A Hybrid Cloud Approach for Secure Authorized Deduplication

Sunita S. Velapure[1], S. S. Barde[2]

Department of Computer Engineering, SKNCOE, Pune, India

[2]Professor, Department of Computer Engineering, SKNCOE, Pune, India

Abstract: ***Data deduplication is one among vital knowledge compression techniques for eliminating duplicate copies of repetition knowledge, and has been wide employed in cloud storage to cut back the amount of storage space and save bandwidth. The main advantage of using cloud storage from the customers' expectation view is that customers will scale back their expenditure in buying and maintaining storage infrastructure whereas solely paying for the quantity of storage requested, which may be scaled-up and down upon demand. To protect the confidentiality of sensitive knowledge whereas supporting deduplication, the focused secret writing technique has been planned to encrypt the information before outsourcing. For raised shield knowledge security, this paper makes the primary plan to formally address the matter of licensed knowledge deduplication. Completely different from ancient deduplication systems, the differential privileges of user's area unit more thought-about in duplicate check besides the info itself. Addition to this we present many new deduplication constructions supporting licensed duplicate check in a hybrid cloud design. Security analysis demonstrates that our theme is secure in terms of the definitions as per the planned security model. As a proof of construct, we have a goal to implement a paradigm of our planned licensed duplicate check theme and conduct test bed experiments using our paradigm. We have a goal to show that our planned licensed duplicate check theme incurs comparatively less overhead compared to traditional operations.***

Keywords: DeDuplication, Fix size Chunks, deduplication Algorithm, Cloud Storage

1. Introduction

Cloud computing has recently emerged as a preferred business model for utility system. The conception of cloud is to supply computing resources as a utility or a service on demand to customers over the web. The concept of cloud computing is kind of the same as grid computing, which aims to achieve resource virtualization. In grid computing, the organizations sharing their computing resources, such as processors, so as to realize the utmost computing capacity, whereas cloud computing aims to supply computing resources as a utility on demand, which may proportion or down at any time, to multiple customers. This makes cloud computing play a serious role within the business domain, whereas grid is popular in tutorial, scientific and engineering analysis. Many definitions of cloud computing are outlined, depended on the individual purpose of read or technology used for system development. In general, we will outline cloud computing as a business model that offer computing resources as a service on demand to customers over the web. Cloud suppliers pool computing resources together to serve customers via a multi-tenant model. Computing resources area unit delivered over the web wherever customers will access them through varied consumer platforms. Customers will access the resources on- demand at any time without human interaction with the cloud supplier. From a customers' expectation view, computing resources area unit infinite, and customer demands will quickly amendment to fulfill business objectives. This is often expedited by the power for cloud services to scale resources up and down on demand leverage the facility of virtualization. Moreover, cloud supplier's area unit able to monitor and management the usage of resources for every client for charge purposes, improvement resources, capability coming up with and different tasks. Cloud storage is one amongst the services in cloud computing which provides virtualized storage on demand to customers. Cloud storage will be utilized in many various ways in which [4]. For example, customers will use cloud storage as a backup service, as opposition maintaining their own storage disks. Organizations will move their depository storage to the cloud which they will reach a lot of capability at the inexpensive, rather than shopping for extra physical storage. Applications running in the cloud additionally need temporary or permanent information storage in order to support the applications. As the quantity of knowledge within the cloud is quickly increasing, customers expect to achieve the on-demand cloud services at any time, whereas suppliers area unit needed to take care of system availability and method an oversized quantity of knowledge. Suppliers would like way to dramatically cut back information volumes, so that they will cut back costs whereas saving energy consumption for running giant storage systems. The same as different storages, storage in cloud environments can even use information deduplication technique.

2. Current Work

Data deduplication could be a technique to cut back cupboard space. By distinguishing redundant knowledge victimization hash values to match data chunks, storing just one copy, and making logical pointers to alternative copies rather than storing alternative actual copies of the redundant knowledge. Deduplication reduces knowledge volume so disc space and network information measure is reduced that reduce prices and energy consumption for running storage systems.

Data deduplication is applied at nearly each purpose which knowledge is hold on or transmitted in cloud storage. Many cloud suppliers provide disaster recovery and deduplication can be wont to create disaster recovery simpler by replicating knowledge when deduplication for dashing up replication time and information measure value savings. Backup and deposit storage in clouds may also apply knowledge deduplication so as to cut back physical capability and network traffic. Moreover in live migration method, we want to transfer an outsized volume of duplicated image knowledge [11]. There are a unit 3 major performance metrics of migration to consider: total knowledge transferred, total migration time and repair period of time. Longer migration time and period of time would be cause service failure. Thus, deduplication will assist in migration [12]. Deduplication can be wont to cut back storage of active knowledge like virtual machine pictures. Factors to think about once victimization deduplication in primary storage is the way to balance the trade-offs between storage space saving and performance impact.

3. Current Work Issues

When activity deduplication, a little of knowledge chunks are rather more vital than others (For example, data chunks that are documented by several files). Ancient deduplication approaches don't implement redundancy of knowledge chunks. Thus, deduplication might scale back the responsible of the storage system because of the loss of many vital chunks that can result in the loss of the many files. As a result, the essential chunks ought to be replicated quite the diminished knowledge chunks so as to boost responsible of the system. The authors in [14], take into account the consequences of deduplication on the reliability of the deposit system. They projected Associate in nursing approach to improve responsible by developing a way to weigh and measure the importance of every chunk by examining the number of knowledge files that share the chunk, and use this weight to identify the amount of redundancy needed for the chunk to guarantee QoS.

Table 1: Survey Table

Sr No	*Paper*	*Year*	*Advantage*	*Disadvantage*
1	A Hybrid Cloud Approach for Secure Authorized De-duplication	2014	1) Works with hybrid cloud 2) Server side compression using de-duplication technique 3) Secure authorization	1) Whole file based matching 2) Few changes in file considered as different file, reduces file compression
2	DupLESS: Server-Aided Encryption for Deduplicated Storage	2013	1) Secure outsourced storage that both supports de- duplication and resists brute-force attacks. 2) DupLESS provides strong security against external attacks which compromise the SS and communication channels.	1) Low performance 2) Increased storage requirement
3	Dynamic Data Deduplication in Cloud Storage	2014	1) A dynamic data de-duplication 2) Scheme for cloud storage, in order to fulfill a balance between changing storage efficiency and fault tolerance requirements 3) Improve performance in cloud storage systems	1) Single metadata server fails then all system fails 2) Less consistency
4	Improving Accessing Efficiency of Cloud Storage Using De-Duplication and Feedback Schemes	2014	1) By compressing and partitioning the files according to the chunk size of the cloud file system, 2) We can reduce the data duplication rate. 3) The backup efficiency can be improved and the load balancing among the nodes is considered.	1) Chunks size influence the system 2) Indexing matter for server performance

4. System Architecture

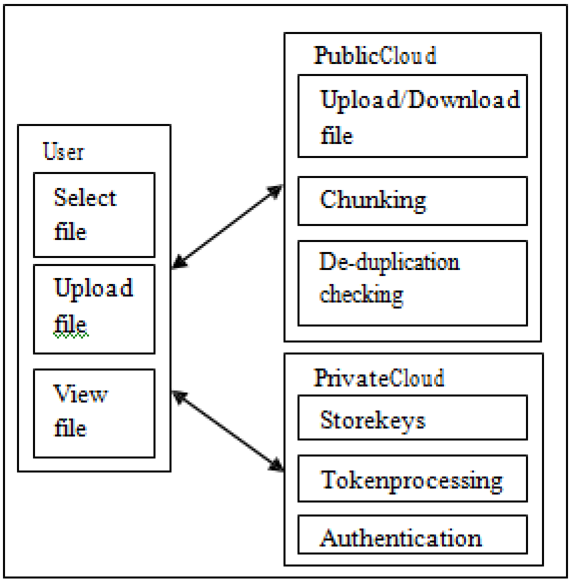

Figure: System Architecture

5. Main Modules

1. User Module

In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

2. Secure DeDuplication System

To support authorized deduplication, the tag of a file *F* will be determined by the file *F* and the privilege. To show the difference with traditional notation of tag, we call it file token instead. To support authorized access, a secret key *kp* will be bounded with a privilege *p* to generate a file token. Let $\phi' F;p$ = TagGen(F, kp) denote the token of *F* that is only allowed to access by user with privilege *p*. In another word, the token $\phi' F;p$ could only be computed by the users with privilege *p*. As a result, if a file has been uploaded by a user with a duplicate token ϕ' $F;p$, then a duplicate check sent from another user

3. Security of Duplicate Check Token

We consider several types of privacy we need protect, that is, i) unforgeability of duplicate-check token: There are two types of adversaries, that is, external adversary and internal adversary. As shown below, the external adversary can be viewed as an internal adversary without any privilege. If a user has privilege *p*, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege *p′* on any file *F*, where *p* does not match *p′*. Furthermore, it also requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forge and output a valid duplicate token with *p* on any *F* that has been queried.

4. Send Key

Once the key request was received, the sender can send the key or he can decline it. With this key and request id which was generated at the time of sending key request the receiver can decrypt the message.

6. Advantages

- It makes overhead to minimal compared to the normal convergent encryption and file upload operations.
- Data confidentiality is maintained.
- One critical challenge of cloud storage services is the management of the ever- increasing volume of data.
- Secure compared to existing techniques.

7. Implementation

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it’s constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Will be successful if and only if he also has the file *F* and privilege *p*. Such a token generation function could be easily implemented as *H*(*F, kp*), where *H*(_) denotes a cryptographic hash function.

8. Conclusion

The well-known data deduplication algorithms are divided into fixed-length chunking and variable- length chunking. The fixed-length chunking is very fast for processing data deduplication but degrades the deduplication performance. However, the variable length chunking can achieve significant data deduplication performance with high computation overhead and longer processing time. In this paper, we suggest a dynamic chunking approach that overcomes the inherent problem of fixed-length chunking by adapting file similarity technique. The key idea of this work is to find several duplicated point by comparing hash key value and file offset within file similarity information. Several issues remain open. First, our work has limitations on supporting simple data file which has redundant data blocks with spatial locality; therefore, if the file has several modifications then overall performance will be degrade. For future work, we plan to build a massive deduplication system with huge number of files. In this case, handling file similarity information needs more elaborated scheme.

References

[1] IEEE Transactions on Parallel and Distributed Systems“A Hybrid Cloud Approach for Secure Authorized Deduplication”Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou DOI10.1109/TPDS.2014.2318320,

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296–312, 2013.

[4] Mokadem, R., Hameurlain, A.: An efficient resource discovery while minimizing maintenance overhead in sdds based hierarchical dht systems. International Journal of Grid and Distributed Computing 4(3), 1–23 (2011)

[5] Mokadem, R., Hameurlain, A.: An efficient resource discovery while minimizing maintenance overhead in sdds based hierarchical dht systems. International Journal of Grid and Distributed Computing 4(3), 1–23 (2011)

[6] http://docplayer.net/5383340-Byte-index-hunking-algorithm-for-data-deduplication- system.html

[7] http://www.hindawi.com/journals/ijdsn/2014/6303