International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 1, January - 2014

A Survey on Privacy Preserving Auditing In Cloud Storage

Vaishnav. S
PG scholar/Department of IT,
Karunya University, Tamilnadu, India.

Esther Daniel
Assistant Professor/ Department of IT,
Karunya University, Tamilnadu, India.

Abstract

Information storage demands for users is increasing each day. Users now want to utilize remote storage facilities such as the cloud. In the cloud there is no direct physical control over the data because the cloud uses its resource pool for the storage. Therefore data integrity protection and auditing is not a simple task. The user needs to depend on a Third Party Auditor (TPA) who is employed as a public auditor for verifying the data integrity and privacy. This paper compares different auditing techniques, using different parameters such as security, computation and communication cost, to improve the data dynamics and better performance.

Key Words: **Cloud computing, Storage auditing, dynamic auditing, privacy-preserving auditing, batch auditing.**

1. Introduction

The cloud is an emerging technology which provides software as a service, platform as a service, also infrastructure as a service and many more. Cloud computing is a utility computing where the resources are available on demand from a resource pool [1]. These are basically the virtual servers which are available over the internet. Wikipedia defines cloud computing as a real time connection of different communication networks. Therefore it is also described as distributed computing and provides the ability to run a program or application on different computers at the same time. This is commonly referred to as network-based services, which is provided by real hardware, and is served by virtual hardware, which is simulated by software running on one or more real machine.

Figure 1. Cloud resource sharing

The cloud provides all these things from the data center, also called the resource pool, so that anyone can access it within or outside the cloud storage. The Cloud Service Provider (CSP) is responsible for providing the services. However, sometimes the CSP does not behave properly towards the users. For example, the CSP will not report to the user about changes in the outsourced data.

Traditional ways of privacy preserving methods using cryptographic methods such as hash function and the digital signature scheme won't work effectively. The auditing process is not a simple task because it is not a practical solution for the users to have a local copy of the data stored at the user site while auditing. Also, there may be many users accessing one resource in the cloud in an enterprise. So it may not be able to be verified by a user. The new schemes allow public auditability, therefore it is not necessary to verify the data only by the owners. Any trusted party which is a public agent can verify its integrity.

Protocols such as POR [2] and PDP [3] which provide some of the simplest verification processes, that is, verification without downloading whole files, which

IJERT

reduces the computation complexity of the client. To ensure the data integrity and also to reduce the online burden we are depending on the Third Party Auditor for the auditing of the outsourced data. The TPA [3], [9], [10] has expertise in this and also has the capability to audit multiple data simultaneously. It can periodically verify the files at a fixed time interval. The periodic verification of the user's data greatly reduces the integrity verification tasks. This paper explains different privacy preserving auditing schemes to increase the performance and auditing techniques.

The sections of this paper are organized as follows: Chapter 2 covers the literature survey. Chapter 3 explains the comparison tables by comparing the results of the different auditing schemes. The final section gives the conclusion for the auditing schemes in cloud for increasing performance.

2. Privacy preserving auditing schemes

Cloud computing allows the use and sharing of large amount of data. In most of the schemes it is based on the following elements:

The cloud user, who has a large amount of data to be stored in the cloud servers. The user has to have the capability to update, delete, and modify the data on the server.

The cloud servers (CS), which possess a large amount of storage space where the data can be stored. This is controlled by the cloud service providers.

*The third party auditor (TPA), i*s the expert in the auditing of large data and also supports the multiple auditing. This section describes the various techniques and protocols used to improve this auditing process while maintaining the security and privacy.

2.1 Demonstrating data possession and uncheatable data transfer

Filho et al., describes protocols based on hash functions [6] with RSA-based secure hash functions. This prevents cheating during the transfer of data. It also reduces the burden on the user by using a trusted third party. This RSA-based hash function also includes cryptographic and elliptic curve cryptography. The protocol is as secure as that of the integer factoring. The advantage is the use of a public key for data protection. It is very flexible with no fixed message size and is easy to implement. The digital signature method [7] in RSA is used to protect the privacy and integrity of the outsourced data in the cloud environment and for the message authentication.

It is restricted if the same data is stored by multiple network users. The performance is slow because it requires 1 to 2 modular multiplications per bit.

2.2 Efficient remote data possession checking in critical information infrastructures

This scheme uses a remote data possession checking protocol [9] which was proposed by F. Sebe´ et al., which allows checking an uncorrupted file that a server can access remotely, and where the verifier does not need to have prior knowledge of the entire file. It is the first protocol that supports an infinite number of verifications. The integrity checking is done using the super file. A super file is defined as a set of files ordered in an arranged manner. Once the super file is identified, the modifications can be easily done.

Unlimited verification is one of the main features. Here the verifier does not want to store the complete files. The disadvantage of this is that the files stored on the server are in bits so it takes time to store the information. This also takes extra time when there is a huge quantity of data to store. In fact there is a small possibility of revealing some data to the auditor.

2.3 Provable data possession at untrusted stores

In this scheme G. Ateniese et al., used a provable data possession with homomorphic verifiable tags [4]. It allows the verification of data without retrieving it from the original source. The model generates probabilistic proofs of possession by sampling random set of blocks of data from the server, which reduce the cost.

The homomorphic verifiable tags computes multiple file blocks which can be combined to form a single file. The client pre-computes the tags and the tags are stored in the Third Party Auditor for verification. The modified file is stored in the server storage. The verification process is done in the requested style generated by the client.

It performs well and supports blockless verification. Its client/server computation is in O (1). Verification and communication takes time. It does not consider the privacy protection of the user's data against the external auditors.

IJERT

2.4 Dynamic provable data possession

C. Erway et al., explained about the Dynamic Provable Data Possession (DPDP) [8]. PDP is mostly applicable for static files. The DPDP is an updated version of the PDP where it supports the updates while storing the data. It can append, modify, or delete the existing blocks of files. This scheme uses rank information to organize the dictionary entities. It supports the verification of files for different users and does not need to download the whole file for verification. It also explains the security and blockless verification of DPDP. Its hashing schemes use ranks based RSA trees. The experimental results show that the block size minimizes the communication and computational overhead.

2.5 Privacy-preserving public auditing for data storage security in cloud computing

C. Wang et al., used a scheme based on a homomorphic authenticator which is uniquely integrated with random masking technique for preserving privacy in auditing. The homomorphic authenticator [10] is the metadata generated from individual data blocks. The bilinear aggregate supports a linear combination of data blocks which can used to handle multiple auditing tasks. These linear combinations of blocks are masked with the pseudo random function (PRF). With this random masking the TPA cannot access the data. This scheme also includes a bilinear map which consists of two phases: setup phase and auditing phase.

It also supports batch auditing where multiple auditing tasks from different users can be simultaneously carried out. The data dynamics supports efficient and scalable auditing in the cloud environment. This motivates public auditing where an external auditor can perform the auditing tasks without the knowledge of the file content. The lightweight process allows the TPA to perform auditing with minimum communication and computation cost.

2.6 PORs: proofs of retrievability for large Files

In this paper A.Juels et al., defined the PORs [11] as using an archive or a backup to help the verifier retrieve the file in the target easily. The user can easily retrieve the file from the backup. The POR is viewed as a kind of cryptographic proof of knowledge (POK), which can support large files. POR protocol reduces the communication cost because it doesn't need to access the file from the server, it can easily be accessed from the archive. This PORs is an unusual security formulation.

The main goal of PORs is that they are used to check the file without downloading the files. It also provides quality of service. Here the pre-processing takes time i.e., encoding the file F is required before storing to the prover. At the time of encoding sentinels are randomly added in specific positions, to constitute the contents of a POR. These sentinels can also retrieved by using the PIR, and it can be reused. It does not consider the privacy of the data against the external auditors. It has computational overhead.

2.7 Efficient audit service outsourcing for data integrity in clouds

The efficient audit service outsourcing by Y. Zhu et al., reduced the storage maintenance and management by providing scalability, low-cost, and location independent platforms. It is based on the interactive zero knowledge proof system and the interactive provable data possession [12] which are used to prevent the leakage of data and also to prevent fraudulence of the storage. This schema also concerns the cost of computation, communication and the storage, as well as the scheduling of the audit process. The audit cost is reduced by the periodic verification and probabilistic queries.

This scheme improves the performance of the audit services and reduces the storage and the network overheads, and also the workload on storage servers. It provides cost effective services and also minimizes the computational overhead. Here the batch auditing process is not utilized for the efficient computation. But for high efficient verification, a periodic verification is also done at regular intervals. The results show that the computation and the communication costs are increasing with that of file size and sampling ratio.

2.8 Privacy-preserving public auditing for storage security in cloud storage

C.Wang et al., proposed a privacy preserving public auditing scheme [15] which is shown in figure 2. This auditing technique consists of four algorithms (KeyGen, SigGen, GenProof, verifyProof). This public auditing system consists of 2 phases Setup and Audit.

Setup: The user first initializes the public and secret keys of the system by executing KeyGen, and pre-processes the data file F by using SigGen to generate the verification metadata. The user then stores the

IJERT

Table 1. Comparison of auditing schemes

Scheme	Computation		Commu-nication	Privacy	Dynamic	Sampling	Prob. Of detection
	CSP	Client					
PDP [3]	O(t)	O(t)	O(t)	Yes	No	No	$1-(1-p)^t$
SPDP [4]	O(t)	O(t)	O(t)	Yes	No	No	$1-(1-p)^t$
DPDP-I [6]	O(t log n)	O(t log n)	O(t log n)	No	No	No	$1-(1-p)^t$
DPDP-II [6]	O(t log n)	O(t log n)	O(t log n)	No	No	No	$1-(1-p)^t$
CPDP [2] [9]	O(t+s)	O(t+s)	O(t+s)	No	No	Yes	$1-(1-p)^{ts}$
IPDP [11] [12]	O(ts)	O(t+s)	O(t+s)	Yes	Yes	Yes	$1-(1-p)^{ts}$
PPPA [8] [13]	O(t log n)	O(t log n)	O(t log n)	Yes	Yes	Yes	$1-(1-p)^t$

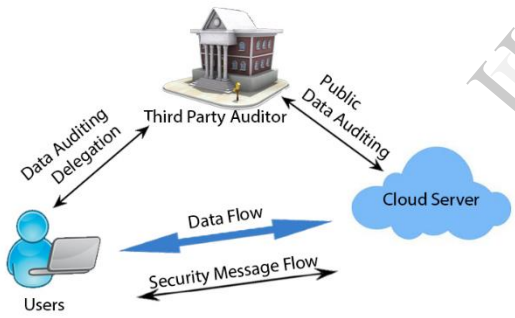

Figure 2. Cloud architecture for secure data storage

metadata and the data to the cloud server. Then it deletes the copy of the data on the user side.

Audit: When the user request the TPA for the verification. The TPA issues an audit message or challenge to the cloud server. The cloud server then sends its metadata to the TPA and the verification starts. The cloud server will generate a response message by executing GenProof using the file F and its verification metadata as inputs. The TPA then verifies the response via VerifyProof by comparing the metadata.

This scheme provides better security by splitting the file from the user side itself and encrypting the data before outsourcing to the cloud storage. This keeps the data secure from the cloud server. The metadata generated from the user side is sent to the TPA. This also provides security to the user data. However, the use of batch auditing reduces the TPA's computational cost, as more than 15 percent of the per task auditing.

3. Comparison Table

The table 1. Shows the detailed comparison between the different schemes and their parameters such as communication, computation of the service provider and the client, privacy, data dynamics etc., here the 'n' is the total number of data blocks of a file, 't' is the number of challenged data blocks in an auditing, Query 's' is the number of sectors in each data block, 'ρ' is the probability of block or sector corruption.

4. Conclusion

Cloud computing is a technology we can use similar to that of utility computing which provides enough storage for the resources. This survey considered parameters like computation and

communication complexity, privacy, data dynamics through different schemes. Some of the schemes try to improve some parameters and some others try to reduce the probability of detection which improves the performance. This paper focused on the privacy preserving techniques on the cloud. By comparing the different schemes and their results we conclude that the Privacy Preserving Public Auditing is better than the rest of the schemes. The Privacy Preserving Public Auditing provides public auditing so that the external auditor can audit the user's data without leakage of data from the TPA. In batch auditing, multiple simultaneous auditing tasks can be performed by the TPA. This also provides security and privacy for improving the performance of auditing.

5. 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, "Above the Clouds: A Berkeley View of Cloud Computing," *Technical Report UCB-EECS-2009-28*, Univ. of California, Berkeley, Feb. 2009.

[2] H. Shacham and B. Waters, "Compact Proofs of Retrievability," *Proc. Int'l Conf. Theory and Application of Cryptology and Information Security: Advances in Cryptology (Asiacrypt)*, vol. 5350, Dec. 2008.

[3] M.A. Shah, R. Swaminathan, and M. Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," *Cryptology ePrint Archive*, Report 2008.

[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)*, 2007.

[5] Ateniese, G., Pietro, R.D., Mancini, L.V., Tsudik, G., 2008. "Scalable and efficient provable data possession." *In: Proceedings of the 4th International Conference on Security and Privacy in Communication Networks*, SecureComm.

[6] D. L. G. Filho and P. S. L. M. Baretto. "Demonstrating data possession and uncheatable data transfer". *IACR ePrint archive*, 2006. Report 2006.

[7] K Govinda, V. Gurunathprasad and H. sathishkumar, "Third Party Auditing for Secure Data Storage in Cloud Through Digital Signature Using RSA", *International Journal of Advanced science and Technical Research*, vol 4,no. 2, August 2012.

[8] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession," *Proc. ACM Conf. Computer and Comm. Security* (CCS '09), 2009.

[9] F. Sebe´ , J. Domingo-Ferrer, A. Marti´nez-Balleste´ , Y. Deswarte, "Efficient Remote Data Possession Checking in Critical Information Infrastructures," *IEEE transactions on knowledge and data engineering*, vol. 20, august 2008.

[10] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Storage Security in Cloud Computing," *Proc. IEEE INFOCOM '10*, Mar. 2010.

[11] A. Juels and J. Burton, S. Kaliski, "PORs: Proofs of Retrievability for Large Files," *Proc. ACM Conf. Computer and Comm. Security (CCS '07)*, Oct. 2007.

[12] Yan Zhu, Hongxin Hu, Gail-Joon Ahn, Stephen S. Yau, "Efficient audit service outsourcing for data integrity in clouds," *The Journal of Systems and Software 85* (2012).

[13] Y. Zhu, H. Hu, G. Ahn, and M. Yu, "Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage," *IEEE Trans. Parallel and Distributed Systems*, vol. 23, no. 12, Dec. 2012.

[14] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S.S. Yau, "Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds," *Proc. ACM Symp. Applied Computing*, W.C. Chu, W.E. Wong, M.J. Palakal, and C.-C. Hung, eds., 2011.

[15] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Storage Security in Cloud Storage," *IEEE transactions on computers*, vol. 62, February 2013.

IJERT