

Optimized Pricing Scheme And Avoiding Duplication Of Data In Cloud

S.Agalya¹,A.Aruna²,M.Selli³,R.Srishila⁴

Department of Information Technology,
A.V.C college of Engineering,Mannampandal,Mayiladuthurai, Tamilnadu,India.

Abstract- IAAS environment is referred as resources with VM instances. Customers can't utilize all resource, but provide full charge for allocated storage and in server side, storage are not utilized, so scalability become degraded. Implement best billing cycle for access and utilize the resources. Data Deduplication is becoming increasingly popular in storage systems as a space-efficient approach to data backup. Present SiLo, a near-exact deduplication system. That effectively and complementarily exploits similarity and locality to achieve high duplicate elimination. The file analyzed the check the auditor. The data secure storing and sharing of the files.

Keywords: Secretkey, deduplication, cloud storage.

I. INTRODUCTION

The cloud computing can be set of hardware, networks, storage, services, and interfaces. That merge to deliver aspects of computing as a service. Traditionally, access controls to data operate on the assumption that data servers can be trusted to keep data confidential and enforce access control policies correctly. cloud data storage where cloud service providers are not in the same trusted domains as end users, and hardware platforms aren't under the direct control of data owners. Cloud Storage provides customers with benefits, ranging from cost saving and simplified convenience, to mobility opportunities and scalable service. These great features attract more and more customers to utilize and store their personal data to the cloud storage. Deduplication file not checking for audio and video, but in this system we improve deduplication. That is effectively and complementarily exploits similarity and locality to achieve high duplicate elimination. In this system, Authorizes person only access the data.

II. RELATED WORK

We mainly focus on integrity and duplication so our work is also revolves around these two sections.

A. Secure Deduplication

Deduplication is a technique, in which user cannot upload the duplicate files. And also they could not make any changes in files. Our approach provides data confidentiality without impacting deduplication effectiveness. Data deduplication is the process by which a storage provider only stores a single copy of a file that is owned by several of its users. We consider client-side deduplication, i.e., the form of deduplication that happens at the client happens at the client side, thus avoiding the need to upload the file and saving network bandwidth.

B. Integrity Auditing

Using SiLo algorithm, we have to check the data as well as name of the file. It includes audio, video, text and images. If an user uploads the file then it will be stored in cloud successfully. Hereafter any other user upload the same file, then this uploading operation is neglected. This is achieved using SILO algorithm. This SILO algorithm compares the file uploaded with the existing file. If it already exist then it neglects and displays the message "content already exist" otherwise it will added into the cloud.

C. Cost Estimation

In the existing systems we have to pay the amount initially then only you are allowed to access the resources. but in our system user can access and use the resource without paying the amount initially. after usage they can pay the amount. this is similar to pay after use model. this in turn reduces the cost lend by customer. because they can pay only for what they use?

III. ALGORITHM

The similarity-based approaches minimize the RAM usage at the cost of potentially missing

large amounts of redundant data which is dependent on the similarity degree of the backup stream. We have examined the similarity degree and the duplicate-elimination measure using this SiLo algorithm. SiLo exploits similarity and locality jointly. It exploits similarity by grouping strongly correlated small files and segmenting large files, while locality is exploited by grouping contiguous segments in a backup stream to preserve the locality layout of these segments.

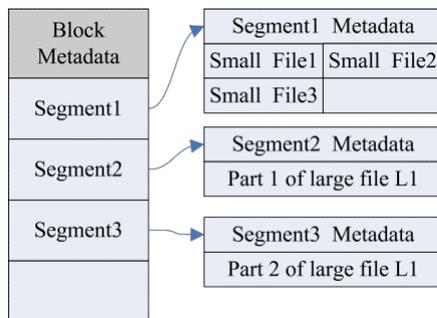


Fig 2: Data structure of the SiLo similarity algorithm

As another salient feature of SiLo, the SiLo locality algorithm groups several contiguous segments in a backup stream into a block and preserves their locality- layout on the disk. It serves to maximize the RAM utilization and reduce frequent accesses to on-disk index by retaining access locality in the backup stream. When SiLo reads the blocks from disk by the similarity detection, it puts the recently accessed block into the read cache. By preserving the backup-stream locality in the read cache, the accesses to on-disk index due to similarity detection can be significantly reduced, which all eviates the disk bottleneck and increases the deduplication throughput. Since it is at the block level where locality is preserved and exploited, the block size is an important system parameter that affects the system performance such as duplicate elimination and throughput. Each block in SiLo has its own Locality Hash Table for chunk filtering. Since a block contains several segments.

IV. SYSTEM MODEL

The data owner has a number of data files and stores them on the cloud server together with the authentication tags. Each owner in our design will also generate its own secret keys and public keys for authentication tag generation and data integrity verification. A user to whom the owner shares the data files can access and check the integrity of data

files using the public key. During the integrity auditing and deduplication processes, the user and the cloud server only use the public key and do not need any help from the data owner. While cloud servers are always equipped with abundant computing resources, data owners and users may have constrained computational power or bandwidth.

Data owner: The data owner is an entity that outsources data storage to the storage server and access the data later. In a client-side data deduplication system, only the first data owner of a file needs to upload while the following data owners of the same file does not require to upload the duplicate data any more.

Users: The entity of users in the deduplication systems makes registration at cloud storage server and has privileges to access some data files shared by some data owners.

Cloud storage server: The cloud storage server is an entity that provides the data storage service for the users. Furthermore, the cloud storage server will also perform duplicate check before upload the file again, which can reduce the storage cost at the server side and save the upload bandwidth at user side.

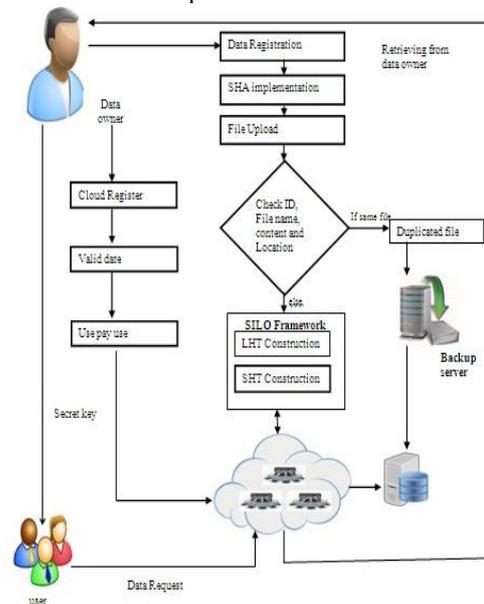


Fig 2: Architecture Diagram

File Upload: To upload a file, the user interacts the cloud storage server and tests if there is any duplicated copy stored in the cloud storage server.

Security Model: we consider the following factors that may impact integrity of data stored

on cloud servers: 1) attackers corrupting data stored on cloud servers; 2) attackers claiming the ownership of file stored on the cloud even if they do not possess the whole file; 3) hardware/software failures of cloud servers and operational errors of system administrator.

File Retrieving: Suppose a user wants to download a file. It first sends a request and the file names to the cloud storage server. Upon receiving the request and file names, the cloud storage server will check whether the user is eligible to download the files. If failed, the cloud storage server sends back an abort signal to the user to indicate the download failure.

Design Goals

To securely and efficiently verify integrity of the shared data on cloud with deduplication,

- **Efficiency:** The communication cost and computational cost for users to verify the integrity of data stored on cloud should be constant.
- **Functionality:** Public data integrity verification and deduplication should be supported at the same time without introducing functionally duplicated authentication tags.
- **Correctness:** The proposed scheme should accept all valid secret keys and public keys, all valid authentication tags, all valid proof information generated based on valid public keys and all valid data blocks.

Advantages of proposed system

In proposed reduce data usage waste.

- Short-job users are satisfied in payment cost.
- Optimal price point can satisfy both users and providers with maximized total utility.
- SiLo is able to remove large amounts of redundant data, dramatically reduce the numbers of accesses to on-disk index.
- Maintain a very high deduplication throughput.

V. Conclusion

In cloud many data are stored in cloud computing once more and again by user. So the user needs more spaces resource one more data. That will remove the memory space of the cloud for the users. To defeat this problem uses the deduplication concept. Data deduplication is a technique for sinking the amount of storage space an association wants to save its data. In

many associations, the resources systems surround duplicate copies of many sections of data. For instance, the similar file might be keep in several divergent places by dissimilar users, two or extra files that aren't the same may still contain much of the similar data. In future can extend our work to handle copies by saving just one copy of the data and return the other copies with pointer that lead reveal to the unique copy. So proposed Block-level deduplication frees up more spaces and demanding category documented as variable block or variable length deduplication has become very accepted. In cloud using the SHT and LHT tables the user easily searches the data and retrieves the searched data from the cloud. And also implemented heart beat protocol to recover the data from corrupted cloud server. New metrics are proved that our proposed draw near provide improved results in deduplication process, multimedia data for deduplication storage. The multimedia data includes audio, image and videos. And also implement heart beat protocol recover each data server and increase scalability process of system.

References

- [1] Y. Kouki and T. Ledoux, "Rightcapacity: Sla-driven cross layer cloud elasticity management," *IJNGC*, vol. 4, no. 3, 2013.
- [2] R. Stefan, K. Holger, M. Pascal, B. Andrew, and L. Mirosław, "Sizing the cloud," Forrester Research, 2011.
- [3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided encryption for deduplicated storage," in *USENIX Security Symposium*, 2013.
- [4] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in *Technical Report*, 2013.