Volume 3, Issue 2, February 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Enhanced Data Security in Cloud Computing with Third Party Auditor

Bhavna Makhija* *Department of Computer Engineering Hasmukh Goswami College of Engineering,Vahelal,Gujarat*

VinitKumar Gupta *Department of Computer Engineering Hasmukh Goswami College of Engineering,Vahelal,Gujarat*

Indrajit Rajput *Department of Computer Engineering Hasmukh Goswami College of Engineering,Vahelal,Gujarat*

Abstract— ***Cloud computing is environment which enables convenient, efficient, on-demand 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. Cloud is kind of centralized database where many organizations/clients store their data, retrieve data and possibly modify data. Cloud is a model where user is provided services by CSP(Cloud Service Provider) on pay per use base. Means here Client has to pay for what he is using or being served. Data stored and retrieved in such a way may not be fully trustworthy so here concept of TPA(Third Party Auditor) is used. TPA makes task of client easy by verifying integrity of data stored on behalf of client. In cloud, there is support for data dynamics means clients can insert, delete or can update data so there should be security mechanism which ensure integrity for the same. Here TPA can not only see the data but he can access data or can modify also so there should be some security mechanism against this.***

Keywords*— *Third Party Auditor, Software as a Service, Cloud Service Provider

I. INTRODUCTION

Cloud computing is a model which provides a wide range of applications under different topologies and every topology derives some new specialized protocols. In this research paper, we will present an introduction to a cloud computing that is expected to be adopted by governments, manufacturers and academicians in the very near future. It directly affects the company, government and convenience to the small user. It is the technology of building a robust data security between CSP and User. This promising technology is literally called Cloud Data Security. In this research, an introduction to the technology of Cloud Computing, TPA, data security and security algorithm of different papers will be presented.

II. THEORETICAL BASELINE

Cloud computing is a Kind of network where user can use services provided by Service provider on pay per use bases. It is a research area which provides a wide range of applications under different topologies where every topology computing that is expected to be adopted by government, manufacturers and academicians in the near future. Here user uses services virtually from CSP. Cloud Computing is the technology of building a robust data security between CSP and user. This technology is literally called Cloud Data Security. In this research, an introduction to the technology of Cloud Computing, TPA, data security and security mechanisms of different existing papers with their merits and demerits will be presented. In this point, the cloud involves Cloud Visual model, Cloud component, TPA..

A. Cloud Computing

Cloud computing is a model which enables convenient, efficient, on-demand 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. In this section we have divided cloud computing into further part i.e. Service models, Cloud Component for more understanding about cloud.

B. Types of Service Models in Cloud

Cloud computing providers offer their services according to three fundamental models[9] Infrastructure as a service (IaaS), and software as a service(SaaS) where IaaS is the most basic and each higher model abstracts from the details of the lower models.

1) *Software as a Service (SaaS):* The capability provided to the consumer is to use the provider‘s applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

2) *Platform as a Service (PaaS):* The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers,

operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

3) *Infrastructure as a Service (IaaS):* The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

C. Deployment Models in Cloud computing

There are four types of cloud available in cloud computing i.e. private cloud, public cloud, hybrid cloud and community cloud as shown in Fig 1. These deployment models describe who owns, manages and is responsible for the services. The detail types of different type cloud are as follows:

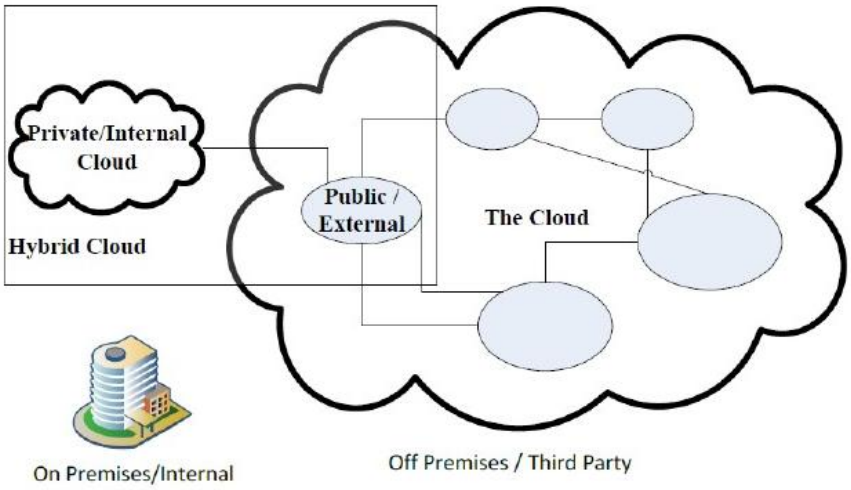

Fig. 1 Types of Cloud

1) *Private cloud:* The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers(e.g. business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.
2) *Public cloud:* The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider [8].
3) .*Community cloud:* The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations) [8]. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.
4) *Hybrid cloud:* Hybrid cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that will be unique entities, but bound together by standardized technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds) [8].

D. Third Party Auditor

Third Party Auditor is kind of inspector. There are two categories: private auditability and public auditability. Although private auditability can achieve higher scheme efficiency, public auditability allows anyone, not just the client (data owner), to challenge the cloud server for the correctness of data storage while keeping no private information. To let off the burden of management of data of the data owner, TPA will audit the data of client. It eliminates the involvement of the client by auditing that 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 help owners to evaluate the risk of their subscribed cloud data services, and it will also be beneficial to the cloud service provider to improve their cloud based service platform [6]. Hence TPA will help data owner to make sure that his data are safe in the cloud and management of data will be easy and less burdening to data owner.

III. LITERATURE SURVEY

Different factors such as integrity of data, data dynamics and data privacy affects The performance of a number of approaches in cloud data storage. Each and every approach has merits and demerits which make them suitable for different applications. In this chapter we will discuss different approaches which are already carried out for cloud data security.

A. Review of existing approaches

A simple approach like message authentication codes (MACs) can be used to protect the data integrity.Data owners

will initially locally maintain a small amount of MACs for the data files which are to be outsourced. The data owner can verify the integrity by recalculating the MAC of the received data file when he/she wants to retrieve data and will compare it to the local precomputed value. Even though this method allows data owners to verify the correctness of the received data from the cloud, but if the data file is large, MACs cannot be employed. A hash tree can be employed for large data files, in which leaves contains hashes of data blocks and internal contains hashes of their children of the tree. To authenticate his received data the data owner has to store the root hash of the tree. But it does not give any assurance about the correctness of other outsourced data. So to perform this thing for data owner TPA can be used.

Various mechanisms are proposed on how to use the TPA so that it can relieve the burden of data owner for local data storage and maintenance, it also eliminates their physical control of storage dependability and security, which traditionally has been expected by both individuals and enterprises with high service-level requirements. This kind of audit service not only helps save data owners'computation resources but also provides a transparent yet cost- effective method for data owners to gain trust in the cloud. The presence of TPA eliminates the involvement of the client by auditing whether his data stored in the cloud are indeed intact, which can be important in achieving economies of scale for Cloud Computing.

Though this method states how to save the computational resource and cost of storage of owner's data buthow to trust on TPA that is not calculated. If TPA modifies data or deletes some dataand if it becomes intrusive and pass information of data owner to unauthorized user than how owner know about this problem is not solved. Thus, new approaches are required to solve the above problem.

The author Abhishek Mohta and R. Sahu [11] have given algorithm which ensures data integrity and dynamic data operations. They have used encryption and message digest to ensure data integrity. Although encryption ensures that data is not leaked while transfer and message digest gives identity of client who has send data. They have designed algorithm for data manipulation, insertion of record and record deletion. Insertion and manipulation algorithms inserts and manipulate data efficiently but in data deletion we can't identify the person who have deleted record, how and when means if any one deletes record then this algorithm can no longer work.

In that case we can use indexing scheme i.e. if we trace every record by index, that when and which user is accessing record then if user tries to delete record then we can identify him as we have traced him by index.

The author Ateniese et al. [6] are the first who have considered the public adaptability in their defined―provable data possession‖ (PDP) method which ensures possession of data files on untrusted storages. For auditing outsourced data their technique utilizes the RSA-based homomorphic authenticators and suggests to randomly sample a few blocks of the file. However,in their scheme the public auditability demands the linear combination of sampled blocks which exposed to the external auditor. When used directly, their protocol is not provably privacy preserving, and thus may leak user data information to the auditor.

The author Cong Wang et al. [7] used the public key based homomorphic authenticator and to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind, it uniquely integrate it with random mask technique.For efficiently handling multiple auditing tasks,the technique of bilinear aggregate signature can be explored to extend the main result into a multi-user setting, where TPA can perform multiple auditing tasks simultaneously.

A keyed hash function hk(F) is used in Proof of retrievability (POR) scheme. The verifier, pre-computes the cryptographic hash of F using hk(F) before archiving the data file F in the cloud storage, and stores this hash as well as the secret key K. The verifier releases the secret key K to the cloud archive to check the integrity of the file F and asks it to compute and return the value of hk(F). The verifier can check for the integrity of the file F for multiple times by storing multiple hash values for different keys, each one being an independent proof.

Although this scheme is very simple and easily implementable the main drawback of this scheme is that it requires higher resource costs for the implementation. Verifier has to store as many keys as the number of checks it wants to perform as well as the hash value of the data file F with each hash key. Computation of the hash value for even a moderately large data files can be computationally burdensome for some clients (PDAs, mobile phones, etc.). Each invocation of the protocol at archive requires the archive to process the entire file F. This processing can be computationally burdensome for the archive even for a lightweight operation like hashing. Furthermore, it requires the prover to read the entire file F - a significant overhead for an archive whose intended load is only an occasional read per file, where every file to be tested frequently [3].

The author Ari Juels and Burton S. Kaliski Jr proposed a scheme "Proof of retrievability" for large files using "sentinels"[3]. In this scheme, only a single key can be used irrespective of the size of the file or the number of files unlike in the key-hash approach scheme in which many number of keys are used.

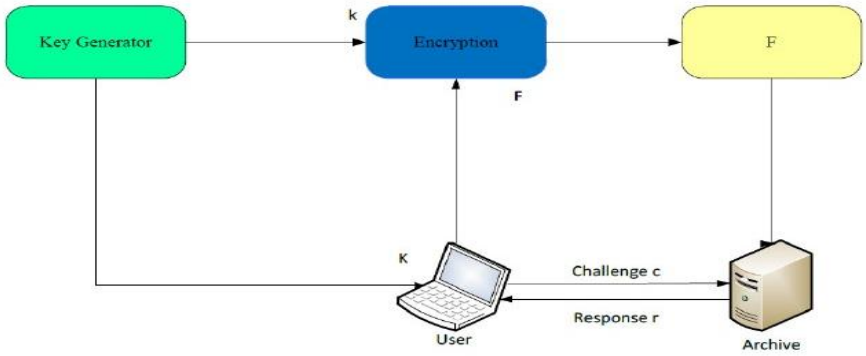

Fig. 2 Schematic view of a proof of retrievability based on inserting random sentinels in the data file F [3].

The archive needs to access only a small portion of the file F unlike in the key-hash scheme which required the archive to process the entire file F for each protocol verification. This small portion of the file F is in fact independent of the length of F. The schematic view of this approach is shown in Fig 2.

In their scheme, Ari Juels and Burton S. Kaliski used special sentinels blocks, which are hidden among other blocks in the data file F. In initial phase, the verifier randomly embeds these sentinels among the data blocks. To check the integrity of the data file F, the verifier challenges the prover (cloud archive) during the verification phase by specifying the positions of a collection of sentinels and asks the prover to return the associated sentinel values. If the prover has modified or deleted a substantial portion of F, then with high probability it will also have suppressed a number of sentinels. Therefore it is unlikely to respond correctly to the verifier. To indistinguish the sentinels from the data blocks, the whole modified file is encrypted and stored in the archive. Here the use of encryption renders the sentinels indistinguishable from other file blocks. This scheme is best suited for storing encrypted files.

It becomes computationally cumbersome to encrypt data file especially when the data to be encrypted is large as this scheme involves encrypting data file. Hence, this scheme has disadvantage that small users are left with limited computational power (PDAs, mobile phones etc.). This method also has storage overhead on the server, partly due to the newly inserted sentinels and partly due to the error correcting codes that are inserted. And the clients need to store all the sentinels with them, what may be storage overhead to thin clients (PDAs, low power devices etc.).

It is not a practical solution to simply download the file for its integrity verification as it requires high cost of input/output and transmission cost across the network. Also it is not easy to check the data thoroughly and compare with our data.

If we consider the large size of the outsourced data and the owner‘s constrained resource capability, the task of auditing the data correctness in a cloud environment can be formidable and expensive for data owners. To fully ensure data security and save data owners‘ computation resources, we propose to enable publicly auditable cloud storage services, where to verify the outsourced data, the data owners can resort to an external TPA when
needed. The TPA provides a transparent and cost-effective approach for establishing trust between client and cloud service provider. Based on the audit report of TPA, the released audit result would help the data owner to evaluate the risk of their subscribed cloud data services, and also beneficial for the CSP to improve their cloud based service platform.

IV. CONCLUSIONS

In this paper we explained different existing paper techniques and their merits and demerits. We discussed their methods of data security and privacy etc. In all those papers some haven’t described proper data security mechanisms, some were lack in supporting dynamic data operations, some were lack in ensuring data integrity, while some were lacking by high resource and computation cost. Hence this paper gives overall clue of all existing techniques for cloud data security and methods proposed for ensuring data authentication using TPA.

ACKNOWLEDGMENT

This work was supported by Assistant professor Mr. Vinit kumar Gupta and Assistant professor Mr. Indrajit Rajput. I would like to thank them for their guide and help.

REFERENCES

[1] http://www.pds.ewi.tudelft.nl/~iosup/research_cloud.html
[2] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li,‖ Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing‖ in IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, Vol. No. 22, Issue 5, MAY 2011.
[3] Cong Wang and Kui Ren, Wenjing Lou, Jin Li,‖Toward Publicly Auditable Secure Cloud Data Storage Services‖ in IEEE Network July/August 2010
[4] M. Krigsman, ―Apple‘s MobileMe Experiences Post-Launch Pain,‖ July 2008; http://blogs.zdnet.com/projectfailures/?p=908.
[5] A. Juels, J. Burton, and S. Kaliski, ―PORs: Proofs of Retrievability for Large Files, Proc. ACM CCS _07, Oct. 2007, pp. 584–97.

[6] G.Ateniese et al., ―Provable Data Possession at Untrusted Stores,‖ Proc. ACM CCS _07, Oct. 2007, pp. 598–609.
[7] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou,‖Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing‖ in IEEE INFOCOM 2010, San Diego, CA, March 2010.
[8] Federal Information Processing Standards (FIPS) 140-2. (2001, May 25). Security Requirements for Cryptographic Modules. Gaithersburg, MD: National Intitute of Standards and Technology (NIST). Retrieved from http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf.
[9] Voorsluys, William; Broberg, James; Buyya, Rajkumar (February 2011). "Introduction to Cloud Computing". In R. Buyya, J. Broberg, A.Goscinski. Cloud Computing: Principles and Paradigms. New York,USA:WileyPress.pp.1–44.ISBN978-0-470-88799-8. http://media.johnwiley.com.au/product_data/excerpt/90/04708879/0470887990-180.pdf.
[10] http://dwachira.hubpages.com/hub/Data-Security-Risks-In-Cloud-Computing.
[11] Abhishek Mohta,Ravi Kant Sahu and LK Awasthi, "Robust Data Security for Cloud while using Third Party Auditor" in International Journal of Advanced Research in Computer Science and Software Engineering, Vol No. 2, Issue 2,Feb 2012.

