A Survey on Secure Cross user Deduplication Techniques on Cloud Storage

Shruthi P. S
M-Tech Dept. of Computer Science & IT
Dayananda Sagar University,
Bangalore, India

Dr. Reeja S. R
Associate Professor, Dept. of Computer Science & Engg
Dayananda Sagar University,
Bangalore, India

***Abstract*— Cloud computing is Associate in Nursing innovative proficiency within the field of data and experience. It provides such a lot of things in terms of "As-A-Service" basis. Cloud Computing is that the long unreal visual image of computing as a utility, wherever users will tenuously store up their records into the cloud therefore on getting pleasure from the on-demand high-quality applications and services from a shared assortment of configurable computing assets. It suggests that of outsourcing records, users may be alleviated as of the difficulty of restricted information storage and maintenance. on the contrary, the understanding that users now may not have physical management of whereby all likelihood of huge volume of outsourced data to create the data reliability enrichment. In Cloud Computing associate degree exceptionally is powerful and doubtless alarming task. For cloud storage, privacy and security are the blazing issues. just the once the user hoard their knowledge on the cloud consequently at hand may be an intimidation of trailing the data, or every so often data may be adapted or restructured. It may not be totally reliable as a result of the consumer doesn't have a duplicate of all stored data within view of the fact to facilitate cloud storage moves the user's records to an outsized data center, which are vaguely located, on which user do not have any control. As the cloud computing technology developed throughout the last decade, outsourcing information to cloud service for storage become an eye-catching development, which profits in precautious hard work on heavy data maintenance and management. On the other hand, since the outsourced cloud storage isn't absolutely trustworthy, it raises security considerations on a way to appreciate information deduplication during a cloud whereas achieving integrity. The intention is on gaining reciprocally the information integrity and deduplication at the same time and notifying the users relating to the modification of information on the cloud.**

Keywords*— *Deduplication, Homomorphic, Pos, Integrity,

I. INTRODUCTION

Cloud computing will be the core of information infrastructure in future. It provides all kinds of services for the users. The considerable service happening by the cloud is nothing but storage competence. Storage outsourcing is fetching more and more attractiveness to both industry and academic outstanding to the advantages of low cost, ease of access, and easy distribution. As one of the storage outsourcing type, cloud storage gains broad interest in recent years. several corporations, such as Amazon, Google, and Microsoft, offer their own cloud storage space for storing services, where users can transfer their files to the servers, access them from a spread of devices, and share them with the others. even though cloud storage service area unit are extensively adopted in current days, there still keep behind several security problems and potential threats.

One vital challenge of cloud storage services is the administration of the ever-increasing degree of data. Thanks to conjure data managing ascendable and still on to diminish the amplified quantity of data within the cloud ,deduplication has been a well-known procedure. Data deduplication is to be found fascinated to follow by that's meant for eliminating duplicate copies of repetitive data in storage. There area unit series of act utilized in the track to advance storage operations and competency of cloud storage. deduplication permits to avoid wasting space for storing and minimize redundant data. Here all the method through the revelation of sequential duplicate of data is stored only one time and we keep pointers or tips to the particular data. Encryption entail numerous users to encrypt their information with their personal keys. Thus, not possible to totally differentiate data copies of numerous users which will guide to different cipher texts, making deduplication impractical.

Deduplication will take consign at what is more the folder level or the chunk level. For file-level deduplication, it eliminates duplicate copies of the identical file. Deduplication also can turn up at the chunk level, that eliminates spare blocks of data that occur in non-identical files. although data deduplication brings a bunch of compensation, security and privacy considerations arise as users sensitive data are vulnerable to both inside business executives and outsider attacks.

Data integrity is one in every of the foremost vital properties once a user outsources its files to cloud storage. Users ought to be convinced that the files hold on the server havent tampered. Thus, researchers introduced Proof of Storage (PoS) for checking the integrity while not downloading files from the cloud server.
moreover, users may oblige many dynamic operations, cherish modification, insertion, and deletion, to update their files, where as maintaining the potential of PoS. Dynamic PoS is projected for such dynamic operations. Therefore, when vibrant operations are executed, users reinforce hash codes (which are used for reliability checking) for the reorganized blocks barely, as an alternative of regenerating for all blocks. In this scheme, for every chunk of a file hash code will be generated which is used for verifying the integrity of that block.

For illustration, a file consists of 1 thousand blocks, and a substitute block which is inserted after the second block of the file. Then, 998 block indexes of the preceding first file which are modified that implies the user is able to generate and send 999 tags for this update. For the sake of separating this challenges several structure are introduced in dynamic POSs, Result points out that the tags are hooked up to the structure rather than the block indexes.

II. MOTIVATION

Motivation supported based on the challenges discussed.

The motivation following the dynamic PoS remains to be increased in a very multi-user atmosphere because the demand of cross-user duplication on the client-side. This means that users can omit the uploading process and gain the possession of files currently as long for the reason that the uploaded files present already at intervals within the cloud server. This technique can trim down space for storing for the cloud server, and store broadcast information gauge for users.

III. LITERATURE SURVEY

The study on proof of storage was introduced by Ateniese *et al.* [1], and Juels and Kaliski [2], correspondingly. The foremost proposal of PoS is to without aim for some data blocks as the check. after, the cloud server ensues the challenged data blocks and their tags because as the retort. Given that the information blocks and the tags can be often amalgamated via homomorphic functions, the communication overheads are square measure reduced. The ensuing workings [3] [4] [5] [6] [7] [8] [9] [10] [11] extended the analysis of PoS, except those mechanisms didn't take any dynamic operations into consideration.

Halevi *et al.* [12] introduced the thought of proof of ownership that may be the key of cross-user deduplication on the client-side. It needs that the user will produce the Merkle tree while not with the help from the cloud server, that may be is a huge challenge in dynamic PoS.

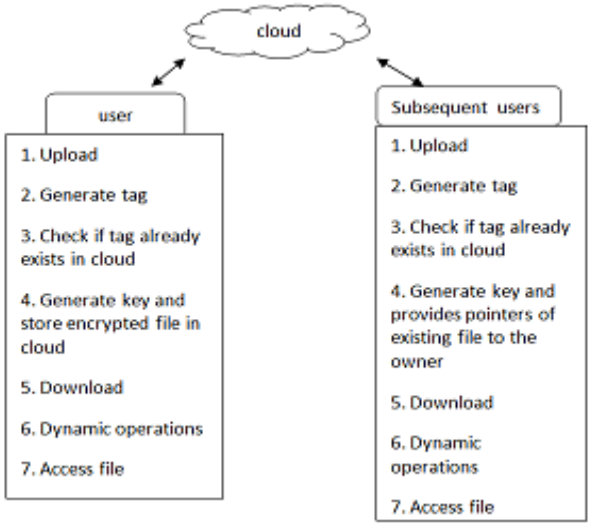
Pietro and Sorniotti [13] projected another proof of ownership theme that improves the potency. Xu *et al.*

[14] anticipated a client-side deduplication style for encrypted data, however theme employs a settled proof algorithm that indicates to facilitate each and every file has a deterministic settled short proof. As a consequence, any person who obtains this proof can get further on the authentication exclusive of possessing the file in the neighborhood. Other deduplication schemes for encrypted data [15] [16] [17] were anticipated for enhancing the security and efficiency.

Note that, all accessible techniques for cross-user deduplication on the client-side were thought of for static files. Once the files are restructured, the cloud server must regenerate the absolutely legitimate structures for these files, that causes serious computation value on the server-side.

Wang *et al.* [18] [19], and Yuan and Yu [20] measured proof of storage for multi-user updates, however those schemes center of attention was the matter of sharing files during a cluster. Deduplication in these eventualities is to deduplicate files among completely different teams. Unfortunately, these schemes cannot inhibit deduplication thanks to structure diversity and private tag generation.

IV. PROPOSED METHODOLOGY



In system model think about 2 kinds of entity: the cloud server and users.

For each file, original user is that user who uploaded the file to the cloud server, where as successive user is the user who proved the possession of the file but did not in reality upload the file to the cloud server.

There are 5 phases in a system:

1) pre-process

2) transfer

3) deduplication

4) update

5) proof of storage within the pre process phase

In the pre-process phase, users mean to transfer their native files. The cloud server decides whether or not these files be believed to transfer. If the transfer method is approved, go in the transfer phase; otherwise, go into the deduplication phase.

In the transfer phase, the files to be uploaded don't live within the cloud server.

The initial users inscribe the native files and transfer them to the cloud server. at intrervals the duplication part,half the files to be uploaded antecedently exist at intervals within the cloud server. The consecutive users hold the files domestically and thus the cloud server stores the authenticated structures of the files.

consecutive users ought to be compelled to sway the cloud server that they own the files destitute of uploading them to the cloud server. Note that, these 3 phases (pre-

process, upload, and deduplication) are dead just once within the life cycle of a file from the outlook of users.

That is these 3 phases occur only users will transfer files. If these phases conclude on the average, i.e., users endtransfering within the upload part, or they bypass the verification within the deduplication part, we are saying that the users have the ownerships of the files.

Some Advantages of Proposed System

1) The second copy files are mapped with a single copy of the file by mapping with the accessible file in the cloud

2) The inclusive supplies in multi-user cloud storage systems are introduced in the model .

V. CONCLUSION

The analysis illustrate that the implementation is proficient, especially when the file size and the number of the challenged blocks are outsized. The great provisions in multi-user cloud storage systems is introduced in the representation. A unique tool usage known as HAT that is correlated in tending inexpensive authentic configuration.

REFERENCES

[1] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in *Proc. of CCS*, pp. 598–609, 2007.

[2] A. Juels and B. S. Kaliski, Jr., "PORs: Proofs of retrievability for large files," in *Proc. of CCS*, pp. 584–597, 2007.

[3] F. Armknecht, J.-M. Bohli, G. O. Karame, Z. Liu, and C. A. Reuter, "Outsourced proofs of retrievability," in *Proc. of CCS*, pp. 831–843, 2014.

[4] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Proc. of ASIACRYPT*, pp. 90–107, 2008.

[5] Y. Dodis, S. Vadhan, and D. Wichs, "Proofs of retrievability via hardness amplification," in *Proc. of TCC*, pp. 109–127, 2009.

[6] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," in *Proc. of CCS*, pp. 187–198, 2009.

[7] C.Wang, Q.Wang, K. Ren, andW. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *Proc. of INFOCOM*, pp. 1–9, 2010.

[8] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," *ACM Transactions on Information System Security*, vol. 14, no. 1, pp. 1–34, 2011.

[9] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multi cloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2231–2244, 2012.

[10] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," in *Proc. of ASIACCS*, pp. 79–80, 2012.

[11] J. Chen, L. Zhang, K. He, R. Du, and L. Wang, "Message-locked proof of ownership and retrievability with remote reparing in cloud," *Security and Communication Networks*, 2016.

[12] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in *Proc. of CCS*, pp. 491– 500, 2011.

[13] R. Di Pietro and A. Sorniotti, "Boosting Efficiency and Security in Proof of Ownership for Deduplication," in *Proc. of ASIACCS*, pp. 81–90, 2012.

[14] J. Xu, E.-C. Chang, and J. Zhou, "Weak leakage-resilient clientside deduplication of encrypted data in cloud storage," in *Proc. of ASIACCS*, pp. 195–206, 2013.

[15] S. Keelveedhi, M. Bellare, and T. Ristenpart, "DupLESS: Serveraided encryption for deduplicated storage," in *Proc. of USENIX Security*, pp. 179–194, 2013.

[16] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1615–1625, 2014.

[17] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1206–1216, 2015.

[18] B. Wang, B. Li, and H. Li, "Public auditing for shared data with efficient user revocation in the cloud," in *Proc. of INFOCOM*, pp. 2904–2912, 2013.

[19] B. Wang, B. Li, and H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud," *IEEE Transactions on Cloud Computing*, vol. 2, no. 1, pp. 43–56, 2014.

[20] J. Yuan and S. Yu, "Efficient public integrity checking for cloud data sharing with multi-user modification," in *Proc. of INFOCOM*, pp. 2121–2129, 2014.