IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

SURVEY ON PRIVACY PRESERVING CLOUD AUDITING FOR SHARED DATA

Kashyap Chetan Kotak*, Ria Wadhwani, Shikha Soneji, Prof. Sunita Sahu
Dept. of Comp. Engg, VESIT, Mumbai, India

DOI: 10.5281/zenodo.266689

ABSTRACT

Cloud is a type of platform which helps to store the data as well as helps in sharing the data. While sharing the huge amount of data, the primary concern comes in mind is data integrity, security of the data. For the cloud server and the user it is not possible to check the integrity, consistency of stored data on cloud. Public Auditing method can help to overcome this problem. Hence the user takes help of the third party auditor (TPA) for auditing their data. Many of the techniques have been proposed by various researchers which improve upon one another. Here we have presented three landmark methodologies on Privacy preserving cloud auditing for shared data on cloud.

KEYWORDS: Cloud computing; public auditing; PDP; WWRL; ORUTA; TPA.

INTRODUCTION

Cloud Computing is an upcoming new technology which provides the on-demand facility of a shared pool of resources (computing resources) (e.g., computer storage, applications and other resources), which can include rapid allocation and freedom with minimum number of efforts. Cloud computing and storage solutions provide individual users and companies with variety of capabilities to store and work on their data in data centers which are not owned by them and the location may be remote, may be across a city or across continents. Cloud computing provides sharing of resources to achieve economy of scales. The users use this cloud for sharing and the collaboration of their data with many other users in the group. Data sharing has become need in today's world and it is provided in most of the cloud storage offerings, via Dropbox, Google.

The data integrity in cloud storage, is a subject to skepticism and inspection, as there is always this fear of data being stored in an environment where there is a chance of getting lost or corrupted[1]. The need of the Third Party Auditor (TPA) is very necessary for ensuring the integrity of the data.

Allowing public auditability for cloud storage is important so that users assign a third- party auditor (TPA) for the checking of the integrity of outsourced data and TPA offers its auditing service with more commanding computation and abilities of the communication than regular users. If we mention information, Wang et al. designed to construct a mechanism of public auditing system for cloud data, so that during public auditing process, the contents of the private data that belong to a personal user is not revealed to the third party auditor.

Sharing data among the multiple users is one of the features in the motivation of cloud storage. A unique problem that is introduced during the whole process of public auditing for the shared data in the cloud is how we should reserve the identity secrecy from the TPA, because the identities of each signer on shared data may show that a particular user among the group of users or special blocks in shared data is amore valuable object than others. Such information is very confidential to the group and should not be shown to any third party as there is data shared. In this paper, we have also studied Oruta , a new privacy-preserving public auditing mechanism for shared data in an untrusted cloud.

PROBLEM DEFINITION

The current/existing system demands public auditability of the data that is shared on cloud. While many techniques fails to preserve the identity privacy or data privacy on shared data during the process of public auditing that has the fear of revealing the significant and the confidential information to public verifiers. So for the protect of this confidential information and identity, it is very essential as critical to preserve identity privacy and data privacy from public verifiers during public auditing.

GENERAL MODEL

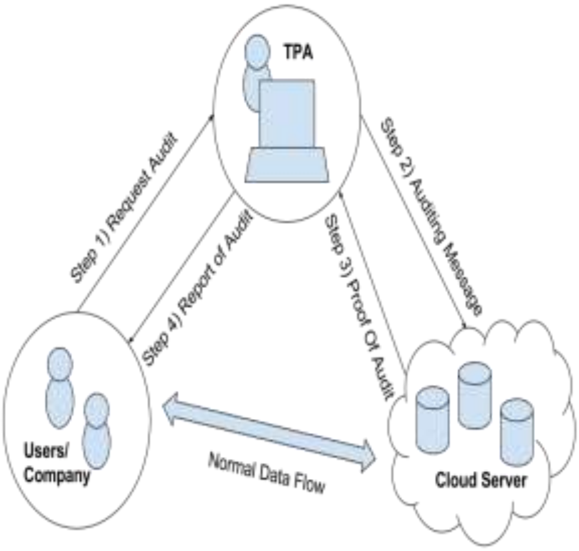

Fig 1: general model of public auditing

As in the illustratration from the Figure 1 , in Third Party Auditing, three parties are involved: The cloud server, The third party auditor (TPA) and The users. There are mainly two types of users in a group: the original user and some group users. The original user and group users, both of them members of the group. Group members are allowed to access as well as modify the data which is shared is created by the original user which is based on policies of the access control. Shared data and its information from verification (i.e. signatures) are both stored in the cloud server. The third party auditor does the verification of the integrity of shared data in the cloud server on behalf of the members of the group.

CONSTRAINTS

Following are the constraints required in our project:-

1. The TPA should not be able to view the contents of the data
2. The TPA should not be able to know about the id- entity of signers of the data.
3. The cloud must support and allow public verifiability.
4. Groups are assumed to be static (i.e. adding new members will require additional work and is not included in this project)

SURVEY

PDP [2]

A model called Provable Data Possession (PDP) was put upon by G.Ateniese. In this model client wants to verify that server retains original data without retrieving the complete data which is stored at untrusted server. This model allows server for accessing the small parts of file in generating the proof. Therefore this model generates probabilistic proofs of possession by sampling of the random sets of blocks from server, which significantly reduces I/O costs.

Requirements & Parameters: The important performance parameters of a PDP scheme include:

1. Computation Complexity:- The computational cost of preprocessing a file at client, to generate a proof of possession at Sender and to verify such a proof at client.

2. Block Access Complexity:-The number of file blocks accessed to generate a proof of possession at sender.
3. Communication Complexity:-The amount of data that is transferred between sender and client.

The Preliminaries of PDP:

The client C wants to store a file S on the server file F which is a finite ordered collection of n blocks: F = (m1,... ,mn). We denote the output that is x of an algorithm A by x ← A. We denote by |x| the absolute value of variable x.

Homomorphic verifiable tags act as verification metadata for the file blocks and, with being unforgeable, they also have the following properties:

Blockless verification:

Use of the HVTs the server can construct a proof that allows the client to verify if the server has the possession of certain file blocks, even when the client does not have access to the actual file blocks. Homomorphic tags: Given are two values Tmi and Tmj , anyone can combine them into a value Tmi+mj which corresponds to the sum of the messages mi + mj .

A PDP is a collection of four polynomial time algorithms:

1. KeyGen(1^k) → (pk, sk) is a probabilistic ind of key generation algorithm that is run by the client for setting up of the scheme. It takes a security parameter that is k as input, and returns a pair of matching public and secret keys (pk, sk).
2. TagBlock(pk, sk, m) → *Tm* is a (possibly probabilistic) algorithm that is run by the client for the generation of verification metadata. It takes as inputs a public key pk, a secret key sk and a file block m, and returns the value of verification metadata Tm.
3. GenProof(pk, F, chal, Σ) → *V* is run by the server for the generation of a proof of possession. It takes as inputs a public key pk, an ordered collection F of blocks, a challenge and an ordered collection Σ which is the verification metadata that corresponds to the blocks in F. It returns a proof of possession V for the blocks in F that are calculated by the challenge chal.
4. CheckProof(pk, sk, chal, V) → {"success", "failure"} is run by the client for validation of a proof of possession. It takes as inputs a public key pk, a secret key sk, a challenge which is chal and a proof of the possession V. It returns whether V is a correct proof of possession for the blocks determined by chal.

Their scheme utilizes the RSA based homomorphic non-linear authenticators which audit the outsourced data and suggests randomly sampling a few blocks of the file.

Advantages:-

1. Can Check how much correct the data is stored in an untrusted server, without retrieving the entire data.

Disadvantages:-

1. The public auditability in their scheme demands the linear combination of sampled blocks that are exposed to external auditor. When used directly, their protocol is not a provably privacy preserving, and thus may leak user data information to the auditor.
2. Data and identity, both are not preserved for the benefit of third party auditing.

WWRL [3]

The problem with the PDP was that it worked properly, but it was not that much protective when it came to the privacy of the data stored in the cloud service as well as the identity of the user signing in. A scheme for public auditing for the preservation of the content of private data belonging to a personal user was proposed by Wang et al . It is opted for preserving the data from any kind of misuse done or malicious activities performed. It efficiently checks for integrity of cloud data without retrieving the local copy of data. This scheme can eliminate the burden from tedious work and also the expensive auditing tasks of the cloud user and efficiently preserves the data of user . There is a third party auditor(TPA) assigned for checking and the verification of cloud data. The mechanism that is proposed utilizes some of the random masking techniques. In this mechanism the private content which belongs to a personal user is not published to the third party auditor. It supports only public auditing and privacy of the data. The TPA checks the data and audits it without retrieving the entire file of the user, thus maintaining the privacy of the data being shared on the cloud. However, there is no guarantee of the TPA not knowing about the identity of the user. That means , the identity of the user who has shared files on the cloud services is not preserved.

Advantages:-

The data is preserved, as the auditor can carry out the auditing process without sneaking into the data the user has shared.

Disadvantages:-

Though the data is preserved, there is no guarantee of the identity of the user being protected from being revealed in front of the TPA.

ORUTA [4]

In this paper a mechanism for public auditing for shared data that is privacy preserving is first proposed for cloud that is shared. Where The Identity of The original Signer of the data is not revealed to the TPA which makes it different from the other two just described. The experimental results of the authors has proved and demonstrated the efficiency and effectiveness of this new technique.

Design Goals:

1. Public Auditability by consuming less Bandwidth: Third party auditor is able to audit the data by auditing aggregate block thus, preventing the download of the whole data and also hides data.
2. Verification of Correctness: The third party auditor should be able to check for correctness without knowing the data and should be able to specify the corrupted block of data.
3. Allowing Authentication of users: The users which do not belong to a group should not be able to generate a valid verification request to be sent to a TPA.
4. .Providing and Ensuring Identity secrecy:- When auditing the TPA takes place it should not reveal private key or any other authentication information that discloses the identity of a particular user in the group which leads to identification of valuable targets.

Techniques used:

The preliminaries that are used in this are bilinear maps and ring signatures. Also something first proposed in this paper is Homomorphic Authenticable Ring Signatures (HARS)

HARS contains three algorithms:

1. KeyGen : for generating user's public key and private key.
2. RingSign : it is an algorithm for signing a particular data block that uses one private key and all other users public key to hide the identity of the signer
3. RingVerify: The verifier can check whether a block is signed by someone from the group that uses this Algorithm.

Oruta is constructed, using the method of HARS and its properties. Using Oruta, the TPA is able to verify the data without retrieving all of them and still maintaining the identity privacy of the users who signed them includes five algorithms:

1. KeyGen: for generating a user's public and private key
2. SigGen : invalid user i.e a member of the group or the original owner of the data can sign a data block which is a combination of its own private key and all other user's public key.
3. Modify: evaluation is a can add delete or modify data and computer new signature on that data
4. ProofGen : used by both the TPA as well as the cloud server to generate a proof possession of shared data.
5. ProofVerify: the cloud sent this proof of procession to the TPA and the TPA verifies it.

Solution To Other Problems Proposed in ORUTA:

The size of the Ring signature is an issue which is solved in this paper by using an aggregate block that is aggregate of many other data packs.

Another issue which is solved in this paper is the dynamic changes in the data by valid users. This was achieved by the use of hashing techniques.

Advantages:

1. Preserves data from third party auditor and achieves data security.
2. Solves the problem of large storage space required for ring signatures.
3. Solves the problem of dynamic operations on data by Hash indexing.
4. Preserves user identity from third party auditor.

Disadvantages:

1. An unsolved challenge in Oruta remains that the groups need to be predefined. Dynamic changes in the group like addition or deletion or modification of user keys and changing the private information of users is not allowed. This remains the future scope of this paper.

OTHER RELATED WORK

To improvise on the efficiency of verification of PDP, Ateniese et al. [5] Constructed scalable and the efficient PDP uses symmetric keys. This mechanism is supportive of executing partially dynamic data operations. Unfortunately, it cannot support public verifiability and it only offers every user with very less number of verification requests.

Juels and Kaliski [6] have defined one more similar model called proof of retrievability (POR), which also checks how much correct the data is that is stored in an untrusted server. The original file is added along with a set of randomly-valued blocks called as the sentinels. The user carries verifies the integrity of data by asking the server to return some of the sentinel values. Shacham and Waters [7] have designed two improved POR, which are built on some random functions and BLS signatures [8].

Wang Et al.[9] in their paper on WWRL stated the Hash Tree for the making of a public auditing mechanism, which can be able to support fully dynamic data. Erway et al.[10] presented a fully dynamic PDP based on the rank-based authenticated dictionary. Zhu et al.[11] exploited index hash tables that supports the fully dynamic data during the public auditing process.

More recently, Wang Et al.[4] first considered the public auditing process for the cloud data along with data privacy. In this mechanism, the third party auditor checks the integrity of cloud data but it is hard to obtain any data that is private. In addition, to operate the multiple users' which are auditing the tasks simultaneously, they have also extended their mechanism for enabling of the batch auditing by leveraging aggregate signatures [12]. Recent work of B. Wang, B. Li, and H. Li, [13] (authors that proposed ORUTA) who proposed another technique called KNOX. This methodology which is called KNOX is able to audit the integrity of shared data for groups that have large number of members/users which was not feasible for their original technique ORUTA.But unfortunately, KNOX cannot support public auditing using the services of TPA.

CONCLUSION

The cloud storage services, it is common for data to be not only stored in the cloud but also is shared across several users. However, public auditing for shared data — while preserving identity privacy — remains to be an open task. In this paper, we survey privacy-preserving mechanism which allows the public auditing on data that is shared which is stored in the cloud.

Here we surveyed PDP which neither provided identity privacy nor let the data remain a secret. Though it was a breakthrough paper for the explanation of the Public auditability of shared data. We also surveyed WWRL that provided data secrecy but could not provide Identity Privacy.Here we found that the third paper that was surveyed, ORUTA, the first privacy-preserving mechanism of public auditing of the shared data in the cloud, is better than all other approaches discussed in this paper. Oruta uses ring signatures for building the homomorphic authenticators, so the TPA audits the integrity of shared data, yet cannot distinguish between the signer on each block, which can accomplish identity privacy. There is another problem also of allowing large groups was solved in KNOX but it did not allow for the public verifiability

An interesting problem for future work is how to powerfully audit the integrity of data shared here with dynamic groups while still preserving the uniqueness of the signer on each block from the third party auditor.

REFERENCES

1. 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, April 2010.

2. G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable Data Possession at Untrusted Stores,” in Proc.ACM CCS , 2007, pp. 598–610.
3. C. Wang, Q. Wang, K. Ren, and W. Lou, “Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,” in Proc. IEEE INFOCOM , 2010, pp. 525–533.
4. B. Wang, B. Li, and H. Li, “Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud,” Proc. IEEE Fifth Int’l Conf. Cloud Computing, pp. 295-302, 2012
5. G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, “Scalable and Efficient Provable Data Possession,” in Proc. ICST SecureComm , 2008.
6. A. Juels and B. S. Kaliski, “PORs: Proofs pf Retrievability for Large Files,” in Proc. ACM CCS, 2007, pp. 584–597.
7. H. Shacham and B. Waters, “Compact Proofs of Retrievability,” in Proc. ASIACRYPT . Springer-Verlag, 2008, pp. 90–107.
8. D. Boneh, B. Lynn, and H. Shacham, “Short Signature from the Weil Pairing,” in Proc. ASIACRYPT . Springer-Verlag, 2001, pp. 514–532.
9. Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, “Enabling Public Verifiability and Data Dynamic for Storage Security in Cloud Computing,” in Proc. European Symposium on Research in Computer Security. Springer-Verlag, 2009, pp. 355–370.
10. C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, “Dynamic Provable Data Possession,” in Proc. ACM CCS , 2009, pp. 213–222
11. 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 , 2011, pp. 1550–1557.
12. D. Boneh, C. Gentry, B. Lynn, and H. Shacham, “Aggregate and Verifiably Encrypted Signatures from Bilinear Maps,” in Proc. EUROCRYPT. Springer-Verlag, 2003, pp. 416–432.
13. B. Wang, B. Li, and H. Li, “Knox: Privacy-Preserving Auditing for Shared Data with Large Groups in the Cloud,” in Proc. ACNS. Spring- Verlag, 2012.