RESEARCHERID
THOMSON REUTERS

[Pankaj, 6(5): May 2019]

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

Remote Data Integrity Schemes: A Survey

Vidyullata Vinayak Devmane
Associate Professor, Computer Engineering Department, Shah & Anchor Kutchhi Engineering College Mumbai

ABSTRACT

Cloud storage is an important service of cloud computing, which allows data owners to move data from their local computing systems to the Cloud. With the high costs of data storage devices as well as the rapid rate at which data is being generated it proves costly for enterprises or individual users to frequently update their hardware. Apart from reduction in storage costs data outsourcing to the cloud also helps in reducing the maintenance. Another reason is that data owners can rely on the Cloud provider which is a more reliable service, so that they can access data from anywhere and at any time. Individuals or small-sized companies usually do not have the resource to keep their servers as reliable as the Cloud does. Cloud storage moves the user's data to large data centers, which are remotely located and on which user does not have any control. However, this feature of the cloud poses many new security challenges which are required to be clearly understood and resolved. From this survey, it has been investigated that, one of the important concerns that needs to be addressed is to assure the customer with the data integrity i.e. correctness of owners data in the cloud storage by designing a framework for efficient, secure and fully dynamic remote data integrity scheme. The objective of this survey is to provide a knowledge of data proving techniques, which will be helpful for the research in this area.

Keywords*:* data Data integrity checking methods in cloud storage, Provable data possession in cloud storage

I. INTRODUCTION

More and more data owners have started hosting their data in the Cloud. The main reason is cost effectiveness, which is particularly true for small and medium-sized businesses. By hosting their data in the Cloud, data owners can avoid the initial investment of expensive infrastructure setup, large equipments, and daily maintenance cost. The data owners only need to pay the space they actually use.By hosting data in the Cloud, it introduces new security challenges. Firstly, data owners would worry for their data could be misused or accessed by unauthorized users. Secondly, the data owners would worry for their data could be lost in the Cloud. Some recent data loss incidents are the Sidekick Cloud Disaster in 2009 [1] and the breakdown of Amazon's Elastic Compute Cloud (EC2) in 2010 [2].Sometimes, the cloud service providers may be dishonest and they may discard the data which has not been accessed or rarely accessed to save the storage space. Moreover, the cloud service providers may choose to hide data loss and claim that the data is correctly stored in the Cloud. As a result, data owners need to be convinced that their data are correctly stored in the Cloud.

The Third party auditing (TPA) is the only choice for the storage auditing, as they will have expertise and capabilities,they can do more efficient work and convince both the cloud service provider and the data owner[7][13]. For the Third party auditing, the system model contains three types of entities: data owners, the cloud server and the

ISSN 2348 – 8034
Impact Factor- 5.070

third party auditor. Data storage auditing is a very resource demanding operation in terms of computational resources, memory space, and communication cost. But the methods for the data owner auditing may not be directly used in the TPA due to the reasons like: (1) Sending data to the third party auditor can leak out the data owners' data, and (2) The third party auditor has no responsibility to store all the metadata.

A representative network architecture for cloud data storage can have three different network entities as follows[7]:

- Client: an entity, which has large data files to be stored in the cloud and relies on the cloud for dat maintenance and computation, can be either individual consumers or organizations.
- Cloud Storage Server (CSS): an entity, which is managed by Cloud Service Provider (CSP), has significant storage space and computation resource to maintain the clients' data.
- Third Party Auditor : an entity, which has expertise and capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request.

Possible Threats to the Data in Cloud Storage

- ·Replay Attack: The Server generates the proof from the previous proof or other information, without querying the actual Owner's data.
- ·Server colluding attack or forge Attack: The Server may forge the metadata of data block and deceive the Auditor.
- Byzantine Failure: During the execution of an algorithm in a distributed system when this failure occurs the system may respond in any unpredictable way.
- Data lost auditing pass attack: The malicious cloud server loses the outsourced data for some internal external reasons and can pass the auditing from the TPA.

II. LITERATURE SURVEY

The different phases of software development demand various different automation tools in the cloud environment. For example, to enable automatic job configure, schedule and monitor tools such as Puppet, Juju, Apache Continuum, Jenkins, and Cobbler enable automatic job are used in different phases [3]. To improve quality and address customer issues rapidly, we require a tool which is capable of continuous integration, test-driven development, and debugging.

Public auditability : to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional online burden to the cloud users.

Storage correctness : to ensure that there exists no cheating cloud server that can pass the TPA's audit without indeed storing users' data intact.

Privacy preserving : to ensure that the TPA cannot derive users' data content from the information collected during the auditing process.

Batch auditing : to enable TPA with secure and efficient auditing

ISSN 2348 – 8034
Impact Factor- 5.070

Dynamic operation support: to allow the clients block-level or file-level (if possible) operations on the data files while maintaining the same level of data correctness assurance.

Lightweight: Cryptographic algorithms should not add an extra burden with respect to time, bandwidth, computation cost etc.

Ateniese et.al.[3] aims to work for authenticity of archival data on cloud storage, archival storage servers retain tremendous amount of data, little of which is accessed. They also hold data for long periods of time during which there may be exposure to data loss from administrative errors as the physical implementation of storage,e.g., backup and restore, data migration to new systems.

Accessing an entire file for checking it's integrity is expensive in I/O costs to the storage server and in transmitting the file across a network. So Ateniese et.al. suggested the methodology that clients need to be able to verify that a server has retained file data without retrieving the data from the server and without having the access to the entire file.

At a later time, the client can verify that the server possesses the file by generating a random challenge against a randomly selected set of file blocks. Using the queried blocks and their corresponding tags, the server generates a proof of possession. The client is thus convinced of data possession, without actually having to retrieve file blocks**.** However, the precomputation of the tags imposes heavy computation overhead that can be expensive for an entire file[10].

Ateniese et.al. have changed the methodology in their further work than the above mentioned work, this work[4] secure provable data possession(PDP) technique based entirely on symmetric key cryptography and authors says that any bulk encryption is not requiring. This PDP technique allows outsourcing of dynamic data which was not possible in previous work as it supports operations, such as block modification, deletion and append.

However, since this work[4] is based upon symmetric key cryptography, it is unsuitable for public (third-party) verification. They suggested a solution like a hybrid scheme of combining elements of [3] and [4] schemes.

The method in[4] has lower overhead than their previous scheme and allows for block updates, deletions, and appends to the stored file. However, their scheme focuses on single server scenario and does not provide data availability guarantee against server failures, leaving both the distributed scenario and data error recovery issue unexplored[10].

Ateniese et.al.[5] have provided a framework for building public-key HLAs. Here in identification protocol, a prover P is possessing of a secret key s_k to prove its identity to a verifier V that possesses the corresponding public key p_k. They consider identification protocols where the prover generates the first message alpha using the public key p_k and randomness r; the verifier sends a random challenge beta and the prover then computes a response using (p_k; s_k), the randomness r, and the verifier's challenge beta, Given the transcript of the protocol, the verifier decides whether to accept or not.

The authors in [5] are the first to propose a partially dynamic version of the prior proof of data possesion scheme, using only symmetric key cryptography but with a bounded number of audits[11]. [17]Though HLA based method allowing efficient data auditing and consuming only constant bandwidth, the direct adoption of these HLA-based

ISSN 2348 – 8034
Impact Factor- 5.070

techniques is still not suitable as the linear combination of blocks may potentially reveal user data information to TPA, and violates the privacy-preserving guarantee.

Ateniese et.al.[6] in their scheme uses homomorphic verifiable tags. Because of the homomorphic property, tags computed for multiple file blocks can be combined into a single value. The client precomputes tags for each block of a file and then stores the file and its tags with a server. At a later time, the client can verify that the server possesses the file by generating a random challenge against a randomly selected set of file blocks. The server retrieves the queried blocks and their corresponding tags, using them to generate a proof of possession. The client is thus convinced of data possession, without actually having to retrieve file blocks.

The authors of [7] aim to achieve storage correctness insurance and data error localization i.e. the identification of misbehaving servers. They also took in to considerations of dynamic operations on data blocks including data update, delete and append. They also proposed the system which will be resilient against Byzantine failure, malicious data modification attack and server colluding attacks.

As [4],[7] do not support the privacy protection of data against external auditors so the authors of [8] gave an effective third party auditing fundamental requirements,cryptanalysis done by authors Kan Yang ,Xiaohua Jia in[19],

- Due to large number of data tags their auditing protocols may incur heavy storage overhead on the server.
- It's vulnerable for eavesdrop on the data and forge a great deal of data.
- Outsider attacker can intercept the data sent from the user to the cloud server in TagBlock step and modify it arbitrarily.

In the cryptanalysis the author of[14] shows that the public auditing scheme proposed by Wang et al. can not resist against existential forgery using a known message attack. Moreover, they have shown that the protocol is vulnerable for attacks by a malicious cloud server and an outside attacker through four specific attacking schemes. And the results shows that the protocol can not provide secure data storage for users[13].

Qian Wang et.al.in[8] proposed methodologies requires large number of data tags, their auditing protocols may incur a heavy storage overhead on the server. Different methods to ensure remote data integrity achieving the public auditability and dynamic data operations are given by Qian Wang et.al.in [9] in their further work. Combined BLS-based HLA with MHT to support fully dynamics and Classic Merkel-Hash tree construction for block tag authentication. used by them.

This ensures the storage correctness without possession of local data by users but focusing on single server scenario models given in[10][16] may leak content to the auditor because it requires the server to send the linear combinations of data blocks to the auditor.

Wang et.al.in [10] aims to provide flexible distributed storage integrity auditing mechanism which will allow users to audit the cloud storage with very lightweight communication and computation cost. The auditing results will not only ensures strong cloud storage correctness guarantee, but also simultaneously it can achieve fast data error localization, i.e., the identification of misbehaving server. Considering the cloud data are dynamic in nature, the proposed design should further support secure and efficient dynamic operations on outsourced data, including block modification, deletion and append.

Erasure-correcting code in the file distribution preparation to provide redundancy parity vectors and guarantee the data dependability is done in this work. By utilizing the homomorphic token with distributed verification of erasure

coded data, their scheme achieves the integration of storage correctness insurance and data error localization. Data corruption can be detected during the storage correctness verification across the distributed servers is said by authors. They say that their scheme is highly efficient and resilient to Byzantine failure, malicious data modification attack, and server colluding attacks.

Wang et al. concludes in[11] that, the system proposed in [10] supports for partially dynamic data storage in a distributed scenario.

Cong Wang et.al in [11] aims to achieve privacy-preserving public auditing, they integrated the HLA with random masking technique. They have used Bilinear property of bilinear pairing instead of mask techniques. This work is extended work of the work done by [8]. In their protocol, the linear combination of sampled blocks in the server's response is masked with randomness generated by the server. With random masking, the TPA will not have all the necessary information to build up a correct group of linear equations and therefore cannot derive the user's data content, no matter how many linear combinations of the same set of file blocks can be collected. Proofs for batch auditing and data dynamics are given but full-fledged implementation of the mechanism on commercial public cloud is not done in this work.

Wenjun Luo, Guojing Bai[12] used HLAs and RSA construction to complete their protocol. The support of public verifiability makes the protocol very flexible, since the user can commit the data

possession to check the TPA.The protocol is based on the RSA problem with large public exponent to enhance the security of the data storage.

The protocol to supports public verifiability without help of a third-party auditor has been implemented with RSA-based homomorphic verifiable tags (HVT) by Zhuo Hao et al. in[16].This protocol is not supporting to data level dynamics. The difficulty is that there is no clear mapping relationship between the data and the tags. Whenever a piece of data is modified, the corresponding blocks and tags can be updated. But,this can bring unnecessary computation and communication costs. All the programs are written in the C++ language with the assistance of Multiprecision Integer and Rational Arithmetic C/C++ Library MIRACL library.

Kan Yang, Xiaohua Jia in [17] have given the solution for the data privacy problem, their method is to generate an encrypted proof with the challenge stamp by using the bilinearity property of the bilinear pairing, such that the auditor cannot decrypt it, but that improves the system performance. By using the homomorphic verifiable tags, no matter how many data blocks are challenged, the server only responses the sum of data blocks and the product of tags to the auditor, whose size is constant and equal to only one data block. Thus, it reduces the communication cost.

Kan Yang ,Xiaohua Jia in [19] say that if the parameters for generating the data tags used by each owner are different then they cannot combine the data tags from multiple owners to conduct the batch auditing.

ISSN 2348 – 8034
Impact Factor- 5.070

III. Conclusion and Challenging issues of data storage auditing

Most of previous works deal with static data which leads to a security flaw when it is tried to apply on a dynamic environment. Some of the schemes with public verifiability lack data privacy preserving. In designing public auditing protocol which is dynamic, unbounded in use of queries and also privacy preserving great care must be taken on its efficiency and security. Hence that designing efficient, secure and fully dynamic remote data integrity scheme is quit challenging.

Efficiency of the protocol is based on the mathematical model used in it. Different mathematical models gives different level of security. The drawback of the BLS method is that, the server needs to send the linear combination of all the challenged data blocks to the auditor and hence data leakage may be easy to the auditor. The drawback of erasure coding is it can be more CPU-intensive, and that can translate into increased latency. Those protocols are using mask technique to ensure data privacy will not be suitable in multicloud batch auditing. Most of the authors in their protocol design have used RSA homomorphic algorithm but it requires bigger size of key, larger size of data tags and operations will be computationally too costly. Due to larger size of data tags in auditing protocol may incur a heavy storage overhead on the server.

Elliptic Curve Cryptosystem (ECC), based on the discrete logarithm problem over points on an elliptic curve. It can be best alternative to RSA and other logarithmic models, as it offers a much shorter word length for a given strength. For example, ECC with a key size of 128—256 bits can offer equal security to that of RSA with key size of 1—2 Kbits. Elliptical curve discrete logarithmic problems are harder than the logarithmic problems and so it can be used to get better security with shorter key size. A number of software implementations of ECC have been reported previously, the advantages of software implementations include ease of use, ease of upgrade, portability, low development cost and flexibility.

REFERENCES

1. Cellan-Jones, R.: The Sidekick Cloud Disaster. BBC News, vol. 1 (2009)
2. Miller, R.: Amazon addresses EC2 power outages. Data Center Knowledge 1 (2010).
3. Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., Song, "Provable data possession at untrusted stores", Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS '07, pp. 598–609. ACM, New York, NY, USA (2007).
4. Ateniese, G., Di Pietro, R., Mancini, L.V., Tsudik, "Scalable and efficient provable data possession", Proceedings of the 4th International Conference on Security and Privacy in Communication Networks, SecureComm '08, pp. 9:1–9:10. ACM, New York, NY, USA (2008).
5. Ateniese, G., Kamara, S., Katz, "Proofs of storage from homomorphic identification protocols",Proceedings of the 15th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology, ASIACRYPT '09, pp. 319–333. Springer, Berlin, Heidelberg (2009).
6. Giusepp Ateniese,"Remote data checking using provable data possession", ACM Transaction on information system security,2011.
7. C.Wang, Q.Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing," in Proc. of IWQoS'09, July2009,pp. 1–9.
8. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing, "in InfoCom2010, IEEE, March 2010.

9. Qian Wang, Cong Wang, Kui Ren, Wenjing Lou and Jin Li, " Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", IEEE Transactions on Parallel and Distributed Systems, Vol. 22, No. 5, May 2011.
10. Qian Wang, Cong Wang, Kui Ren, Wenjing Lou and Jin Li , "Toward Secure and Dependable Storage Services in Cloud Computing", IEEE Transactions on services computing Vol. 5, No. 2, April-June 2012.
11. Qian Wang, Cong Wang, Kui Ren, Wenjing Lou and Jin Li , "Privacy-preserving public auditing for Secure Cloud Storage", IEEE Transactions on Computers,Vol.62,No.5,Feb.2013.
12. Wenjun Luo, Guojing Bai," Ensuring the data integrity in cloud data storage", Proceedings of IEEE CCIS2011.
13. Solomn Guadie worku, Zhong Ting, Qin Zhi-Guang, "Survey on Cloud Data Integrity Proof Techniques", Seventh Asia Joint Conference on Information Security, 2012.
14. XU Chun-xiang, HE Xiao-hu, Daniel Abraha,"Cryptanalysis of auditing protocol proposed by Wang et al. for data storage security in Cloud Computing",University of Electronics Science and Technology of China 2006.
15. Spring Semester 2014 Cryptography 2 (2WC13) / Cryptographic Protocols 1 (2WC17)Lecture Notes Cryptographic Protocols Version 1.0, February 3, 2014 Berry Schoenmaker Department of Mathematics and Computer Science,Technical University of Eindhoven, P.O. Box 513, 5600 MB Eindhoven,TheNetherlands.berry@win.tue.nl,l.a.m.schoenmakers@tue.nl,www.win.tue.nl/~berry/2WC13/ www.win.tue.nl/~berry/2WC17/
16. Zhuo Hao, Sheng Zhong, Nenghai Yu, "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability", IEEE Transaction on Knowledge and Data Engineering, Vol. 23, No. 9, September 2011.
17. Kan Yang, Xiaohua Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing", IEEE Transaction on Parallel and Distributed Systems, Vol. 24, No. 9, September 2013.
18. Kan Yang ,Xiaohua Jia, "Data storage auditing service in cloud computing:challenges, methods and opportunities", Springer Science+Business Media, LLC 2011.
19. Kan Yang ,Xiaohua Jia, "Security for Cloud Storage Systems", ISSN 2191-5768,Springer,Year-2014.
20. Giuseppe Ateniese, Seny Kamara, Jonathan Katz, "Proofs of Storage from Homomorphic Identification Protocols", Research supported by NSF grant #042668,Year 2009.