# Certificate of Excellence

is hereby presented to

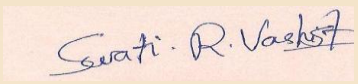
Ajeet Singh Rajput

for publication of his/her/their research paper titled

Privacy Preservation and Data Integrity Verification in Cloud Computing

in Volume II and Issue VIII of International Journal of Computational Linguistics and Natural Language Processing.

ISSN 2279-0756



**Editor,IJCLNLP**

**24th September 2013**

# Privacy Preservation and Data Integrity Verification in Cloud Computing

Ajeet Singh Rajput
Computer Science and Engineering Department
S.D.B.C.T, Mhow Road,Indore,(M.P), India
ajeetsinghrajput@gmail.com
M.E.(CSE), S.D.B.C.T, Indore

***Abstract*—Cloud computing is the long dreamed vision of computing as a utility, where users can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. By data outsourcing, users can be relieved from the burden of local data storage and maintenance. Thus, enabling public auditability for cloud data storage security is of critical importance so that users can resort to an external audit party to check the integrity of outsourced data when needed. To securely introduce an effective third party auditor (TPA), the following two fundamental requirements have to be met: 1) TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user.**

**Keywords**: Privacy verification, Data Integrity verification, Third Party Auditing

## I. Introduction

Cloud computing is an Internet-based computing, where shared resources, software and information, are provided to computers and devices on-demand. It provides people the way to share distributed resources and services that belong to different organization. Since cloud computing uses distributed resources in open environment, thus it is important to provide the security and trust to share the data for developing cloud computing applications. In this paper we assess how can cloud providers earn their customers' trust and provide the security, privacy and reliability. Cloud computing is revolutionizing the way companies are implementing their information systems. Cloud computing has elevated IT to newer limits by offering the market environment data storage and capacity with flexible scalable computing processing power to match elastic demand and supply, whilst reducing capital expenditure [1]. As a result, Cloud adoption is spreading rapidly and represents a new opportunity that companies should not ignore given its profound impact. Due to its perceived open nature, Cloud Computing raises strong security, privacy and trust concerns, namely: How data safely stored and handled by Cloud providers? How data privacy being managed adequately? Are Cloud providers adhering to laws and regulations? How is business disruption or outage kept to its minimum? Are Cloud providers sufficiently protected against cyber-attacks? Cloud computing share distributed resources via the network in the open environment, thus it makes security problems important for us to develop the cloud computing application. Cloud computing is a convenient, on-demand model for network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction . The cloud element of cloud-computing derives from a metaphor used for the Internet, from the way it is often depicted in computer network diagrams. Conceptually it refers to a model of scalable, real-time, internet-based information technology services and resources, satisfying the computing needs of users, without the users incurring the costs of maintaining the underlying infrastructure. Examples in the private sector involve providing common business applications online, which are accessed from a web browser, with software and data stored on the "cloud" provider's servers

## II. Literature survey on Data Auditing

**Third Party Data Auditing**

Data auditing is a one-time or periodic event to evaluate security, data integrity , computational accuracy and privacy preserving.On behalf of the cloud client , a third party auditor (TPA) have been used to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA eliminates the direct involvement of the client through the auditing of whether his data stored in the cloud.

**Demerits of Data Auditing Techniques**

1. Localization of the error may be found with very difficulty.
2. There is a tradeoff between Data dynamics, privacy preservation and public verifiability.
3. None of the scheme support multiple TPA for cross check and cross authenticate the data integrity verification and privacy preservation.
4. Some of the schemes do not support batch auditing.

## III. Problem Formulation and Approach Used

Cloud computing is not secure by nature. Security in the Cloud is often intangible and less visible, which inevitably creates a false sense of security and anxiety about what is actually secured and controlled. The off-premises computing paradigm that comes with cloud computing has incurred great concerns on the security of data, especially the integrity and confidentiality of data, as cloud service providers may have complete control on the computing infrastructure that underpins the services. Accordingly, the various security challenges as shown above, related to Cloud computing are worth of a deeper attention and can relate to many different aspects.

### *A. Information security policy*

Cloud users typically have no control over the Cloud resources used and there is an inherent risk of data exposure to third parties on the Cloud or the Cloud provider itself. Some policies are

a) Whether there exists an Information security policy, which is approved by the management, published and communicated as appropriate to all employees.

b) Whether it states the management commitment and set out the organizational approach to managing information security.

c) Whether the Security policy has an owner, who is responsible for its maintenance and review according to a defined review process.

d) Whether the process ensures that a review takes place in response to any changes affecting the basis of the original assessment, example: significant security incidents, new vulnerabilities or changes to organizational or technical infrastructure.

### *B. Information security infrastructure*

The cloud computing architecture requires the adoption of identity and access management measures. Some issues:

a)Whether there is a management forum to ensure there is a clear direction and visible management support for security initiatives within the organization.

b) Whether there is a cross-functional forum of management representatives from relevant parts of the organization to coordinate the implementation of information security controls.

c) Whether responsibilities for the protection of individual assets and for carrying out specific security processes were clearly defined.

d) Whether there is a management authorization process in place for any new information processing facility. This should include all new facilities such as hardware and software.

e) A specific individual may be identified to co-ordinate in-house knowledge and experiences to ensure consistency, and provide help in security decision making.

f) Whether appropriate contacts with law enforcement authorities, regulatory bodies, information service providers and telecommunication operators were maintained to ensure that appropriate action can be quickly taken and advice obtained, in the event of a security incident.

### *C. Security of third party access*

To access the third party some challenges:

a) Whether risks from third party access are identified and appropriate security controls implemented.

b) Whether the types of accesses are identified, classified and reasons for access are justified.

c) Whether security risks with third party contractors working onsite was identified and appropriate controls are implemented.

d) Whether there is a formal contract containing, or referring to, all the security requirements to ensure compliance with the organization's security policies and standards.

e) Whether security requirements are addressed in the contract with the third party, when the organization has outsourced the management and control of all or some of its information systems, networks and/ or desktop environments.

The contract should address how the legal requirements are to be met, how the security of the organization's assets are maintained and tested, and the right of audit, physical security issues and how the availability of the services is to be maintained in the event of disasters The tasks arrived to the data center are scheduled for execution using energy-aware "TPA" scheduler. This "TPA" scheduler tends to group the workload on a minimum possible amount of computing servers. The servers left idle are put into a sleep mode (DNS) while on the under-loaded servers the supply voltage is reduced

## IV. Proposed Work

In this paper, we utilize the public key based homomorphic authenticator and uniquely integrate it with random mask technique to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind. To support efficient Handling of multiple auditing tasks, we further explore the technique of bilinear aggregate signature to extend our main result into a multi-user setting, where TPA can perform multiple auditing tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient. We also show how to extent our main scheme to support batch auditing for TPA upon delegations from multi-users. In proposed work, we propose an effective and flexible distributed scheme with explicit dynamic data support to ensure the correctness of users' data in the cloud with more security. We rely on erasure correcting code in the file distribution preparation to provide redundancies and guarantee the data dependability. This construction drastically reduces the communication and storage overhead as compared to the traditional replication-based file distribution techniques. By utilizing the homomorphic token with distributed verification of erasure-coded data, our scheme achieves the storage correctness insurance as well as data error localization: whenever data corruption has been detected during the storage correctness verification, our scheme can almost guarantee the simultaneous localization of data errors, i.e., the identification of the misbehaving server(s).

1. Compared to many of its predecessors, which only provide binary results about the storage state across the distributed servers, the challenge-response protocol in our work further provides the localization of data error.

2. Unlike most prior works for ensuring remote data integrity,

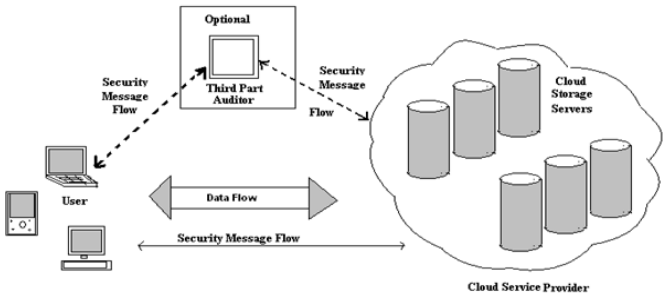



Fig. 1. Introduction of TPA

the new scheme supports secure and efficient dynamic operations on data blocks, including: update, delete and append.

3. Extensive security and performance analysis shows that the proposed scheme is highly efficient and resilient against Byzantine failure, malicious data modification attack, and even server colluding attacks.

### *A. Steps Involved in Designing Above Architecture*

**Step 1: Client Application**

In this step, the client sends the query to the server. Based on the query the server sends the corresponding file to the client. Before this process, the client authorization step is involved. In the server side; it checks the client name and its password for security process. If it is satisfied and then received the queries form the client and search the corresponding files in the database. Finally, find that file and send to the client. If the server finds the intruder means, it set the alternative Path to those intruders.

**Step 2: System Modules**

Representative network architecture for cloud data storage is illustrated in Figure 1. Three different network entities can be identified as follows: User: A users who have data to be stored in the cloud and rely on the cloud for data computation, consist of both individual consumers and organizations. Cloud Service Provider (CSP): A CSP, who has significant resources and expertise in building and managing distributed cloud storage servers, owns and operates live Cloud Computing systems,.

Third Party Auditor (TPA): An optional TPA, who has expertise and capabilities that users may not have, is Trusted to assess and expose risk of cloud storage services on behalf of the users upon request.

**Step 3: Cloud data storage Module**

Cloud data storage, a user stores his data in a cipher text form through a public key infrastructure mechanism with hash function with additional padding bit through a CSP into a set of cloud servers, which are running in a simultaneous, the user interacts with the cloud servers via CSP to access or retrieve his data. In some cases, the user may need to perform block level operations on his data.. users should be equipped with security means so that they can make continuous correctness assurance of their stored data even without the existence of local copies. In case that users do not necessarily have the time, feasibility or resources to monitor their data, they can delegate the tasks to an optional trusted TPA of their respective choices. In our model, we assume that the point-to-point communication channels between each cloud server and the user is authenticated and reliable, which can be achieved in practice with little overhead.

**Step 4: Cloud Authentication Servers**

The Authentication Server (AS) functions as any AS would with a few additional behaviors added to the typical client-authentication protocol. The first addition is the sending of the client authentication information to the masquerading router. The AS in this model also functions as a ticketing authority, controlling permissions on the application network. The other optional function that should be supported by the AS is the updating of client lists, causing a reduction in authentication time or even the removal of the client as a valid client depending upon the request.

**Step 5 :Unauthorized data modification and corruption module**

One of the key issues is to effectively detect any unauthorized data modification and corruption, possibly due to server compromise and/or random Byzantine failures. Besides, in the distributed case when such inconsistencies are successfully detected, to find which server the data error lies in is also of great significance.

**Step 6:Adversary Modules**

Security threats faced by cloud data storage can come from two different sources. On the one hand, a CSP can be self-interested, untrusted and possibly malicious. Not only does it desire to move data that has not been or is rarely accessed to a lower tier of storage than agreed for monetary reasons, but it may also attempt to hide a data loss incident due to management errors, Byzantine failures and so on. On the other hand, there may also exist an economically motivated adversary, who has the capability to compromise a number of cloud data storage servers in different time intervals and subsequently is able to modify or delete users' data while remaining undetected by CSPs for a certain period. Specifically, we consider two types of adversary with different levels of capability in this paper:

**Weak Adversary**: The adversary is interested in corrupting the user's data files stored on individual servers. Once a server is comprised, an adversary can pollute the original data files by modifying or introducing its own fraudulent data to prevent the original data from being retrieved by the user.

**Strong Adversary:** This is the worst case scenario, in which we assume that the adversary can compromise all the storage servers so that he can intentionally modify the data files as long as they are internally consistent. In fact, this is equivalent to the case where all servers are colluding together to hide a data loss or corruption incident.

### *B. Algorithm*

A public auditing scheme consists of four algorithms (KeyGen, SigGen, GenProof, VerifyProof).

**KeyGen:** key generation algorithm that is run by the user to setup the scheme

**SigGen:** used by the user to generate verification metadata, which may consist of MAC, signatures or other information used for auditing

**GenProof:** run by the cloud server to generate a proof of data storage correctness

**VerifyProof:** run by the TPA to audit the proof from the cloud server.

## V. MODULES

We have following modules.

### *A. Privacy-Preserving Public Auditing Module :*

Homomorphic authenticators are unforgeable verification metadata generated from individual data blocks, which can be securely aggregated in such a way to assure an auditor that a linear combination of data blocks is correctly computed by

verifying only the aggregated authenticator. Overview to achieve privacy-preserving public auditing, we propose to uniquely integrate the homomorphic authenticator with random mask technique. In our protocol, the linear combination of sampled blocks in the server's response is masked with randomness generated by a pseudo random function (PRF). The proposed scheme is as follows: 1. Setup Phase 2. Audit Phase

### *B. Batch Auditing Module:*

With the establishment of privacy-preserving public auditing in Cloud Computing, TPA may concurrently handle multiple auditing delegations upon different users' requests. The individual auditing of these tasks for TPA can be tedious and very inefficient. Batch auditing not only allows TPA to perform the multiple auditing tasks simultaneously, but also greatly reduces the computation cost on the TPA side.

### *C. Data Dynamics Module:*

Hence, supporting data dynamics for privacy-preserving public risk auditing is also of paramount importance. Now we show how our main scheme can be adapted to build upon the existing work to support data dynamics, including block level operations of modification, deletion and insertion. We can adopt this technique in our design to achieve privacy-preserving public risk auditing with support of data dynamics.

## VI. Conclusion

We propose a privacy-preserving public auditing system for data storage security in Cloud Computing. We 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, we further extend our privacy- preserving public auditing protocol into a multi-user setting, where the TPA can perform multiple auditing tasks in a batch manner for better efficiency. Extensive analysis shows that our schemes are provably secure and highly efficient.. We leave the full-fledged implementation of the mechanism on commercial public cloud as an important future extension, which is expected to robustly cope with very large scale data and thus encourage users to adopt cloud storage services more confidently

## References


[1] Zhifeng Xiao and Yang Xiao. " Security and Privacy in Cloud Computing". *IEEE COMMUNICATIONS SURVEYS and TUTORIALS, ACCEPTED FOR PUBLICATION* in December 2012.

[2] F. Sebe, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater. "Efficient Remote Data Possession Checking in Critical Information Infrastructures". *IEEE Trans. Knowledge and Data Eng*,vol. 20, no. 8, pages. 1034-1038, Aug. 2008.

[3] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou. "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing,". *Proc. 14th European Conf. Research in Computer Security (ESORICS)*, IEEE, 2009.

[4] C. Wang, Q. Wang, K. Ren, and W. Lou "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing". In *Proc. IEEE INFOCOM*, IEEE, 2010.

[5] K.D. Bowers, A. Juels, and A. Oprea "HAIL: A high-availability and integrity layer for cloud storage," *," Proc. 16th ACM conference On Computer and communications security,*, pages 187-198, 2009.

[6] Zhuo Hao, Sheng Zhong and Nenghai Yu. "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability," IEEE Transection on Knowledge and Data Engineering, VOL. 23, NO. 9, September 2011.