC AUDITING

The User always has following general doubts while using Cloud Storage:

1. Will my data be available when I will need it? Many CSPs although offering good space might be running internal implementation where-in storage is recovered for the data not used for long enough time.[1]
2. Will my data be uploaded through secure channel? When user is uploading public data like internet saved news/articles/images/files, he is less concerned about the channel being used to upload the data. But when user is uploading private information like confidential documents, finance related data; secured information etc, the question of channel being used is primary one.
3. The data should not be available to non-authorized user. No other Cloud user should get the access to private data.
4. How much cost effective it will be?
5. How much fast data recovery will user get for lost data?

As the cloud technology growing on it is difficult for user to map the pros and cons of this technology, basically there are two main areas of cloud computing where lots of challenges are arriving with benefits: cloud storage and cloud backup/recovery.

While both these services in cloud are used for to improve manageability, and can be integrated easily to backup most all aspects of a businesses' data requirements, from server to laptop.

As the IT people are having higher-performance, more scalable and cheaper storage business benefits from cloud. But following are the challenges meet by the end user if the proper cloud services are not used by the client/CSPS (cloud service provider).[10]

III. CURRENT DATA STORAGE CHALLENGES IN CLOUD

A cloud storage service provider should base its pricing on how much storage capacity a business has used, how much bandwidth was used to access its data, and the value-added services performed in the cloud such as security.

Unfortunately, all the CSPS are not functioning in equal manners'. Data storage paradigm in "Cloud" brings about many challenging design issues because of which the overall performance of the system get affected. Most of the biggest concerns with cloud data storage are:

- Data integrity verification at un-trusted servers-

For example, the storage service provider, which experiences Byzantine failures occasionally, may decide to hide the data errors from the clients for the benefit of their own. What is more serious is that for saving money and storage space the service provider might neglect to keep or

deliberately delete rarely accessed data files which belong to an ordinary client. Consider the large size of the outsourced electronic data and the client's constrained resource capability, the core of the problem can be generalized as how can the client find an efficient way to perform periodical integrity verifications without the local copy of data files.

- Data accessed by unauthorized users:

The confidentiality feature can be guaranteed by the Owner via encrypting the data before outsourcing to remote servers. For verifying data integrity over cloud servers, researchers have proposed provable data possession technique to validate the intactness of data stored on remote sites.

- Location Independent Services*:*

The very characteristics of the cloud computing services are the ability to provide services to their clients irrespective of the location of the provider. Services cannot be restricted to a particular location but may be requested from any dynamic location as per the choices of the customer.

- Infrastructure and security:

The infrastructure that is used for these services should be secured appropriately to avoid any potential security threats and should cover the life time of component.

- Data recovery /Backup:

For data recovery in cloud the user must concern the security as well as the bandwidth issue in consideration.

IV. PERFORMANCE CRITERIA IN STORAGE AUDITING PROTOCOL

Data storage auditing is a very resource demanding operation in terms of computational resource, memory space, and communication cost. There are three performances criteria in the design of storage auditing protocols:

- ***Low storage overhead***: The additional storage used for auditing should be as small as possible on both the Auditor and the cloud server.
- ***Low communication cost***: The communication cost required by the auditing protocol should be as low as possible.
- ***Low computational complexity***: The computational complexity for storage auditing should be low, especially on the Auditor

V. DATA STORAGE AUDITING MODEL

In this section, we describe the system model and threat model of data storage auditing protocol in cloud computing. Some models are discussed here:

- ***Data Owner Auditing:***

In recent years, with the development of distributed storage systems and online storage systems [3], the data storage auditing problem becomes even more significant and many protocols have been proposed: e.g., Remote Integrity Checking (RIC) protocols [1,3], Proof of Retrievability (POR) protocols [3] and Provable Data Possession (PDP) protocols . However, most of the existing protocols only allowed data owners to check the integrity of their remote stored data. We denote this type of auditing protocols as the Data Owner Auditing.

Challenge-response Protocol

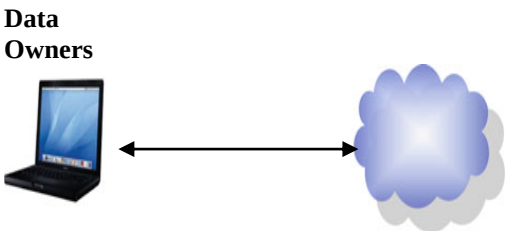

Fig .1: System model of the data owner auditing.

***Third Party Auditing*:**

For the Third Party Auditing, the system model contains three types of entities: data owners, the cloud server and the third party auditor, as shown in Figure 2. During the system initialization, data owners compute the metadata of their data and negotiate the cryptographic keys with the third party auditor and the cloud server. Each auditing query is conducted via a challenge-response auditing protocol, which contains three phases: Challenge, Proof and Verification.

When the third party auditor wants to check the correctness of data owners' data stored on the cloud server, it generates and sends a challenge to the cloud server. The cloud server generates a proof of data storage and sends it back to the third party auditor. Then, the third party auditor runs the verification to check the correctness of the proof from the cloud server and extracts the result on this audit query.

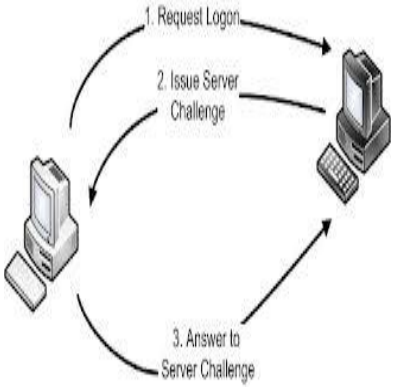

Fig.2: Auditing Query via the Challenge-response Protocol

VI. EXISTING ALGORITHM

In the existing cloud architecture listed below are the some of the methods available for the dynamic data storage on remote machine to access remote data.

- Algorithms for Audit System

Firstly, we present the definition of two algorithms for the tag generation process as follows:

ü Key Gen: takes a security parameter _ as input, and returns a public/secret key pair.

T agGen (sk, F): takes as inputs the secret key sk and a file F, and returns the triple, where denotes the secret used to generate the verification tags, verification parameters u and index-hash and _ denotes the set of tags is a set of public.

- Fragment Structure and Secure Tags

To maximize the storage efficiency and audit performance, general fragment structure is introduced into our audit system for outsourced storage. An instance for this framework which is used in this scheme is showed in Figure 3: an outsourced file F is split into n blocks, and each block mi is split into s sectors. The fragment framework consists of n block-tag pair, where _i is a signature tag of block mi generated by some secrets. Finally, these block-tag pairs are stored in CSP and the encryption of the secrets _ (called as PVP) is in

TTP.[7][12]

- Index-Hash Table

Simple index-hash table (IHT) to record the changes of file blocks as well as generate the hash value of block in the verification process. The structure of our index-hash table is similar to that of file block allocation table in file systems. Generally, the index-hash table _consists of serial number block number, version number, random integer, and so on.

Different from the common index table, we must assure that all records in this kind of table differ from one another to prevent the forgery of data blocks and tags. In addition to record data changes, each record i in table is used to generate a unique Hash value, which in turn is used for the construction of signature tag i by the secret key sk. This kind of relationship must be cryptographic secure, and we can make use of it to design our verification protocol depicted and the checking algorithm.

- **Data Owner**

In this architecture, we consider a data storage service involving four entities: data owner (DO) the client (data owner) uses the secret key sk to pre-process a file, which consists of a collection of n blocks, generates a set of public verification parameters (PVP) and index-hash table (IHT) that are stored in TPA, transmits the file and some verification tags to CSP, and may delete its local copy.

VII. PROPOSED ALGORITHM

As we have discussed multiple available cloud storage services in above section here we would like to propose the service model which allow data owner to get benefits from CSPs and maintain trust worthy relation between them.

For that three factors are in consideration:

i. Allow the data owner to outsource their sensitive data to a CSP, and perform full block-level dynamic operations on the outsourced data, i.e., block modification, insertion, deletion, and append.
ii. Authentication and authorization.
iii. Build mutual trust between the data owner and CSPs.

Main components as illustrated in Fig. 1: (i) a data owner that can be an organization generating sensitive data to be stored in the cloud and made available for controlled external use; (ii) a CSP who manages cloud servers and provides paid storage space on its infrastructure to store the owner's files and make them available for authorized users; (iii) authorized users – a set of owner's clients who have the right to access the remote data; and (iv) a trusted third party (TTP),

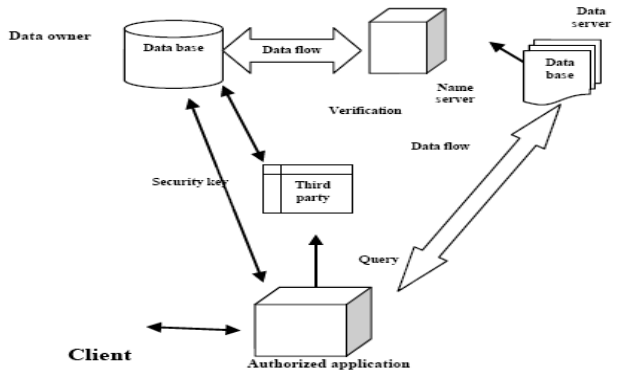

Fig. 3: Cloud computing data storage system model.

Client

In Fig. 3, the relations between different systems components are represented by double-sided arrows, For example, the data owner, the authorized users, and the CSP trust the TTP. On the other hand, the data owner and the authorized users have mutual distrust relations with the CSP. Thus, the TTP is used to enable indirect mutual trust between these three components.[1][16]

TTP is the third trusted party auditor who will audit the data of data owner or client so that it will let off the burden of management of data of data owner. TTP eliminates the involvement of the client through the auditing of whether his data stored in the cloud are indeed intact, which can be important in achieving economies of scale for Cloud Computing. The released audit report would not only help owners to evaluate the risk of their subscribed cloud data services, but also be beneficial for the cloud service provider to improve their cloud based service platform [1][5].

However, any possible leakage of data towards the TTP must be prevented to keep the outsourced data private. The TTP and the CSP are always online, while the owner is intermittently online. The authorized users are able to access the data file from the CSP even when the owner is offline.

Data access by data owner to modified / updated data result in false assumptions as the CSPs are un-trusted. For this reason security requirement is essential.[7][8][12] It checks the integrity of data and also maintaining consistency at cloud data storage for CSP and Client.

For deletion of record

- Client sends a request to CSP to delete the record.
- CSP ask client for authentication just like login page
- Client sends a message like and as to CSP for Deletion and denotes for File name.
- CSP will delete the file.

For updating records:

Client Side

1. Client request to access a file from CSP.
3. Client authenticates CSP by his password
5. Client decrypts the file by applying decryption algorithm [12].
6. If client modify the file he will send file to CSP and TTP with a message like *Md* as (*F*',$,*M*) and *F* ' here M denotes for modification F ' for encrypted file, *Md* for message digest file [12] and $ for signature.
11. If file is same as previous one, drop this packet and move to step 1 or step 13.
12. Else ask CSP to follow step 11 again.
13. Exit ' F

CSP Side

2. CSP ask client for authentication just like login page.
4. Verify password if correct send a file that he wants to access. Else move to step 2.
7. CSP check the signature for authenticity and compute the message digest to find encrypted file which is com-pare with encrypted file of another message.
8. If correct it will change previous file with this one and move to step12.
9. Else ask the client to follow the step 8.
10. CSP sends a same message to client after addition of his signature.

Fig.4: Table for User Operations

From updating of record and insertion of record algorithm, TTP already have encrypted file. So it will check this encrypted file with the encrypted file of CSP. If there is mismatch in file than it send the error report to data owner.

Depending upon these algorithms now days so many cloud

storage providers are available where people are increasingly using cloud storage. Whether you've got important documents, photos, music or other files that need to be shared across more than one device, using a cloud storage option is often the easiest way to do it.

With so many cloud services along with the recent introduction of Google Drive, you may be wondering which one could be right for you. Check out our list below to get a general summary of each popular cloud service and their major features.

Cloud Storage Providers	Features	Specs
Google Drive	**FREE STORAGE**	5 GB
	Approximate annual price for 20 GB	$29.88
	APPROXIMATE ANNUAL PRICE FOR 100 GB	$59.88
	Max file size allowed	10 GB
	Desktop apps	WINDOWS, MAC
	Mobile apps	ANDROID, IOS COMING SOON
Drop box	**FREE STORAGE**	2 GB
	Approximate annual price for 20 GB	$99
	APPROXIMATE ANNUAL PRICE FOR 100 GB	$199
	Max file size allowed	300 MB VIA BROWSER, UNLIMITED VIA DESKTOP
	Desktop apps	WINDOWS, MAC, LINUX
	Mobile apps	ANDROID, IOS, BLACKBERRY
Apple I Cloud	**FREE STORAGE**	5 GB
	Approximate annual price for 20 GB	$40
	APPROXIMATE ANNUAL PRICE FOR 100 GB	$100
	Max file size allowed	25 MB FOR FREE USERS, 250 MB FOR PAID USERS
	Desktop apps	WINDOWS, MAC
	Mobile apps	IOS ONLY
Microsoft SkyDrive	**FREE STORAGE**	7 GB
	Approximate annual price for 20 GB	$10
	APPROXIMATE ANNUAL PRICE FOR 100 GB	$50
	Max file size allowed	2 GB
	Desktop apps	WINDOWS, MAC
	Desktop apps	WINDOWS, MAC
	Mobile apps	IOS, WINDOWS PHONE
Box	**FREE STORAGE**	5 GB
	Approximate annual price for 20 GB	$120
	APPROXIMATE ANNUAL PRICE FOR 100 GB	$240
	Max file size allowed	25 MB
	Desktop apps	WINDOWS, MAC
	Mobile apps	ANDROID, IOS, BLACKBERRY

Fig.5: Table of Storage Provider comparison

VIII. CONCLUSION

We presented a construction of dynamic audit services for un-trusted and outsourced storage. We also presented an efficient method for periodic sampling audit to minimize the computation costs of third party auditors and storage service providers with the survey of current cloud storage providers. Our experiments showed that our solution has a small, constant amount of overhead, which minimizes computation and communication costs.

ACKNOWLEDGMENT

We sincerely thank the Staff members and Colleagues who have directly or indirectly contributed for completion of this Paper. We are Grateful to our Institute and Management for lending support without which this would have not been possible. Last but not least we will extend

our gratitude to our family members.

REFERENCES

[1] Ayad Barsoum ,Anwar Hasan, Ontario, Canada "Enabling Dynamic Data and Indirect Mutual Trust for Cloud Computing Storage Systems" Digital Object Indentifier 0.1109/TPDS.2012.337 1045-9219/12/$31.00 © 2012 IEEE

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS '07, 2007, pp. 598–609.

[3] F. Seb´e, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient remote data possession checking in critical information infrastructures," IEEE Trans. on Knowl. And Data Eng., vol. 20, no. 8, 2008.

[4] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, 2008, pp. 1–10.

[5] C. Erway, A. K¨upc¸ ¨ u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in Proceedings of the 16th ACM Conference on Computer and Communications Security, 2009, pp. 213–222.

[6] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in Proceedings of the 14th European Conference on Research in Computer Security, 2009, pp. 355–370.

[7] Francesc Sebe´ , Josep Domingo-Ferrer, Antoni Martı´nez-Balleste, Yves Deswarte, Jean-Jacques Quisquater 'Efficient Remote Data Possession Checking in Critical Information Infrastructures', 1041-4347/08/$25.00 _ 2008 IEEE Published by the IEEE Computer Society

[8] Kan Yang · Xiaohua Jia" Data storage auditing service in cloud computing:challenges, methods and opportunities", WorldWide Web (2012) 15:409–428 DOI 10.1007/s11280-011-0138-0

[9] Ayad F. Barsoum and M. Anwar Hasan "On Verifying Dynamic Multiple Data Copies over Cloud Servers"

[10] Amazon elastic compute cloud (Amazon EC2), http://aws.amazon.com/ec2/.

[11] Reza Curtmola, Osama Khan, Randal Burns, Giuseppe Ateniese" MR-PDP: Multiple-Replica Provable Data Possession", 1063-6927/08 $25.00 © 2008 IEEE DOI 10.1109/ICDCS.2008.68

[12] Raghul Mukundan, Sanjay Madria, Mark Linderman" Replicated Data Integrity Verification in Cloud", IEEEcase number 88ABW-2012-6360

[13] Qian Wang, Student Member, IEEE, Cong Wang, Student Member, IEEE, Kui Ren, Member, IEEE,Wenjing Lou, Senior Member, IEEE, and Jin Li'Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing" 1939-1374/12/$31.00 _ 2012 IEEE Published by the IEEE Computer Society).

[14] Smitha Sundareswaran, Anna C. Squicciarini, Member, IEEE, and Dan Lin" Ensuring Distributed Accountability For Data Sharing in the Cloud", 1545-5971/12/$31.00 _ 2012 IEEE Published by the IEEE Computer Society

[15] Kan Yang, Student Member, IEEE, Xiaohua Jia, Senior Member, IEEE" An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing ",Digital Object Identifier 10.1109/TPDS.2012.278 1045-9219/12/$31.00 © 2012 IEEE