International Journal of Science Engineering and Advance Technology

Secure Deduplication Technique For Privilege Data Access Using Hybrid Cloud

V.Priyanka#1, J.S.V.Gopala Krishna#2
#1Student of M.Tech (CSE) ,#2 Associate. Prof, Department of Computer Science and Engineering, Sir C R Reddy College of Engineering, Eluru, W.G.DIST.

Abstract:
Cloud is moderated technology in now a days but this type of technology is failure in many times. Cloud computing having secure and confidential problems because of data irretrievability. So we are deliberate on faithful information. In this time the technology is focus on duplicate copies. The major idea is we did not accept any duplicate copy like same name of file are same content of file. Then we are successfully gain needed data on cloud Data deduplication is a demanding technology to support dispose of redundant information as an option of enthralling records; it provisions simply distinct copy of file. Together with the whole associations reposting and numerous associations encase more bits of copy in sequence. Holder in point different clients stores corresponding documents in a few preferred spaces. Deduplication abolish the extra duplicates by trimming the duplicate data and restore alternate duplicates alongside pointers that flipside to the first duplicate. It represents the information pressure procedure to build the data transfer capacity proficiency. Data deduplication is tremendously utilizing as a part of distributed computing currently. Data deduplication protects the privacy of touchy data. It lives Up to expectations with merged encryption system to scramble the information before convey. Administrations frequently utilize Deduplication technique for reinforcement and disaster recuperation functions. Here We attempt to accept the deduplication technique, to assemble with simultaneous encryption to supervise the cost of security for sensitive information with half and half distributed computing.

Keywords : Deduplication, authorized duplicate check, confidentiality, hybrid cloud.

.I. Introduction:

Cloud computing technique which is most widely used todays.In that, computing is done over the large communication network like Internet.It is an important solution for business storage in low cost.Cloud computing provide vast storage in all sector like government, enterprise,also for storing our personal data on cloud. Without background implementation details,platform user can access and share different resources on cloud.The most important problem in cloud computing is that large amount of storage space and security issues.

One critical challenge of cloud storage to management of ever -increasing volume of data.To improve scalability,storage problem data deduplication is most important technique and has attracted more attension recently.It is an important technique for data compression ,It simply avoid the duplicate copies of data and store single copy of data.Data deduplication take place in either block level or file level.In file level approach duplicate files are eliminate,and in block level approach duplicate blocks of data that occur in non - identical files.Deduplication reduce the storage needs by upto 90 95% for backup application,68% in standard file system.Important issues in data deduplication that security and privacy to protect the data from insider or outsider attack.For data confidentiality,encryption is used by different user for encrypt there files or data ,using a secrete key user perform encryption and decryption operation.For uploading file to cloud user first generate convergent key,encryption of file then load file to the cloud.To prevent unauthorize access proof of ownership protocol is used to provide proof that the user indeed owns the same file when deduplication found.After the proof,server provide a pointer to subsequent user for accessing same file without needing to upload same file.

II. Related Work

Information deduplication innovation is technique for expanding the utilization of given information repositing. Deduplication distinguishes similarity among different records to free circle space. Point of interest in deduplication is to minimize repositing cost by putting away more information on circles. Deduplication decreases the info/yield proportion and serves to bring down the expense of capacity and data existance. Deduplication likewise serves to retrieve the record or whole framework at certain point in time. If there should be an occurrence of deduplication on records, whole document is utilized for approval if whatever other record with comparative information is available or not. In the event that comparable duplicate is discovered then another duplicate of same document is not put away. Point of inclination in document level deduplication is it needs less

metadata data and nearly simple to actualize and keep up. In the event of piece level deduplication, document is isolated into pile of same sizes or different sizes. In deduplication, every piece is utilized for approval. On the off chance that comparable piece (of same or other record) is discovered then deduplication just stores a reference to this pile rather than its genuine substance.

Conventional encryption (Encryption) does not work completely for deduplication as client encode their information with their individual keys and therefore duplicate information will have different figure content and deduplication becomes complicated. Merged encryption [2] is generally used to accomplish deduplication and not too bad privacy of information. In merged key encryption same key is utilized to encode and alter the information as key is produced utilizing crypto realistic hash estimation of information. Since centralized key is obtained from information, it will produce the identical figure content for comparative information. This serves to accomplish deduplication on cloud. Downside of centralized encryption is liable to animal power assault for information or records falling into known sets. In run of the mill repositing framework with deduplication, first customer will just send the hash estimation of the record then server will check if that hash esteem as of now exists in its database. In the event that document is as of now present on the server it requests that customer not to send the record and imprints customer extra proprietor of the record. In this way customer side deduplication prompts security issue and could discover other customer who has same document of sensitive/delicate data. This issue can be tended to by confirmation of proprietorship convention (PoW)[3]. POW is in two sections and it's between two players on basic info document. In first step verifier rundowns to itself and create sort data "v". In later step prover and verifier take part in intelligent convention where verifier has sort data "v" and prover has record "F" toward the end verifier either acknowledges or rejects it. Restriction with POW convention can't assist with the differential duplicate check, which is valuable in numerous applications. In framework with approval and deduplication, client is appointed as situated with benefits when users are included. Every record added to cloud is likewise relegated situated of benefits that indicate which sort of client is permitted to execute duplicate check and permitted to get to the documents.

III. Methods Used In Secure Deduplication

Following are the secure primitive used in the secure deduplication

3.1 Symmetric Encryption

Symmetric encryption uses a common secret key k to encrypt and decrypt information. A symmetric encryption scheme made up of three primary functions.

- KeyGen SE (1λ)→ k is the key generation algorithm that generates k using security parameter 1λ ;
- Enc SE (k, M)→ C is the symmetric encryption algorithm that takes the secret k, and message M and then outputs the ciphertext C, and
- Dec SE (k, C) → M is the symmetric decryption algorithm that takes the secret k and ciphertext C and then outputs the original message M.

3.2 Convergent Encryption

Convergent encryption provides data confidentiality in Deduplication. A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derive tag for the data copy, such that to detect duplicates tag will be used Here, we assume that the tag holds the property of correctness , i.e., if two data copies are the same, the tags of the data also same. The user first sends the tag to the server side to check if the identical copy has been already stored for detect duplicates.[4].

3.3 Proof of Ownership

The notion of proof of ownership (PoW) enables users to prove their ownership of data copies to the storage server. Specifically, Proof of ownership is implemented as an interactive algorithm run by a user and a storage server.

Conclusion:

n this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format,and also in private cloud our key is store with respective file.There is no need to user remember the key.So without key anyone can not access our file or data from public cloud.

In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format,and also in private cloud our key is store with respective file.There is no need to user remember the key.So without key anyone can not access our file or data from public cloud

References:

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with
encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.

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

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked
encryption and secure deduplication. In *EUROCRYPT*, pages 296–
312, 2013.
[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for
identity-based identification and signature schemes. *J. Cryptology*,
22(1):1–61, 2009.
[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes:
Proofs of security against impersonation under active and concurrent
attacks. In *CRYPTO*, pages 162–177, 2002.
[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin
clouds: An architecture for secure cloud computing. In *Workshop
on Cryptography and Security in Clouds (WCSC 2011)*, 2011.
[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.
Reclaiming space from duplicate files in a serverless distributed
file system. In *ICDCS*, pages 617–624, 2002.
[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15th
NIST-NCSC National Computer Security Conf.*, 1992.
[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.
[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of
ownership in remote storage systems. In Y. Chen, G. Danezis,
and V. Shmatikov, editors, *ACM Conference on Computer and
Communications Security*, pages 491–500. ACM, 2011.
[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication
with efficient and reliable convergent key management. In *IEEE
Transactions on Parallel and Distributed Systems*, 2013.
[13] libcurl. http://curl.haxx.se/libcurl/.
[14] C. Ng and P. Lee. Revdedup: A reverse deduplication storage
system optimized for reads to latest backups. In *Proc. of APSYS*,
Apr 2013.
[15] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication
protocols in cloud storage. In S. Ossowski and P. Lecca, editors,
Proceedings of the 27th Annual ACM Symposium on Applied Computing,
pages 441–446. ACM, 2012.
[16] R. D. Pietro and A. Sorniotti. Boosting efficiency and security
in proof of ownership for deduplication. In H. Y. Youm and
Y. Won, editors, *ACM Symposium on Information, Computer and
Communications Security*, pages 81–82. ACM, 2012.
[17] S. Quinlan and S. Dorward. Venti: a new approach to archival
storage. In *Proc. USENIX FAST*, Jan 2002.

V. RIYANKA is Pursuing M.Tech (Computer Science and Engineering), Sir C.R.Reddy College of Engineering Eluru, W.G.DIST., Andhrapradesh, India.

J.S.V.Gopala Krishna had Received M.Tech(CSE), JNTU Hyderabad. He is currently working as Associate. Professor in Sir C R Reddy College of Engineering , in the Department of Computer Science and Engineering, Eluru, W.G Dist, Andhra Pradesh, India. His Area of Specialization is Data mining and artificial intelligence.