ISSN: 2321-7782 (Online)
Volume 2, Issue 12, December 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Public Auditing of Big Data with Fine Grained Updates on Cloud

Surapriya Swain[1]
Student of ME (Computer)
G S Moze College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra – India

Saurabh Gupta[2]
Assistant Professor
G S Moze College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra – India

Abstract: ***Today data are created by individual and organization in a rapid rate. The rate data is growing and the need to maintain them is in high demand. Various data generated by the organization, government or business industry are managed by a external storage provider called CSS (Cloud Storage Service). CSS provide the facility of data virtualization by spreading it in various distributed data centre and data on demand service. The communication between CSS and cloud user is monitored and controlled by TPA(Third Party Auditor). AS the TPA is an external agent, it can also be malicious or it can misuse the user's data stored in Cloud storage. So this paper focuses on checking the authenticity of the TPA. Another issue of the current public data auditability is fine grained data updates. In social networking sites always there is need of small changes which incurs a large communication overhead if we follow the previous methods. In this paper methods are suggested which support fine-grained data update with less communication overhead. Measures are taken to calculate the efficiency of the proposed system which is better than some existing ones.***

Keywords: CLOUD COMPUTING, BIG DATA, AUTHORIZED AUDITING, FINE-GRAINED DATA UPDATES.

I. INTRODUCTION

In today's world of digitization, cloud computing has emerged as a concept of handling Big data. This paper focuses on the nature of Big Data, origin of Big data and security related issues with big data.

Data are originated from various domains like science, education, industry, healthcare and many more. The feature of data generated from different source's are different. The definition of Big data includes 5 V: Velocity, Volume, Variety, Value, Veracity.

Big data is supported by new infrastructure and tools. Cloud based infrastructure, storage, network, high computing performance helps to manage the feature of Big data . New data centric security models for trusted infrastructure and data processing and storage are also proposed for the above purpose.

Big Data is not a simple Database rather it contains large scale data processing and data analytics. The most important part of Big data is its support to Dynamicity. Big data require different data centric operational models and protocols .Sometimes object or event related data go through a number of transformations and became more distributed between traditional security domains.

Today's Big data architecture framework comprises of 5 different aspects of Big data definitions.

i. Data Models, Structures and Type

ii. Big Data management

iii. Big Data Analytics and Tool

iv. Big Data Infrastructure

v. Big Data Security

In this paper we will concentrate only on the security part of it which include Data security in-rest, in-move and in trusted processing environment.

Different transformation of data starts once the data is published. It means the data is uploaded in some web site. There should be provision of allowing other legitimate user to audit and reproduce the data in other environment. Here comes the security aspect of Big Data.

The second major concern is Big data is usually distributed at the collection side as well as at processing side. Linking of these distributed data is one of the problem to be solved.

Here comes the concept of Cloud computing. It is an emerging technology.

A cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers .

The necessary characteristics of Cloud computing are

(i) *On-Demand Self Service*: A consumer can acquire one-sided computing resources by pay per-service basis.

(ii) *Resource Pooling*: The service provider's pooled the computing resources to serve diverse users, with different the physical and virtual resources dynamically.

(iii) *Selection of Provider*: One must make sure that the provider is consistent, well-reputed for their customer service and should have a verified path record in IT- related ventures.

(iv) *Rapid Elasticity*: Capabilities can be rapidly and elastically provisioned to the cloud users dynamically and automatically.

(v) *Measured Service*: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service.

Cloud Computing offer various services like IaaS (Infrastructure as a Service), SaaS (Software as a Service), PaaS (Platform as a Service). IaaS is the way of providing on-demand computing resources like Server, storage array, virtualized data centres etc. PaaS is providing a higher level Software application which can be compatible to the different user's requirement. SaaS is the way of providing some specific applications as fully or partially remote services. It may include web based application or network interactions.

The above services provided by CSS makes the cloud user to be relaxed from burden of storing, managing and providing on-demand service to the client. The overhead incurred by implementing these entire infrastructure by own is reduced somehow. So now days Cloud computing is in great demand. It is an emerging technology in the world of IT.

But still there are some hidden problems that we will be analysed in our paper. The issues related to security, integrity and availability of data. There is no direct control of user on cloud. But data integrity can be verified without possession of actual data. Verification done by a trusted third party (TPA) called data auditing. TPA can be anyone challenging the integrity of data stored in CSS.

Our research work aim to authenticate the TPA which authenticate the cloud user and Cloud service provider. Another aim is to allow fine-grained dynamic data update in CSS. Its advantages, efficiency calculation will be analysed in our paper.

II. Related Theory

Cloud storage moves the user's data to large data centres, which are remotely located, on which user does not have any control. This feature poses security challenges which needs to be resolved . In cryptography, a Trusted Third Party Auditor (TPA) is an entity which facilitates secure interactions between Cloud user and cloud service provider and both parties trust on TPA . The purpose of a TPA is to provide end-to-end security services, which are scalable, and useful across different domains, geographical areas and specialization sectors. The cloud environment includes the Trusted Third Party Auditor (TTPA) for verifying the data storage correctness of cloud server in a timely manner. As described by Castell, ''A Trusted Third Party is an impartial organization delivering business confidence, through commercial and technical security features, to an electronic transaction.

The involvement of TPA demands retrieval of user data which does not remain secret and also TPA has to remember the keys for transactions. For static data it is ok. But to deal with dynamic data this feature cannot ensure data secrecy and integrity. Ateniese et al. [4] first considered public auditability in their defined "provable data possession" (PDP) model for ensuring possession of files on un trusted storages. In "proof of retrievability" (PoR) model spot-checking and error-correcting codes are used to ensure both "possession" and "retrievability" of data files on archive service systems .Remote data auditing(RDA) refers to a group of protocols to securely ,frequently, and efficiently verify the correctness of the data over a cloud managed by untrustworthy provider without having to retrieve the data. By using the BLS signature the integrity and authenticity of the updated data can be verified.

When the feature of Big data is implemented in any social networking site, or government portal for storing statistical data, there is always a requirement of small changes as the time changes. Also another need for managing dynamic data i.e. data used in running application. Previous research work suggested to access the whole block where the data resides and change the exact data and again restore it in cloud. To better support scalability and elasticity of cloud computing, some recent public data auditing schemes do support data dynamics.

III. Literature Review

Cloud computing is on high demand today because of its features like scalability, elasticity and efficiency in supporting dynamic data. Cloud users are able to conveniently scale up/ down their virtual allocated resources according to their current need with minimal management effort and service interruption. The most existing problem in cloud is the data security and privacy. Integrity verification for outsourced data storage is the main area of today's research. Jules et al. [1] proposed a model based on POR which is only applicable to static data storage. Ateniese, et al. proposed a similar model based on PDP which verifies the integrity of a proportion of the outsourced file through verifying a combination of pre- computed file tags which they call homomorphic verifiable tags(HVT). Shacham, et al. [2] proposed another model which is based on BLS signature scheme. BLS signature is shorter than RSA signature. In 2009, Erway, et al. proposed the first PDP scheme based on skip list that support full dynamic data updates. In any of this proposal public auditability and variable sized data block are not supported by default. Wang, et al. [3] proposed a scheme based on BLS Signature which supports the above but does not provide the facility of fine-grained data updates and authorized auditing.

IV. Proposed System

It is important to assure customer about the integrity of their data in cloud. Storage correctness to ensure that there exists no cheating cloud server that can pass the TPA's audit without indeed storing users' data intact.

The relationship between the cloud user and cloud service provider is transparent.

Cloud user will utilize the resource of cloud on pay as you use basis. The SLA signed between the user and service provider is not transparent. This agreement includes the Cloud service provider's quality of service, Standard of the service, service

monitoring and controlling. TPA is there to audit the SLA and check if CSS is violating any rule to hide its fault. TPA has the list of auditing strategy and can check the integrity of data stored in cloud storage. Privacy preserving ensure that TPA cannot derive user's data content from the information collected during auditing process.

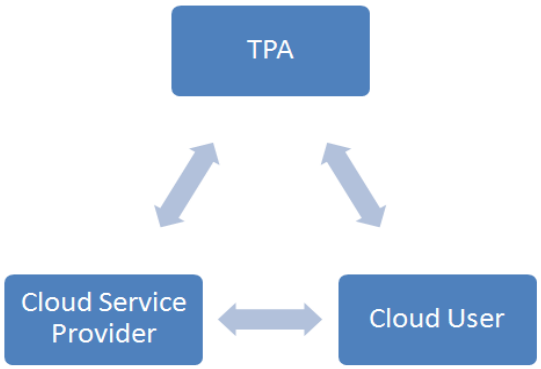

Fig 1 : TPA in a Cloud environment

As the TPA is an external entity there is also chance that it can steal users sensitive data. So in our proposed system we will implement some methodology to verify TPA's integrity.

The process is like both CSS and Cloud user will authorize the third part auditor by running various algorithms keeping the public key of client as common parameter. Once TPA is authorized it can send challenge to Cloud server for checking user's data integrity.

A. Frame Work and Definition

To achieve fine-grained data updates different operations are carried out. Block level operation in fine-grained data updates contain 5 types of operations.

- Partial Modification(PM) :- It is to update a consecutive part of a certain block.
- Whole Block Modification(M) :- A whole block is replaced by new set of data
- Block Deletion(D) :- The whole block is deleted from the tree structure.
- Block Insertion(I):- A whole block is inserted to carry new set of data
- Block Splitting(SP) :- Some part of the existing block is taken out and new block is created to be added to the tree structure.

B. Authorization Of TPA

To verify TPA's authenticity our scheme proposes 3 steps. They are Setting up the environment, Fine-grained update verification and Challenge, verification and proof generation.

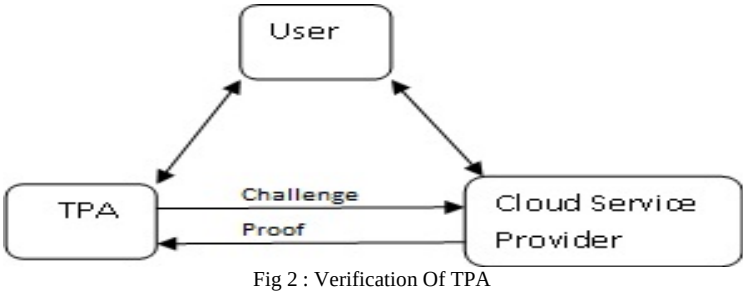

Fig 2 : Verification Of TPA

Setup :- This phase is based on the BLS signature scheme. The client generate keying material by KeyGen and Fileproc . Then client upload the data to CSS. The client store a RMHT as a metadata and authorize TPA by sharing the value of sig_{AUTH} .

After all the parties finished with the negotiation operation the client run the key generation algorithm. This algorithm outputs a secrete key and a public key.S_{max} denotes the number of segments per block. After the setup phase, the client prepare for authorization by asking TPA for its IDVID which is used for authorization.TPA return its ID by encrypting with clients public key. After comparing the value of Sig_{AUTH} with secrete key ,client sends its auditing request to TPA.

Fine-grained Update Verification :- This process occurs between client and CSS. The client send fine-grained update request to CSS via PerformUpdate and client runs verifyUpdate algorithm to check whether CSS has performed the update correctly on the data block as well as in corresponding authenticator.

To update a certain portion of the data block client has to adopt the PM process i.e partial modification. This involves the following steps:

First client compose a Update Request and send to CSS and CSS run the PerformUpdate (UpdateRequest,F) algorithm .

Secondly CSS send the P_{update} to client and client run the VerifyUpdate ($P_{k,}$; P_{update}) algorithm .

Finally if the algorithm returns true value it update the old value with new ones otherwise if it returns false then client send the Update request again.

Challenge, Verification And Proof generation:

This is the last step in verifying TPA's authenticity. In this phase TPA has to show that it is the genuine one who is challenging the CSS for data integrity checking. TPA runs the GenChallenge() algorithm with private key and signature as parameters. Then a challenge message is generated with TPA's new ID selected randomly from the set of total blocks. This VID is encrypted with CSS's public key. After this process TPA can send challenges to CSS.

When CSS get the challenges it will run another algorithm to verify the signature, VID and client's public key. If the algorithm returns a true , then CSS will send a proof "p" to TPA and TPA will run the algorithm Verify(pk, challenge, p) otherwise if the algorithm run by CSS returns false , the request is rejected.

For TPA authorization, a signature scheme is chosen which can be forged by malicious TPAs . No malicious TPA cannot make the CSS to respond to its challenge which contain an integrity proof for a subset of existing file in CSS.

C. Fine-Grained Data Updates

The second major concern of our research work is to be able to do fine-grained updates in contrast to coarse-grained updates. Fine-grained update in dynamic data is the provision of doing small changes in the corresponding data block instead of accessing and changing the whole block. This method can reduce the communication overhead as occurring in previous methods.

We can define the fine-grained update request as the set of 3 variables .

- Starting address of the update in the file
- The length of the file
- The new message to be inserted into the File

For this update, we have to sure that the new message is not exceeding the maximum block size after updation.

V. CONCLUSION

Cloud computing is a big computing paradigm. Cloud data security is the important aspect for the cloud user. A trusted third party can ensure the security and integrity of data. This paper presents an overview of trusting a third party. It focuses on privacy-preserving which means TPA cannot derive user's data during the process of public data auditing. The proposed system uses a signature scheme which cannot be forged so that it will prevent malicious TPAs. It provides a feature of fine-grained dynamic data update which increases the efficiency of update process.

ACKNOWLEDGEMENT

I would like to express my thanks to my guide Prof. Saurabh Gupta for his highly appreciable support and encouragement also to my HOD Prof. Ratnaraj Kumar . Their guidance is a force behind the completion of this paper. I am grateful for all the suggestions and hints provided by him. My acknowledgment of gratitude to all who supported to make it possible.

References

1. A. Juels and B.S. Kaliski Jr., ''PORs: Proofs of Retrievability for Large Files,'' in Proc. 14th ACM Conf. on Comput. and Commun. Security (CCS), 2007, pp. 584-597.
2. H. Shacham and B. Waters, ''Compact Proofs of Retrievability,'' in Proc. 14th Int'l Conf. on Theory and Appl. of Cryptol. and Inf. Security (ASIACRYPT), 2008, pp. 90-107.
3. Q. Wang, C.Wang, K. Ren,W. Lou, and J. Li, ''Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing,'' IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5,pp. 847-859, May 2011.
4. G.Ateniese, R.B. Johns,R. Curtmola, J.Herring, L. Kissner,Z. Peterson, and D. Song, ''Provable Data Possession at Untrusted Stores,'' in Proc. 14th ACM Conf. on Comput. and Commun. Security (CCS), 2007, pp. 598-609.
5. C. Wang, Q. Wang, K. Ren, and W. Lou, ''Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,'' in Proc. 30st IEEE Conf. on Comput. and Commun. (INFOCOM), 2010, pp. 1-9.
6. Cong Wang, Sherman S.-M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," IEEE Transactions On Cloud Computing, Year 2013.
7. C. Wang, K. Ren, W. Lou, and J. Li,"Toward publicly auditable secure cloud data storage services," IEEE Network , vol. 24, no. 4, pp. 19–24, 2010.

AUTHOR(S) PROFILE

Mrs Surapriya Swain, received the B.E degree in Computer Science and Engineering from B.I.E.T in 2005, affiliated to Biju Pattanaik University Of Technology, Odisha, India. She is now pursuing her M.E (Computer Engineering) in G. S. Moze College of Engineering under Savitribai Phule Pune University.

Mr. Saurabh Gupta, working as an Assistant Professor with G. S. Moze College of Engineering under Savitribai Phule Pune University (Maharashtra, India). He has received his B.E(IT) degree from Agra University(UP, India) in 2004 and M.E from BITS Pilani(Rajstan, India) in the year 2009.

