Indian Journal of Science and Technology, Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/107981, December 2016

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Hybrid Cloud Based Deduplication Scheme Based on User's Privileges

M. Sharmila*, P. Balakrishnan

School of Computing, SASTRA University, Thirumalaisamudram, Thanjavur - 613401, Tamilnadu, India;
sharmila30711@gmail.com, baskrish1977@gmail.com

Abstract

Background/Objectives: This research work reviews different techniques to protect the data and discussed about their merits and demerits which help to realize a secured cloud deduplication environment. **Methods:** Hybrid cloud based deduplication approach is proposedto avoid the data duplication and to ensure the confidentiality in the cloud. Further, the security investigations ensure that the proposed security framework enhances the data security with an insignificant additional overhead. Mention the data source accessed and keywords used; inclusion and exclusion criteria etc. **Findings: Provide your findings. Applications:** This approach makes us to understand the elements of good cloud deduplication framework which can be directly used by any cloud storage service provider to reduce their storage costs.

Keywords: Cloud Computing, Data Deduplication, Hybrid Cloud, Private Cloud, Public Cloud,

1. Introduction omit textbook information

2. Methodologies

Data deduplication is a technique for uprooting copy duplicates of information, and has been broadly connected in distributed storage to decrease storage space as well as bandwidth[1]. However, secure deduplication in distributed storage is a quite challenging scenario. Although convergent encryption has been broadly gained for secure deduplication, the real challenge lies in handling enormous number of joined keys. [2]Encryption has been presented to uphold information privacy while making deduplication feasible. It scrambles/decodes an information duplicate with a merged key. After the information encryption, clients hold the keys and send the cipher text to the cloud. [3]This research work demonstrated a safe focalized encryption technique. The proposed method highlights the encryption and record level deduplication. [4]The proposed model exhibits polynomial-based verification labels and homomorphism direct authenticators. It permits deduplication of both documents and their relating confirmation labels. Here, information trustworthiness inspection and capacity deduplication is accomplished at the same time. It also portrays a constant correspondence and computational expense on the client side. Subsequently, the proposed approach outflanks existing POR and PDP plans while giving the extra usefulness of deduplication. In[5] Sedic influences the extraordinary components of Map Reduce to consequently parcel a figuring work as indicated by the security levels of the information and organize the calculation over a hybrid cloud. In particular, modified Map Reduce methodology conveyed document framework to deliberately repeat information, moving disinfected information pieces to public cloud clients. This approach permits the clients to interface with our framework similarly they work with Map Reduce and specifically run their legacy code in the system. Legitimacy expands the security of the information on hybrid clouds. However, it doesn't address the approved deduplication issue over information out in the public cloud. [6]The proposed approach enables either security verifications or assaults for an extensive number of personality based distinguishing proof. [7]This work proposes a design that gives secure deduplicated stockpiling opposing savage power assaults, and acknowledge it in a

*Author for correspondence

framework called DupLESS. Here, the customers scramble the data using message-based keys got from a key-server. It empowers customers to store encoded information with a current administration, enabling them to perform deduplication for their benefit along with privacy. The primary advantage is to ensure the information classification by changing the anticipated message into eccentric message. However, the drawback is expected to rely on the outsider key server. [8]It gives definitions for both protection as well as label consistency. In light of this establishment, it makes both practical and hypothetical commitments. On the practical side, it gives ROM security investigations of a characteristic group of MLE plans that includes conveyed plans. However on the hypothetical side, it gives the standard model arrangements and associations with deterministic encryption for various classes of message sources. The primary advantage is the information classification in deduplication. But, is does not consider the issues of the key-administration. In[9] this work, the clients have the capability to expand the velocity of reinforcements thereby decreasing the storage capacity. This calculation strengthens the customer side per-client encryption which is important for secret individual information. It underpins a component which permits quick recognition of regular sub trees, evading the need to inquiry the reinforcement framework for each record. The primary benefit is the reduction in reinforcement times and capacity prerequisites. However, the downside is it does not permit client to make the copy check. In[10], an authentic server is proposed to manage the recorded storehouse that can be shared by various customer machines and applications. An incremental reinforcement application will most likely be unable to decide precisely which pieces have changed. However, these copy squares will be disposed of and contains the duplicate of the information will be held. Indeed, even copy information from various applications and machines can be wiped out, if the customers compose the information utilizing the same piece size and arrangement. The benefit is reduction in storage space. However, the important drawback is that there is a possibility for data leakage. In[11] a deduplication strategy for private information stockpiling is presented and formalized. Naturally, a private information deduplication convention permits a customer who holds that information has the capability to demonstrate to a server which holds a synopsis string of the information which reveals that he/she is the owner of that information without uncovering additional data to the server. This private information deduplication convention is provably secure accepting that the fundamental hash capacity is crash strong, the discrete logarithm is hard and the deletion coding calculation can eradication up to division of the bits in the vicinity of vindictive enemies in the vicinity of noxious foes.

3. System Model

Software design is the procedure of characterizing the engineering, segments, modules, and information for a framework to fulfill determined necessities. One could consider it to be the utilization of framework to accomplish deduplication[12]. The Hybrid cloud is shown in Figure 1 is a way to deal with the issue of deduplication with differential benefits in distributed computing. Consequently the symmetric encryption, proof of possession and mixture way to deal with accomplish a protected duplication check with various benefits. Here to stay away from the key sharing among the client we kept and mange the keys in the private cloud. The Novel encryption key used to maintain a strategic distance from the beast power assaults in broad daylight mists and give a safe duplication check. To productively taking care of the issue of deduplication with differential benefits in cloud computing. In this work is shown in Figure 2, demonstrating the hybrid cloud design

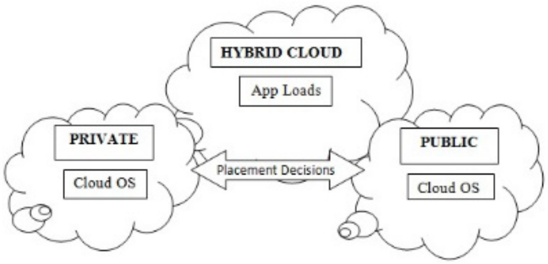

Figure 1. Deployment models.

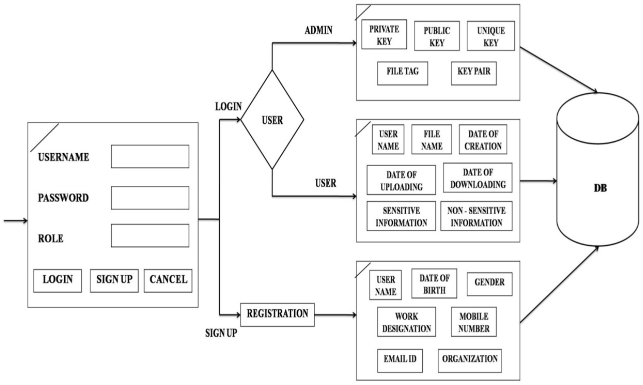

Figure 2. Proposed Architecture.

component. Utilizing this way to deal with perform the protected copy check. For the most part in existing framework like a Convergent encryption. It encodes/decodes a Data duplicate with a merged key, consequently having opportunity to create same focalized key for same indistinguishable duplicates. Essential objective of the cross breed cloud way to deal with taking care of the issue of deduplication with differential benefits in cloud computing. Consequently the Symmetric encryption, Proof of possession and cross breed way to deal with accomplishes a protected duplication check with various benefits. Here to keep away from the key sharing among the client we kept and manage the keys in the private cloud. For give a protected duplication check to the users.

3.1 User Registration/Authentication

The data owner gets signed up in the cloud by giving subtle elements. After signing up, the private cloud administrator assigns a login id and password to that user.

3.2 Key Pair Registration

In this module, the client send request to CSP (cloud service provider). The cloud administrator verify the client and then create the key pair for the client using novel encryption key and concurrent key technique. The combined key contains a client id, open key and private key. The private cloud that stores private key and keep up the benefits table for all the client tokens. Finally the clients get the id, benefits and open key from the CSP. The administrator produces the focalized key for duplication check in the cloud.

3.3 User File N token Development

To check the duplicates for a document, first create the novel encryption key and concurrent key used to delineate information. After that, the client sends a demand token to the private cloud, then the private clouds that check the client open key and id, access rights by the privileage table. If token match to privilege table then produce the document token that contain the general population key, id, record label, private key. In the record token private key was limited by the general population key, id and file tag.

3.4 Duplicate Check by POW

This module utilizes the document token to get to the information out from the public cloud. In public cloud, the document token was confirmed by the POW (proof of ownership). The POW that verifies the data owner by referring it from the private cloud. Then the record token send to the POW which has the general population key ,id, private key ,document tag to confirm the client in private cloud. The private cloud cross verifies and returns the time stamp and guide for the information in public cloud.

3.5 Convergent Key and Novel Key Encryption

Here, novel encryption key and joined key was scrambled to the clients in public cloud after POW confirmation. The scrambled merged key contains the private key, time stamp, and document tag. The POW checks that whether the record label is same for the officially existing document label. If so, it allows the user to access the current record area. On the other hand, if the record tag is new means then permit to include the new document in the cloud thereby the proposed approach accomplishes the distinctive approved copy check.

4. Experimental Results

System implementation is the stage in the project where the theoretical design is turned into a working system. The most critical stage is achieving a successful system and in giving confidence on the new system for the user that it will work efficiently and effectively. The average upload time for test sets of different file size and format (.txt, .jpeg, .pdf, .mpeg) are plotted in Figure 3. The average download time for test sets of different file size and format (.txt, .jpeg, .pdf, .mpeg) are plotted in Figure 4. The proposed approach develops a user friendly system, which has a menu-based graphical user interface for the end user. After coding and testing, the project is to be installed on the necessary system. Firstly, the executable file is to be created

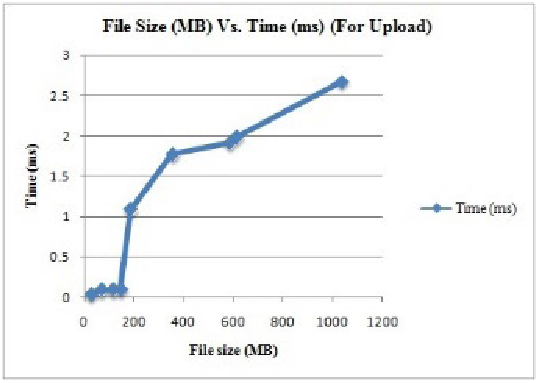

Figure 3. File Size vs. Time (For Upload).

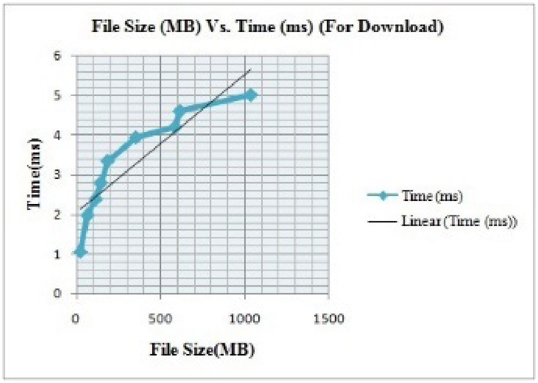

Figure 4. File Size vs. Time (For Download).

and loaded in the system. Secondly, the code is again tested in the installed system. In the existing system, traditional data deduplication is susceptible to both inside and outside attacks. Convergent encryption operation is deterministic in nature that has the identical copies of cipher text which leads to unauthorized access. Since the convergent cryptographic key produces same key for identical file content, it generates similar cipher text for the different cloud owners. The proposed approach uses a novel encryption key generation algorithm which produces different convergent keys. The set of privileges can be assigned by the data owners in two ways: Role-based privileges and Time-based privileges. This improves the scalability and high security to hybrid cloud which in turn reduces the data leakage and enhances the data deduplication to cipher text also.

5. Conclusion

In this research work, a secure deduplication framework was proposed to avoid the data duplication as well as to ascertain the confidentiality in the cloud environment. Besides, the proposed approach reduces the amount of storage space as well as cost of the cloud storage service providers. Additionally, the security investigation promises that the proposed approach improvises the data security with negligible additional overhead. In future, we are planning to develop a tenant based deduplication framework with which we can reap out more benefits in terms of storage space and cost for an organization on the whole.

6. References

1. Li J, Chen X, Li M, Li J, Lee P, Lou W. Secure deduplication with efficient and reliable convergent key management. In Proceeding IEEE Transactions on Parallel Distributed Systems. 2014 Jun; p. 1615–25.
2. Reclaiming space from duplicate files in a serverless distributed file system. Date Accessed: 2/07/2002: Available from: http://ieeexplore.ieee.org/document/1022312//
3. Xu J, Chang E C, Zhou J. Weak leakage-resilient client side deduplication of encrypted data in cloud storage, ASIA CCS '13 Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, USA. 2013; p. 195–206.
4. Secure and constant cost public cloud storage auditing with deduplication. Date Accessed: 14/10/2013: Available from: http://ieeexplore.ieee.org/document/6682702/.
5. Zhang K, Zhou X, Chen Y, Wang X, Ruan Y. Sedic: Privacy-aware data intensive computing on hybrid clouds, CCS '11 Proceedings of the 18th ACM conference on Computer and communications security, USA. 2011; p. 515–26.
6. Bellare M, Namprempre C, Neven G. Security proofs for identity-based identification and signature schemes. Journal of Cryptology. 2009 Jan; 22(1):1–61.
7. Bellare M, Keelveedhi S, Ristenpart T. Dupless: Serveraided encryption for deduplicated storage. Washington, DC: SEC'13 Proceedings of the 22nd USENIX conference on Security. 2013; p. 179–94.
8. Bellare M, Keelveedhi S, Ristenpart T. Springer Berlin Heidelberg: Message-locked encryption and secure deduplication. 2013 May; p. 296–312.
9. Anderson P, Zhang L. Fast and secure laptop backups with encrypted deduplication. USA: LISA'10 Proceedings of the 24th international conference on large installation system administration. 2010.
10. Quinlan S, Dorward S. Venti: A new approach to archival storage. USA: FAST '02 Proceedings of the Conference on File and Storage Technologies. 2002; p. 89–101.
11. Ng WK, Wen Y, Zhu H. Private data deduplication protocols in cloud storage. USA: SAC '12 Proceedings of the 27th Annual ACM Symposium on Applied Computing. 2012; p. 441–46.
12. Venish A, SivaSankar K. Framework of Data Deduplication: A Survey. Indian Journal of Science and Technology. 2015 Oct; 8(26):1–7.