System for Secure Storage and Auditing of Cloud Data

Pooja Jaywant Patil

ME Computer, MMCOE Pune, Maharashtra, India

Abstract: ***Cloud computing is compilation of existing technique and technologies, packaged within a new infrastructure paradigm that offers improved scalability, elasticity, business agility, faster startup time, reduced management costs, and just-in-time availability of resources. Cloud computing is an Internet-based computing and use of computer technology which provides access to resources as a service such as storage, network, server and processors etc. By moving data into the cloud offers great ease to users. In general, data center in cloud holds information that users have stored on their computers. The security concern arises because the outsourced data is used by the user. Cloud computing is the growing field now a days. The platform provides help to reduce the cost as well as make the effective utilization of the hardware as well as software. Data storage is the main most desirable aspect of the cloud computing, but it comes with some security challenges too. The end users store their data on cloud server are always in worry that either their data stored is secure or not? As the data stored is large enough so users can not check its integrity periodically. Sometimes cloud service providers may behave unhonestly and delete customers data Or they fail to make changes on the data which is updated by the users frequently. So to overcome these challenges the Trusted Third Party Auditor plays the vital role on behalf of customers. As they assure to customers that the data hosted on the server is secure. TPA provides more easier and affordable way for users to their storage correctness in cloud, which also helpful for the cloud service providers to improve their cloud based service platform. In other way we can say auditing scheme play a significant role in establishment of secure cloud platform in users mind and increase the cloud economy, where users accesses the risk and apply their trust in the cloud to store data more precisely.***

Keywords: Cloud, Third party auditor, Privacy, Data storage, Public auditing.

1. Introduction

Cloud computing has been envisioned as the next generation information technology (IT) architecture for enterprises, due to its long list of unprecedented advantages in the IT history: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk. As a disruptive technology with profound implications, cloud computing is transforming the very nature of how businesses use information technology. One fundamental aspect of this paradigm shifting is that data are being centralized or outsourced to the cloud. From users' perspective, including both individuals and IT enterprises, storing data remotely to the cloud in a flexible on-demand manner brings appealing benefits: relief of the burden for storage management, universal data access with location independence, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc., [3].

While cloud computing makes these advantages more appealing than ever, it also brings new and challenging security threats toward 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. Examples of outages and security breaches of noteworthy cloud services appear from time to time. Second, there do exist various motivations for CSP to behave unfaithfully toward the cloud users regarding their outsourced data status. For examples, CSP might reclaim storage for monetary reasons by discarding data that have not been or are rarely accessed, or even hide data loss incidents to maintain a reputation. In short, although outsourcing data to the cloud is economically attractive for long-term large-scale storage, it does not immediately offer any guarantee on data integrity and availability. This problem, if not properly addressed, may impede the success of cloud architecture.

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. In particular, simply downloading all the data for its integrity verification is not a practical solution due to the expensiveness in I/O and transmission cost 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 tasks of auditing the data correctness in a cloud environment can be important and expensive for the cloud users. Moreover, the overhead of using cloud storage should be minimized as much as possible, such that a user does not need to perform too many operations to use the data (in additional to retrieving the data). In particular, users may not want to go through the complexity in verifying the data integrity. Besides, there may be more than one user accesses the same cloud storage, say in an enterprise setting. For easier management, it is desirable that cloud only entertains verification request from a single designated party.

To fully ensure the data integrity and save the cloud users' computation resources as well as online burden, it is of critical importance to enable public auditing service for cloud data storage, so that users can ask to an independent third-party auditor (TPA) to audit the outsourced data when needed. The TPA, who has expertise and capabilities that users do not, can periodically check the integrity of all the data stored in the cloud on behalf of the users, which provides a much more easier and affordable way for the users to ensure their storage correctness in the cloud. Moreover, in addition to help users to evaluate the risk of their subscribed cloud data services, the audit result from TPA would also be beneficial for the cloud service providers to improve their cloud-based service platform, and even serve for independent arbitration purposes. In a word, enabling public auditing services will play an important role for this nascent cloud economy to become fully established; where users will need ways to assess risk and gain trust in the cloud.

Recently, the notion of public auditability has been proposed in the context of ensuring remotely stored data integrity under different system and security models. Public auditability allows an external party, in addition to the user himself, to verify the correctness of remotely stored data. However, most of these schemes do not consider the privacy protection of users' data against external auditors. Indeed, they may potentially reveal user's data to auditors. This severe drawback greatly affects the security of these protocols in cloud computing. From the perspective of protecting data privacy, the users, who own the data and rely on TPA just for the storage security of their data, do not want this auditing process introducing new vulnerabilities of unauthorized information leakage toward their data security [3]. Moreover, there are legal regulations, such as the US Health Insurance Portability and Accountability Act (HIPAA), further demanding the outsourced data not to be leaked to external parties. Simply exploiting data encryption before outsourcing could be one way to mitigate this privacy concern of data auditing, but it could also be an overkill when employed in the case of unencrypted/public cloud data (e.g.outsourced libraries and scientific data sets), due to the unnecessary processing burden for cloud users. Besides, encryption does not completely solve the problem of protecting data privacy against third-party auditing but just reduces it to the complex key management domain. Unauthorized data leakage still remains possible due to the potential exposure of decryption keys.

Therefore, how to enable a privacy-preserving third party auditing protocol, independent to data encryption, is the problem we are going to tackle in the proposed system. It is among the first few ones to support privacy-preserving public auditing in cloud computing, with a focus on data storage. Besides, with the prevalence of cloud computing, a foreseeable increase of auditing tasks from different users may be delegated to TPA. As the individual auditing of these growing tasks can be tedious and cumbersome, a natural demand is then how to enable the TPA to efficiently perform multiple auditing tasks in a batch manner, i.e., simultaneously

A. Motivation and challenges

1) Efficient Data Distribution: It provides an efficient method for distributing the data file on distributed server more efficiently by removing extra load on server due to transferred and stored data in the Cloud.
2) Efficient Data Recovery: This system also provide an efficient data recovery method in order to retrieval of the lost data in Cloud with help of file distribution technique.
3) The proposed system inherits the support of data dynamics, and supports public verifiability and privacy against third-party auditor.
4) To authorize the CS to respond to the audit delegated to TPA's, the user can issue a certificate on TPA's public key, and all audits from the TPA are authenticated against such a certificate
5) Security analysis of the proposed system, which shows that it is secure against the untrusted server and private against third party verifiers.

2. Literature Review

a) Privacy preserving public auditing for data storage security in cloud computing

In this paper, authors have proposed a privacy-preserving public auditing system for data storage security in cloud computing. It will utilize the homomorphic linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task, but also alleviates the users fear of their outsourced data leakage.

Considering TPA may concurrently handle multiple audit sessions from different users for their outsourced data files. Also support privacy-preserving public auditing protocol into a multiuser setting, where the TPA can perform multiple auditing tasks in a batch manner for better efficiency[1]. They haveproved the security and justify the performance of proposed schemes through concrete experiments and comparisons with the state-of-the-art.

b) Public auditability for storage security

In this paper, authors have studied the problem of ensuring the integrity of data storage in Cloud Computing. It considers the task of allowing a third party auditor, to verify the integrity of the dynamic data stored in the cloud. This paper achieves both public auditability and dynamic data operations. It first identifies the difficulties and potential security problems of direct extensions with fully dynamic data updates from prior works and then shows how to construct an elegant verification scheme for the seamless integration of these two silent features in protocol design[3].

c) Privacy preserving data integrity checking

Authors of this paper have proposed protocol, that allow a third-party auditor to periodically verify the data stored by a service and assist in returning the data intact to the customer. The protocols are privacy-preserving i.e. it never reveals the data contents to the auditor. This solution removes the burden of verification from the customer, alleviates both the customers and storage services fear of data leakage, and provides a method for independent arbitration of data retention contracts. The solution provides storage service accountability through independent, third-party auditing and arbitration. The protocol have three important operations : Initialization, Audi and Extraction. This protocol primarily focuses on audit and extraction. For audits,the auditor interacts with the service to check that the stored data is intact. For extraction, the auditor interacts with the service and customer to check that the data is intact and return it to the customer [8].

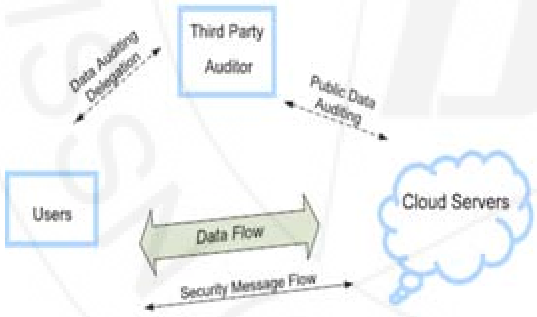
d) Dynamic provable data possesion

As outsourcing the data and services and sharing the resources through network becomes important aspect, there is problem of proving the integrity of stored data at untrusted server and which will receive great attention. PDP is earlier scheme of dynamic PDP where user preprocesses the data and then it will be sent to cloud server for storage. While doing this, user will some amount of metadata with them. This stored data will be checked by user by asking server to prove that stored data has not been changed or deleted without retrieving the whole data.for auditing. However, this scheme is not considering dynamic data,It applies only to static data. Hencedynamic PDP scheme covers limitation of normal PDP scheme by extending it with support to provable updates of stored data. It uses new version of authenticated dictionaries based on rank information

3. System Architecture

As shown in below system architecture, proposed system consist of three major components:

- **User**: User can be individual entity or organization who will store their data in cloud and rely on cloud for data computation.
- **Cloud Service Provider**: CSP has significant resources and expertise in building and managing distributed cloud storage servers.
- **Third Party Auditor**: TPA has expertise and capabilities that users may not have. TPA is trusted to assess and expose risk of cloud storage services on behalf of users request



In the cloud paradigm, by putting the large data files on the remote servers, the clients can be relieved of the burden of storage and computation. As clients no longer possess their data locally, it is of critical importance for the clients to ensure that their data are being correctly stored and maintained. That is, clients should be equipped with certain security means so that they can periodically verify the correctness of the remote data even without the existence of local copies. In case that clients do not necessarily have the time, feasibility or resources to monitor their data, they can delegate the monitoring task to a trusted TPA. In the proposed scheme, we only consider verification schemes with public auditability: TPA in possession of the public key can act as a verifier. We assume that TPA is unbiased while the server is untrusted. For application purposes, the clients may interact with the cloud servers via CSP to access or retrieve their pre-stored data. More importantly, in practical scenarios, the client may frequently perform block-level operations on the data files. The most general forms of these operations we consider are modification, insertion and deletion

4. Proposed System

There are 5 modules in proposed systemas given below:

1. File Distribution
2. Metadata Key Generation
3. Public verifiability
4. Privacy against Third Party Verifiers
5. Data Dynamics
   - Block Insertion
   - Block Modification
   - Block Deletion

A. File distribution module

It will distribute users data file on different servers with help of Rotated reed solomon technique[6].Whenever user want to upload any file on cloud then it sends that file to TPA. In which (m+k; k) R. Reed-Solomon erasure correcting codes [6] is used to create (k) redundancy parity vectors from (m) data vectors in such a way that , the original (m) data vectors can be reconstructed from any (m) out of the (m + k) data and parity vectors. By placing each of the (m + k) vectors on a different server, the original data file can survive the failure of any (k) of the (m+k) servers without any data loss. For support of efficient sequential I/O to the original file, this file layout is systematic, i.e., the unmodified (m) data file vectors together with (k) parity vectors is distributed across (m + k) different servers.

B. Metadata Key Generation

Let the user U wishes to store the file F on cloud server. First file will be divided into n file blocks. This n will be selected by user only. Each of the n file blocks will be having m bits in them. Now each block of file F is encrypted using RSA algorithm.All encrypted blocks are concatenated together. This concatenated blocks will be appended to the file F and now file F with concatenated encrypted file blocks will be stored on cloud server.

C. Public Verifiability

As, users are storing their important data on cloud, they are always in worry whether there data is intact or not. Hence data integrity checking becomes challenging issue. As sometimes cloud service providers may behave unhonestly and delete customers data Or they fail to make changes on the data which updated by the users frequently. So to overcome these challenges the Trusted Third Party Auditor plays the vital role on behalf of customers. As they assure to customers that the data hosted on the server is secure.

D. Privacy against third party verifier

While checking integrity of users data on cloud, third party auditor should not learn any knowledge about the data content stored on the cloud server. Hence privacy should be maintained while publically auditing the data.

E. Data dynamics

Data dynamics means after clients store their data at the remote server, they can dynamically update their data at later times. At the block level, the main operations are block insertion, block modification and block deletion.

1) Block Insertion: Inserting new block after some specified positions in the data file F.
2) Block Deletion: Deleting the specified block.
3) Block Modification: Replacement of specified blocks with new ones.

5. Conclusion

Cloud computing is a internet based computing which enables sharing of services and many users place their data in the cloud. However, users are always in worry whether their data is secure or not. So, data integrity is challenging issue. To overcome this, initially users were downloading the complete data for auditing which is actually not feasible in terms of transmission cost and communication overhead. Hence we proposed public auditing where one dedicated party audit the data from cloudon behalf of users request. while auditing the data, one important aspect is that our system is maintaining privacy such that TPA would not learn users data.

By using Reed Solomon technique, users file will be distributed among different server to achieve data recovery in order to retrieve the lost data. Also it removes extra load on one server . Our system also supports data dynamics where user can perform modify, delete and insertoperation on data which is stored on cloud. In future , datadynamics willshowwhat modification is done in the client file by server to the client and both current and previous versions of the data file and corresponding metadata will be audited on demand.

6. Acknowledgment

We take this opportunity to thank Mr. Ram Joshi sir for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this paper. We are also thankful to all the staff members of the Department of Computer Engineering for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

References

[1] C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-Preserving Public Auditing for Storage Security in Cloud Computing”, Proc. IEEE INFOCOM ’10, Mar. 2010

[2] K.D. Bowers, A. Juels, and A. Oprea,”Proofs of Retrievability:Theory and Implementation”, Proc. ACM Conf. Computer and Comm. Security (CCS ’07), pp. 584-597,2009

[3] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling PublicAuditability and Data Dynamics for Storage Security in Cloud Computing”, IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 5, pp. 847-859,Mar 2010

[4] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song,”Provable Data Possession at Untrusted Stores”, Proc. 14th ACM Conf. Computer and Comm. Security (CCS ’07), pp. 598-609,2007

[5] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia,” Dynamic Provable Data Possession”, Proc. ACM Conf. Computer and Comm. Security (CCS ’09), pp. 213-222,2009

[6] Osama Khan, Randal Burns,James Plank, William Pierce,Cheng Huang Rethinking Erasure Codes for Cloud File Systems:Minimizing I/O for Recovery and Degraded Reads

[7] A.L. Ferrara, M. Green, S. Hohenberger, and M. Pedersen, “Practical Short Signature Batch Verification,” Proc. Cryptographers Track at the RSA Conf. 2009 on Topics in Cryptology (CT-RSA), pp. 309-324, 2009

[8] Z Hao, S Zhong, and N Yu, “A privacy-preserving remote data integrity checking protocol with datadynamics and public verifiability,” IEEE Transactions on Knowledge and DataEngineering, vol. 99, 2011.

[9] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik,” Scalable and Efficient Provable Data Possession”, Proc. Int’l Conf. Security and Privacy in Comm. Networks (SecureComm ’08), pp. 1-10,2008

[10] C. Wang, K. Ren, W. Lou, and J. Li, “Towards Publicly Auditable Secure Cloud Data Storage Services", IEEE Trans. Service Computing, vol. 5, no. 2, 220-232,July/Aug. 2010