International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

Efficient Cross user Deduplication In cloud Storage

Ananda J[1], KumaraSwamy S[2], Dr. Kavitha K S[3], Dr. Kavitha C[4]

[1]*PG Student, Department of CSE, Global Academy of Technology, Bengaluru, Karnataka, India*
[2]*Associate Professor, Department of CSE, Global Academy of Technology, Bengaluru, Karnataka, India*
[3]*Professor, Department of CSE, Global Academy of Technology, Bengaluru, Karnataka, India*
[4]*Professor & HOD, Department of CSE, Global Academy of Technology, Bengaluru, Karnataka, India*

Abstract – *Deduplication on record is a valuable primitive that gives a client to check the veracity of transferred documents and to effectively refresh the records in a cloud server. In spite of the fact that specialists have proposed numerous dynamic deduplication plots in single user situations, the issue in multi-client conditions has not been ascertained adequately. A handy multi-client distributed storage framework needs the protected customer side cross-client deduplication strategy, which enables a client to get a responsibility for record immediately, if the comparative document is now exist in the cloud or whatever other client is transferred document before. To the best of our insight, none of the current dynamic deduplication method can bolster this strategy. In this venture, we present the idea of productive deduplicatable dynamic evidence of capacity and propose an effective development to accomplish secure cross-client deduplication, all the while.*

***Key Words*: Deduplication, Cross User, Veracity, Cloud storage, Multi User.**

1.INTRODUCTION

Data integrity is one of the most important properties when a user outsources its files to cloud storage. Many companies, such as Amazon, Google, and Microsoft, provide their own cloud storage services, where users can upload their files to the servers, access them from various devices, and share them with the others. Deduplication remains to be improved in a multi-user environment, due to the requirement of cross-user deduplication on the client-side. This indicates that users can skip the uploading process and obtain the ownership of files immediately, as long as the uploaded files already existing the cloud server. when a file is updated by a user, the cloud server has to update the corresponding authenticated structure in dynamic PoS, and construct a new authenticated structure for deduplication.

1.1 System model

Our framework display considers two sorts of substances: the cloud server and clients . For each document, unique client is the client who transferred the record to the cloud server, while resulting client is the client who demonstrated the responsibility for document yet did not really transfer the document to the cloud server. There are five stages dedeplicatable dynamic preprosess, transfer, dedplication, refresh and confirmation of capacity. Three stages (pre-prepare, transfer and deduplication) are executed just once in the lifecycle of a record from the point of view of clients. That is, these three stages seem just when clients expect to transfer records. The refresh stage and the confirmation of capacity stage can be executed different circumstances in the life cycle of a file. Once the possession is confirmed, the clients can subjectively enter the refresh stage and the verification of capacity stage without keeping the first records locally.

1.2 Threat model

We present the threat model as follows. The cloud server and users do not fully trust each other. A malicious user may cheat the cloud server by claiming that it has a certain file, but it actually does not have it or only possesses parts of the file. A malicious cloud server may try to convince users that it faithfully stores files and updates them, whereas the files are damaged or not up-to-date. The goal of deduplicatable is to detect these misbehaviors with overwhelming probability.

2. LITERATURE SURVEY

Deduplication is best when various clients outsource similar information to the distributed storage, however it raises issues identifying with security and possession. Confirmation of-possession plans permit any proprietor of similar information to demonstrate to the distributed storage server that he claims the information vigorously. Information deduplication includes finding and expelling duplication inside information without bargaining its devotion or respectability. The objective is to store more information in less space by dividing documents into little factor measured pieces (32–128 KB), recognizing copy lumps, distributed computing had turned into a profoundly requested administration or utility because of the upsides of high processing power, shoddy cost of administrations, superior, versatility, openness and in addition accessibility. also, keeping up a solitary duplicate of each piece. Excess duplicates of the piece are supplanted by a reference to the single duplicate.

3. PROPOSED SYSTEM

Our structure considers two sorts of components: the cloud server and customers, as in system building. For each archive, special customer is the customer who exchanged the record to the cloud server, while resulting customer is the customer who showed the duty regarding record however did not so much exchange the report to the cloud server. There are five phases in a deduplicatable dynamic PoS structure: pre-prepare, deduplication, transfer ,revive, and confirmation of limit and respectability check.

In the pre-handle stage, clients mean to transfer their nearby documents. Records will be separated into the squares for every last piece hashcode will be produced and squares will be sent to the deduplication stage. In the deduplication stage, it will checks every one of the squares of the document whether the piece is now transferred to the cloud or not in view of the hashcode created to each of the piece. On the off chance that as of now the piece is transferred then the client will get the responsibility for square, new pieces will be sent to the transfer phase. In the transfer stage, squares will be transferred to the distributed storage with the assistance of File Transfer Protocol. In the refresh stage, clients may adjust, embed, or erase a few squares of the documents. On the off chance that the new piece is embedded or refreshed just that square will be transferred to the distributed storage and record will get refreshed with the recently embedded or refreshed piece , without influencing to the deduplicated document.

In the proof of storage and integrity check phase, while downloading user will get the recent updated file content.

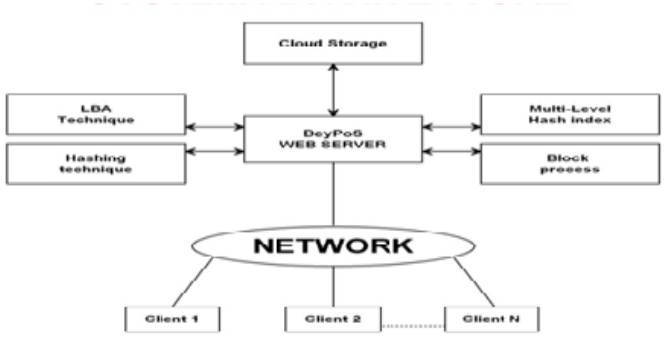

Fig -1: Proposed System Architecture

The architecture is split into following modules as shown in Fig-1:

3.1 File Uploading Process:

In the file uploading process module, user will select the file to be uploaded from the client system and write the file in the application folder ,then file will be send to the chunks formation module after that process finally file will be uploaded to the cloud storage by File Transfer Protocol and uploaded details will be stored in the database and Logical Block Addressing will be maintained in the database.

3.2 File Chunks Formation Process :

In the chunk formation process file will be divided in to the blocks based on the packet size, and generate the hash code for each blocks and maintain those details in the database.

3.3 Multilevel Indexing Process:

In the Multi Level Indexing process will be takes place for the deduplication check process, hash code will be divided in to three parts ,firstly compare the first part of hash code with the first level indexing , if it is matched then comparison will be takes place with second level of indexing ,else it will consider the block as the new block , if second level of indexing matched then the comparison will be takes place with the third level of indexing, else it will consider the block as the new block , if third level of indexing matched then it will consider the block is already exists and Based on the Map Reducing Technique it will increase the instance of the block in the database.

3.4 File Downloading Process

In the file Downloading process module, user will select the file to be downloaded, get the Logical Block Addressing from the database, based on the Logical Block Addressing; it will download the respective blocks from the cloud to the application.

3.5 Chunks Merging Process

In the Block merging process downloaded chunks of file will be merged and it will write the contents of the file it in the single file and finally file will be downloaded to the client system.

4 CONCLUSIONS

The proposed scheme is the first to address multiple copies of dynamic data. The interaction between the authorized users and the CSP is considered in our scheme, where the authorized users can seamlessly access a data copy received from the CSP using a single secret key shared with the data owner.

REFERENCES

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

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

[3] Z. Xiao and Y. Xiao, "Security and privacy in cloud computing,"IEEE Communications Surveys Tutorials, vol. 15, no. 2, pp. 843–859, 2013.

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