International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 5(2), September - October 2015 ISSN 2278-6856

Big data privacy-preserved public auditing for shared data in cloud

Gopika Mangalassery[1] , Prof. Manilal D.L[2]

[1]College of Engineering Cherthala ,Cochin University, Computer and information science
Cherthala,Kerala,688541.

[2] Associate Professor, College of Engineering Cherthala , Cochin University,Computer Science & Engineering
Cherthala,Kerala,688541.

Abstract

Cloud storage is common place for share data's across single users as well as multiple users. Open challenge in this shared data is public auditing while preserving identity privacy. This paper propose the privacy preserving mechanism that allows public auditing on shared data stored in the cloud and how to audit the integrity of shared data in the cloud with dynamic groups [a new user can be added into the group or revoked during data sharing] while preserving identity. In order to audit the integrity of shared data exploit the ring signatures to compute the verification information. With this mechanism, the identity of the signer on each block in shared data is kept private from a third party auditor (TPA),who is still able to verify the integrity of shared data without retrieving the entire file. While cloud is a storage space for huge data, implementing this in hadoop for better performance and big data processing simpler and faster. For auditing the big data parallely batch auditing is implementing using Hadoop.

Keywords: Public auditing, Cloud computing, Shared data, Big data.

1. INTRODUCTION

Cloud computing is a Internet based computing which enables sharing of services. Using Cloud Storage, users can remotely store their data and enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. The advantage of cloud is cost saving. The prime disadvantage is security. To ensure security, cryptographic techniques cannot be directly adopted. Sometimes the cloud service provider may hide the data corruptions to maintain the reputation. To avoid this problem, introduce an effective third party auditor (TPA) to audit the users outsourced data when needed. Security in cloud is achieved by signing the data block before sending to the cloud. Moreover, users should be able to just use the cloud storage as if it is local, without worrying about the need to verify its integrity. Thus, enabling public audit ability for cloud storage is of critical importance so that users can resort to a Third Party Auditor (TPA) to check the integrity of outsourced data.TPA audit the cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user.

With data services in the cloud, users can easily modify and share data as a group. To ensure data integrity can be audited publicly, users need to compute signatures on all the blocks in shared data. Different blocks are signed by different users due to data modifications performed by different users. For security reasons, once a user is revoked from the group, the blocks, which were previously signed by this revoked user, must be resigned by an existing user. The straight forward method ,which allows an existing user to download the corresponding part of shared data and re-sign it during user revocation, is inefficient due to the large size of shared data in the cloud. To protect the integrity of data in an untrusted cloud, a number of mechanisms have been proposed. In these mechanisms, a signature is attached to each block in data, and the integrity of data relies on the correctness of these signatures. One of the most significant and common features of these mechanisms is their ability to allow not only the data owner, but also a public verifier, such as a third party auditor (TPA), to check data integrity in the cloud without downloading the entire data, referred to as public auditing. Most of the previous works focus on auditing the integrity of personal data. Different from these works, our recent work focuses on how to preserve identity privacy from the TPA when auditing the integrity of shared data.

The traditional approach for checking data correctness is to retrieve the entire data from the cloud, and then verify data integrity by checking the correctness of signatures (e.g., RSA) or hash values (e.g., MD5) of the entire data. Certainly, this conventional approach is able to successfully check the correctness of cloud data. However, the efficiency of using this traditional approach on cloud data is in doubt. ORUTA explains the privacy in public auditing but the problem it deals with static group members and the big data auditing is not much faster.

In this paper, to solve the above issues on shared big data, I propose big data privacy preserved public-auditing for shared data in cloud. More specifically, we utilize ring signatures to construct homomorphic authenticators, so that a public verifier is able to verify the integrity of shared data
without retrieving the entire data while the identity of the signer on each block in shared data is kept private from the
public verifier. In addition, we further extend our mechanism to support big data processing faster by batch auditing, which can perform multiple auditing tasks simultaneously and improve the efficiency of verification for multiple auditing tasks.

2. RELATED WORKS

Data encryption before outsourcing is the simplest way to protect data privacy and combat unsolicited access in the cloud and beyond. But encryption also makes deploying traditional data utilization services such as plain text keyword search over textual data or query over database is a difficult task. The trivial solution of downloading all the data and decrypting it locally is clearly impractical, due to the huge bandwidth cost resulting from cloud-scale systems.[1]

Fully homomorphic encryption (FHE) has shown the general results of secure computation out sourcing to be viable in theory. But applying this general mechanism to everyday computing tasks is still far from practical due to FHE operations extremely high complexity, which cannot yet be handled in practice.

Anonymity and Traceability: Anonymity guarantees that group members can access the cloud without revealing the real identity. Although anonymity represents an effective protection for user identity.[2]

After registration, user i obtains a private key (xi; Ai; Bi) which will be used for group signature. User revocation is performed via public available revocation list (RL), based on how many group members is connected to the server socket and ensure the confidentiality against the revoked users. [6]

The technique of providing more security by using the Third Party Auditor (TPA) .The TPA allows the user to know the information about the data stored in the cloud. When anyone tries to modify or steal the data TPA informs the user by verifying the data. The TPA doesn't even allows the CSP to read the data of the user. Here utilize the public key based homomorphic authenticator and uniquely integrate it with random mask technique.TPA can perform multiple auditing tasks simultaneously. [4]

This method use leveraged homomorphic tokens to ensure the correctness of erasure codes-based data distributed on multiple servers. The major contribution of this mechanism is able support dynamic data, identify misbehaved servers. Drawback is leakage of identity privacy to public verifiers. [8]

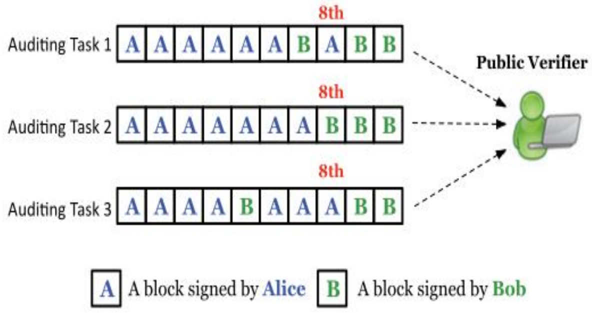


Figure 1: Alice and Bob share a data file in the cloud, and a public verifier audits shared data integrity.

Signing of users is public, so privacy is less [5] .PORs scheme which is also able to check the correctness of data on an untrusted server. The original file is added with a set of randomly-valued check blocks called sentinels. The verifier challenges the untrusted server by specifying the positions of a collection of sentinels and asking the untrusted server to return the associated sentinel values. Sentinel based POR protocol is amenable to real-world application [9].

Drawbacks are:

- ➢Only focus on personal data in the cloud.
- ➢ Integrity Threats: While privacy of user is preserved, it's not considering dynamic number of users. And also about the running time of Big data's.

3. SCOPE OF THE PROJECT

This will lead to a high secured auditing process, which can be used in highly confidential soft wares like bank auditing process etc. Batch auditing process will be efficient and faster.

3.1 Problem Statement

Identify the method of supporting dynamic users and implement batch auditing using hadoop to improve the efficiency"

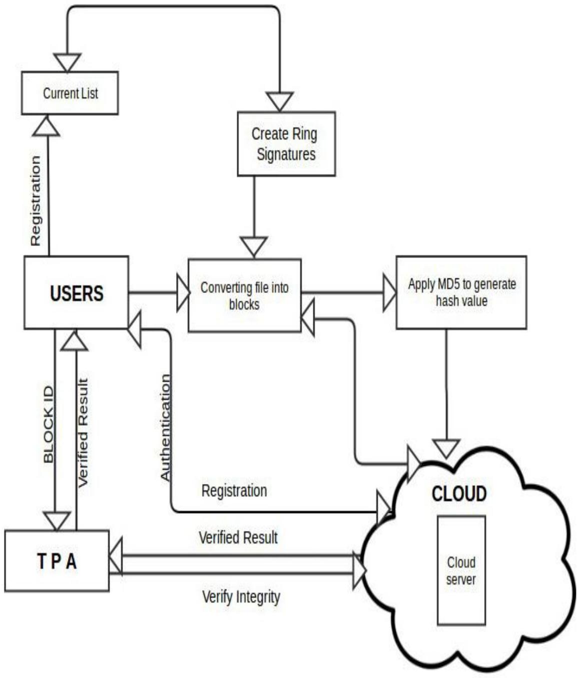
4. PROPOSED SYSTEM

A.Users

1. Each registration of users lead to an entry to current list. Which supports dynamic users (Algorithm 1)

2. During uploading file it will splits into equal sized blocks of data. Which helps public verifier to audit shared data integrity without retrieving the entire data (Algorithm 2)

3. Generate Ring signature for each users for signing. Which cannot distinguish who is the signer on each block.



F

figure 2 : Architecture

Structure of ring signature contains private key of the signer and public key of other users in that group, which is visible as a hex code (Using ECDH).

Algorithm 1: User registration

```
for each user 1, 2....n do
    Add user to current list;
    compute ring signature;
end for
while current list not empty do
    for each user deleted do
        recompute the ring signatures;
    end for
end while
Stop
```

4. Apply MD5 algorithm in each block for finding hash value. For verifying integrity of data (Using MD5).

5. Upload file to cloud for sharing it with group.

6. Send a request to TPA for verifying integrity by passing ring signature and block id.

Algorithm 2: File Upload

```
Upload file of any size
Divide the size of the file by 5 and store the resultant value in variable int splitSize
splitCount=0;
while (splitCount<5) do
    Read the file till the splitSize value;
    Write the partially read file to the destination path;
    splitCount=splitCount+1;
end while
Stop
```

B. TPA

7. Invoke verification algorithm (MD5) on randomly selected blocks of data.

8. TPA has the role of auditing the integrity of files that are shared within a group. The client will request the TPA on which file and block to check. The TPA will request the cloud server for that block and will download that block. After download, the hash function will be applied on that block. If the newly computed hash value and the hash value in the cloud server are the same, it indicates that the file has not been modified (corrupted) and will reply to the client that the block is safe. If the values do not match, it indicates modification (corruption) and the TPA will reply to the client that the block is modified and to re-upload the file.

C. Cloud Server

9. In cloud a database is integrated to store files uploaded which is shared to other users.

10. For simulating the corruption in data, modification can be done.

D. Batch Auditing

Here in this module the batch auditing for multi-client data is explained. And broaden our method to allow for provable data updates and verification in a multi-client system. The ring signature method allows the creation of signatures on arbitrary distinguishable messages. Furthermore, it supports the aggregation of multiple signatures by distinct signers on distinct messages and it greatly reduces the communication cost while giving efficient verification for the authenticity of all datas.

11. Parallely comparing MD5 value by TPA in the case of Big datas using the technique hadoop .In hadoop mainly three classes are there. Give inputs to Mapper class split into different jobs with a key value and given to reducer class to verify (comparing hash value).Then the partitioner class return the output.

5. RESULT

In this work, I have presented a big data privacy-preserved public auditing for shared data in cloud. I have considered the processing speed of big data. It is important to consider the time taken to verify big datas. So by using batch auditing it is done successfully. And also support dynamic users. There by efficiency of system improved.

6. CONCLUSION

This system is dealing with public auditing of shared data in cloud. Utilize ring signatures, so that a public verifier is able to audit shared data integrity without retrieving the entire data, yet it cannot distinguish who is the signer on each block. And also supports dynamic group of users while preserving privacy. To improve the efficiency of Big data auditing process by implementing batch auditing as a multiple auditing tasks.

References

[1] K. Ren, C. Wang, and Q. Wang, Security Challenges for the Public Cloud, IEEE Internet Computing, vol. 16, no. 1, pp. 69-73, 2012.

[2] Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud Xuefeng Liu, Yuqing Zhang, Member, IEEE, Boyang Wang, and Jingo Yan.

[3] Proxy Re-Signatures: New Definitions, Algorithms, and Applications Cloud.

[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. 598-610, 2007.

[5] C. Wang, Q. Wang, K. Ren, and W. Lou, Ensuring Data Storage Security in Cloud Computing, Proc. 17th Intl Workshop Quality of Service (IWQoS09), pp. 1-9, 2009.

[6] Juels and B.S. Kaliski, PORs: Proofs of Retrievability for Large Files, Proc. 14th ACM Conf. Computer and Comm. Security (CCS07),pp. 584-597, 2007.

[7] B. Wang, B. Li, and H. Li, Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud, Proc. IEEE Fifth Intl Conf.Cloud Computing.

[8] M. Armbrust, A. Fox, R. Griffith, A. D.Joseph, R. H.Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A View of Cloud

Computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, Apirl 2010.

[9] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S. S.Yau, "Dynamic Audit Services for Integrity Verification of Outsourced Storage in Clouds," in Proc. ACM Symposium on Applied Computing (SAC), 2011, pp. 1550–1557.

[10] D. Boneh and D. M. Freeman, "Homomorphic Signatures for Polynomial Functions," in Proc. International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT). Springer-Verlag, 2011, pp. 149–168.

AUTHOR

Gopika Mangalassery received the M-tech post graduation in Computer and information science from College of engineering Cherthala in the duration 2013-2015. During the year 2014-15 she done literature review about the Big data auditing challenges.