Survey Based on Differential Authorized Deduplication Using Hybrid Cloud Approach

Madhu Ramteke
CSE, CSIT,Durg, India.

K.L. Sinha
CSE, CSIT,Durg, India.

Abstract – Now a day's volume of data is increasing day by day to reduce the amount of storage space, data deduplication is one of the techniques which compress the data by eliminating redundant copies of the data. In previous system, Symmetric encryption technique uses a common secret key to encrypt and decrypt data. In the existing deduplication system, each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for a file, the user needs to take the file and his own privileges as inputs. The user is going to find a duplicate for the file if and only if there is a copy of this file and a matched privilege stored in cloud. Symmetric encryption technique cannot be used with authorized deduplication check scheme at same time which is more important in many applications. To enhance the data security, convergent key encryption technique which uses common secret key to encrypt and decrypt data has been proposed to solve the issue of authorized data deduplication.

Index Terms – Deduplication, hybrid Cloud, Convergent key, and Encryption.

1. INTRODUCTION

Cloud computing is the new rising trends in the new generation technology. Each client has tremendous amount of data to share to store in a rapidly accessible secured place. The idea of deduplication is arrived here to effectively use the bandwidth and disk utilization on cloud computing. To maintain a strategic distance from the duplication duplicates of similar data on cloud may cause lose of time, bandwidth usage and space . Cloud computing is web based, a network of remote servers associated over the Internet to store, share control, recover and processing of data, rather than a local server or PC. The advantage of cloud computing are tremendous. It enables us to work from anyplace.

The most vital thing is that client doesn't have to purchase the asset for data storage. With regards to Security, there is a possibility where a pernicious client can infiltrate the cloud by imitating a legitimize client, there by influencing the whole cloud in this manner tainting many clients who are sharing the contaminated cloud.

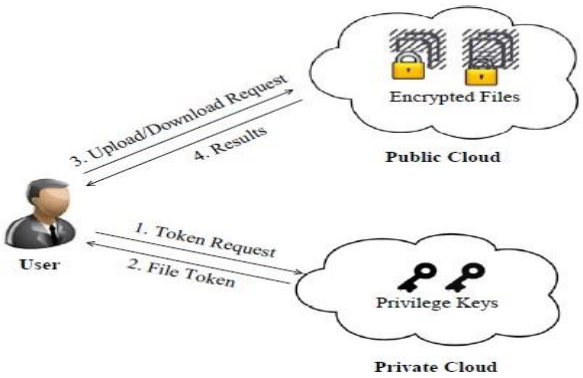

Fig. 1. Architecture for Authorized Deduplication

There is additionally huge issue, where the duplicate copies may transfer to the cloud, which will incite to misuse of band width and disk use. To deal with this issue there should be an average level of encryption gave, that restrictive the customer should have the ability to get the data and not the legitimate User.

1.1 Deduplication

Data deduplication (also called "intelligent compression" or "single-instance storage") is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data.

Deduplication takes out these extra copies by saving just a single copy of the data and supplanting interchange copies with pointers that lead back to the main copy. Associations frequently use deduplication in backup and calamity recovery applications, be that as it may it can be used to free up space in primary storage as well.

Data deduplication is a technique for decreasing storage needs by taking out excess data. Just a single special occurrence of

the data is really held on capacity media, for example, disk or tape. Excess data is supplanted with a pointer to the unique data duplicate. For instance, a commonplace email system may contain 100 occasions of a similar one megabyte (MB) file attachment. On the off chance that the email stage is backed up or archived, every one of the 100 occurrence is spared, requiring 100 MB storage space. With data deduplication, just a single instance of the attachment is certainly stored; each consequent occurrence is simply referenced back to the one spared duplicate. In this example, a 100 MB storage demand could be reduced to only one MB.

Data deduplication offers different advantages. Lesser storage space requirements will spare money on disk consumptions. The more productive utilization of disk space likewise takes into account longer disk maintenance periods, which gives better recuperation time objectives (RTO) for a more drawn out time and decreases the requirement for tape backups. Data deduplication likewise decreases the data that must be sent over a WAN for remote backups, replication, and disaster recovery. In genuine practice, data deduplication is regularly utilized as a part of conjunction with different types of data reduction, for example, traditional compression and delta differencing. Taken together, these three strategies can be exceptionally successful at upgrading the utilization of storage space.

Data deduplication can for the most part work at the file or block level. File deduplication eliminates duplicate files; however this is not an extremely proficient method for deduplication. Block deduplication searches inside a record and spares unique iterations of every block. Every chunk of data is prepared utilizing a hash algorithm, for example, MD5 or SHA-1. This procedure produces a one of a kind number for every piece which is then put away in an index. In the event that a file is updated, just the changed data is spared. That is, if just a couple of bytes of a document or presentation are changed, just the changed blocks are spared; the changes don't constitute an altogether new document. This conduct makes block deduplication significantly more proficient. Be that as it may, Block deduplication takes additionally handling power and uses a much bigger index to track the individual pieces.

In its simplest form, deduplication happens on the file level; that is, it eliminates duplicate copies of a similar file. This sort of deduplication is sometimes called file level deduplication or single instance storage (SIS). Deduplication can also happen on the block level, disposing of copied blocks of data that happen in non-indistinguishable files. Block level deduplication liberates more space than SIS, and a specific sort known as variable block or variable length deduplication has turned out to be extremely popular. Frequently the expression data deduplication is utilized as an equivalent word for block level or variable length deduplication.

All in all, deduplication technology can be conveyed in one of two basic ways:

i. at the source or

ii. at the target.

In source deduplication, data copies are wiped out in essential storage before the data is sent to the backup system. The advantage of source deduplication is that is diminishes the bandwidth requirements and time necessary for backing up data.

On the drawback, source deduplication devours more processor assets, and it can be hard to coordinate with existing systems and applications.

By differentiation, target deduplication happens inside the backup system and is frequently much less demanding to send. Target deduplication comes in two sorts:

a) in-line or

b) post-process.

In-line deduplication happens before the backup copy is composed to disk or tape. The advantage of in-line deduplication is that it requires less storage space than post-process deduplication, yet it can back off the backup process.

Post-process deduplication happens after the backup has been composed, so it requires that associations have a lot of storage space accessible for the original backup. Be that as it may, post-process deduplication is normally faster than in-line deduplication.

The essential advantage of data deduplication is that it decreases the amount of disk or tape that associations need to purchase, which thus diminishes costs. NetApp reports that now and again, deduplication can lessen capacity necessities up to 95 percent, yet the kind of data you're attempting to deduplicate and the measure of record sharing your association wills impact your own deduplication proportion. While deduplication can be connected to data put away on tape, the generally high expenses of disk storage make deduplication an extremely famous alternative for disk based systems. Wiping out additional duplicates of data spares money on direct disk equipment costs, as well as on related costs, like electricity, cooling, maintenance, floor space, and so forth.

Deduplication can likewise reduce the amount of network bandwidth required for backup process, and at times, it can accelerate the backup and recovery process.

2. RELATED WORK

According to Li Jin et al., 2015 [1] to protect the confidentiality of available data while supporting deduplication, proposed the convergent encryption technique

to encrypt the data before outsourcing. To protect data security, this paper does make the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in dup licate check besides the data itself. We also present different new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model.

An arising challenge is to perform secure deduplication in cloud storage. Agrawal N.O. et al., 2014 [12] introduced a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, a new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers. Security analysis demonstrates that Dekey is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication.

Distinguishing regular lumps of data both inside and putting away in just once, according to Storer M.W. et al., 2013 [11] de-duplication can yield cost funds by expanding the utility of a given measure of storage. De-duplication abuses indistinguishable substance; while encryption endeavors to make all substance seem irregular. A similar substance encoded with two diverse keys brings about altogether different figure content. Joining the space proficiency of de-duplication with the mystery parts of encryption is tricky. Built up an answer that gives both information security and space effectiveness in single-server stockpiling and appropriated stockpiling frameworks. United encryption to empower encryption while as yet permitting the de-duplication basic on lumps.

The user, playing the role of a prover, can then identity itself to a verifier in a protocol in which the verifier begins by knowing only the claimed identity of the prover and the master key of the authority. Bellare M. et al., 2011 [5] gives either security confirmations or attacks for countless based recognizable proof and mark plans characterized either unequivocally or certainly in existing. This casing works that one hand it helps explain how the plans are inferred and empowers the secluded security investigation. IBI is a power having an ace open key and an ace mystery key.

This power can furnish a client with a mystery key in light of its personality. The client, assuming the part of a prover, can then personality itself to a verifier in a convention in which the verifier starts by knowing just the asserted character of the prover and the ace key of the power. IBS plan is comparable aside from that the client signs messages, instead of distinguishing itself and check of a mark requires learning just of the character of the endorser and the ace open key.

According to Anderson Paul et al., 2010 [2] Encrypting data refutes the de duplication, two indistinguishable information pieces, encoded with various keys. The encryption key for the information square is gotten from the substance of the information utilizing a capacity is like the has work. Two indistinguishable information squares yield indistinguishable encoded pieces which can be copied in the ordinary way. Every piece has a different encryption key. To depict customary reinforcement calculation which takes favorable circumstances of the information which is normal between clients in increment the speed of reinforcements and diminishes the capacity prerequisite. This calculation bolsters customer end-client encryption which is vital for classified individual information. It additionally underpins a one of a kind component which permits prompt location of regular sub trees, maintaining a strategic distance from the need to inquiry the reinforcement framework for each life. To descries a model execution of this calculation for Apple OS X, and present an investigation of the potential adequacy, utilizing genuine information got from an arrangement of common clients.

3. CONCLUSION

In this paper, the idea of approved data deduplication was proposed to ensure the data security by including differential benefits of clients in the duplicate check. In this extend we perform a few new deduplication developments supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate check tokens of files are produced by the private cloud server with private keys. As a proof of idea in this project we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. From this project we demonstrate that our authorized duplicate check scheme brings about negligible overhead compared to convergent encryption and network transfer.

ACKNOWLEDGEMENT

With immense pleasure, we are publishing this paper as a part of the curriculum of M.Tech. Computer Science & Engineering. It gives us proud privilege to complete this paper

work under the valuable guidance of Principal for providing all facilities and help for smooth progress of paper work. We would also like to thank all the Staff Members of Computer Engineering Department, Management, friends and family members, Who have directly or indirectly guided and helped us for the preparation of this paper and gives us an unending support right from the stage the idea was conceived.

REFERENCES

[1] Li, Jin, et al. "A Hybrid Cloud Approach for Secure Authorized Deduplication.",IEEE 2015
[2] Anderson, Paul, and Le Zhang. "Fast and Secure Laptop Backups with Encrypted De-duplication." LISA. 2010.
[3] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "DupLESS: server-aided encryption for deduplicated storage." Proceedings of the 22nd USENIX conference on Security. USENIX Association, 2013.
[4] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "Message-locked encryption and secure deduplication." Advances in Cryptology–EUROCRYPT 2013. Springer Berlin Heidelberg, 2013. 296-312.
[5] Bellare, Mihir, Chanathip Namprempre, and Gregory Neven. "Security proofs for identity-based identification and signature schemes." Journal of Cryptology 22.1 (2009): 1-61.
[6] Bugiel, Sven, et al. "Twin clouds: Secure cloud computing with low latency." Communications and Multimedia Security. Springer Berlin Heidelberg, 2011.
[7] Bugiel, Sven, et al. "Twin clouds: An architecture for secure cloud computing." Proceedings of the Workshop on Cryptography and Security in Clouds Zurich. 2011.
[8] Halevi, Shai, et al. "Proofs of ownership in remote storage systems." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.
[9] Di Pietro, Roberto, and Alessandro Sorniotti. "Boosting efficiency and security in proof of ownership for deduplication." Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. ACM, 2012.
[10] Yuan, Jiawei, and Shucheng Yu. "Secure and constant cost public cloud storage auditing with deduplication." Communications and Network Security (CNS), 2013 IEEE Conference on. IEEE, 2013.
[11] Stanek, Jan, et al. A secure data deduplication scheme for cloud storage. Technical Report, 2013.
[12] Agrawal N.O, et al. "Secure deduplication with efficient and reliable convergent key management." 2014.
[13] Nithya, A., B. Ramakrishnan, and Resul Das. "A Novel Approach for Data Privacy Using Attribute Based Scheme Algorithm for Cloud Computing." International Journal of Computer Networks and Applications, 3(4), PP: 70 – 77, 2016, DOI: 10.22247/ijcna/2016/v3/i4/48567.

Authors

Ms. Madhu Ramteke, M.Tech Student, Dept of CSE, Chhatrapati Shivaji Institute of Technology, Durg. Received B.E in Computer Science and Engineering from Chhatrapati Shivaji Institute of Technology, Durg. Interesting Areas are Analysis and Design of Algorithms and Cloud Computing.

Mr. K.L. Sinha, Associate Professor, Dept of CSE, Chhatrapati Shivaji Institute of Technology, Durg. Received B.E in Information Technology from Chhatrapati Shivaji Institute of Technology, Durg. Received M. Tech degree in 2012 from Chhatrapati Shivaji Institute of Technology, Durg, His interests are Digital Image Processing, Operating Systems and Data Mining.