International Journal For Technological Research In Engineering
Volume 3, Issue 7, March-2016 *ISSN (Online): 2347 - 4718*

SURVEY ON HYBRID CLOUD APPROACH FOR SECURE AUTHORIZED DEDUPLICATION

A.S. Raghuram[1], Hamsaveni.M[2]
[1]M.Tech Student, [2]Assistant Professor
Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysore.

Abstract: The cloud storage has become very popular in these days due its advantages like sharing of data among different geographical locations .But in most of the organizations, the storage systems can contain many duplicate copies of many pieces of same data. For example, same file can be saved by the different users using different names of the same document, or two or more files that do not match the name sometimes will be consisting of same as that of previous file. Deduplication eliminates these extra copies by keeping one copy of the data and making the other copies to point with pointers that point to the original copy. Companies regularly uses this deduplication process in case of backup and disaster recovery applications, but this isalso be used to free up space in storage as well. To avoid this duplicate data and to maintain the confidentiality in the cloud there are many methods proposed. To protect the sensitive data while supporting deduplication, many more techniques have been proposed.
Keywords: Deduplication, confidentiality.

I. INTRODUCTION

Cloud computing is a process of providing unlimited virtualized resources to the cloud users in the form of service which they are connected across the Internet, with hiding platform and the implementation details. In these days cloud service providers (CSP) offers both high storage capacity and parallel computing relatively at lower costs. As the cloud computing becoming prevalent, there is a significant increase in amount of data being stored in cloud and the data stored is shared by users with the specified privileges, which also define the access rights of the stored data. One of the critical challenges in the cloud storage is the management of the increasing data in the cloud. In cloud computing, the data must be scalable to achieve this well-known technique is used called data deduplication. Data deduplication is a process of data compression technique for eliminating the duplicate copies of repeating data in storage. The data compression technique is used to improve utilization of storage and can also be applied to network data transfers to reduce the number of bytes sent. rather maintaining multiple data copies with the same data, this process eliminates same data by keeping only one physical file and then referring other copied data to that copy. Deduplication can also be applied at the file level or the block level. The file level deduplication eliminates the duplicate copy of the same file. This process can also be used at block level, which removes the redundant blocks of data that occur in non-identical files.

II. DIFFERENT APPROACHES:

A. Hybrid cloud

The data deduplication is the process of removing the extra copy of same data making link to the new user, in this new user is also considered as the owner of that data. In this approach the owner who uploads the data first generates the hash value based on the document using the function File Tag(File) - It computes SHA-1 hash of the File as File Tag. Once the hash value is generated the owner uploads the hash value to the public cloud where it checks whether the hash value exists or not if it exists, then return with the file exists and ask the user to upload the membered and password. If the file does not exist then it returns file does not exist upload the file with memid, file and hash value generated. File Encrypt (File) - It encrypts File with the Convergent key Encryption technique using 256-bit AES algorithm. In the cipher block chaining (CBC) mode, the convergent key is taken from SHA-256 Hashing of the file. Once the file is encrypted then it is uploaded using the function File Upload Req (FileID, File, Token) this uploads File to the public Server if the file is not uploaded or Unique and updates the File Token stored. Once the document is stored we have to take care of security. Figure 2.1 explains the whole deduplication process. In the public cloud all the user of that space can access that document. to overcome this problem the convergent key is encrypted with the public key each of the user who has to have an access to the document, then that encrypted convergent key is uploaded is stored in the private server. This can be done using the function the function used is Share Token Req (Tag, {Priv.}) – this function is used requests the private server to generate the Share File. The user whose key stored in the private server can be used to view the document. by this we are achieving both security and deduplication.

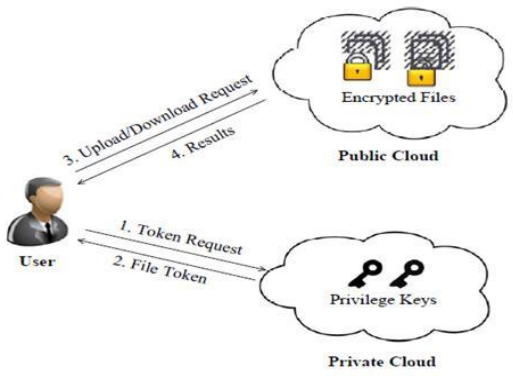

Figure 2.1 :-Architecture for Authorized Deduplication

B. Proof of data by data owner

In this paper [1][7]they achieve deduplication by checking the proof of data by the data owner. At the time of uploading a file to the cloud the file is provided with the set of privileges so that only a privileged user can access the file or check the duplicate files. Before checking the duplicate copy of the file the user has to take the file by his own privileges as his input. The user can get the duplicate copy if and only if it is present and the user has the privilege to access it.

Here they are using the common secret key k to encrypt as well as decrypt data. This common key is used for both converting the plain text to the cipher text andvice versa.

Here we have used three functions,

KeyGenSE: k is called the key generation algorithm that generates κ.

Once the key is generated using that key the files are encrypted and then that files are uploaded to the cloud.

EncSE (k, M): C is the symmetric encryption algorithm that takes the secret key K and message to be encrypted M and it gives the output cipher text value C;

Once the data is encrypted, to access the file the user has to use the same key decryption or to check the file is another copy or not.

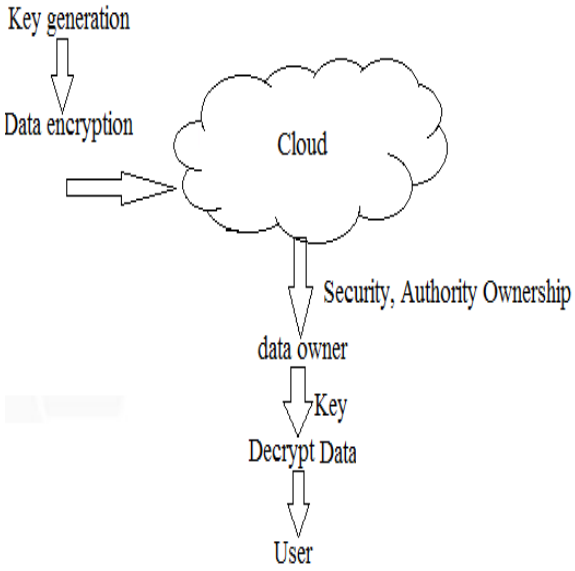

Figure 2.2 encryption and decryption of file

DecSE (k, C): M is the symmetric decryption that inputs secret keykand ciphertext C generated while encryption and then outputs the original message M.

In addition to the convergent key generated for both encryption and decryption it also generates a tag which is useful for finding the duplicate files in the cloud thus removing it from the cloud by making reference to the same file.

C. Message locked Encryption:

The convergent encryption [2][6] techniques provide confidentiality of data to the user's outsourced data which will be stored on the public clouds. While providing the confidentiality to the data it must also compatible with the data deduplication. In this technique encryption key is derived from the message. Itsupports data deduplication, because when we use same file, generate the same key so it will get same ciphertext which makes data deduplication possible. In this paper they proposed a new cryptographic technique which overcomes the drawback of previous encryption technique Massage Locked Encryption. In this key for encryption and decryption is derived from the message itself. The key is generated for a specific text which is uploaded file then with that key which is called a convergent key. With that the file is encrypted and it is called message locked encryption (MLE) .once the encryption is done then it is uploaded in to the cloud and with the same key the file is decrypted.

D. Proof of ownership

In this paper [3][8] of proof of ownership describes proofs of retrievability (PORs). A POR scheme enables archive/ back-up service (prover) to get a concise proof that a user (verifier) can retrieve a target file F, which is, that archive retains and transmits file data sufficient for the user to recover F in its entirety. A POR can be viewed as a cryptographic proof of knowledge (POK), but one specially designed to handle a large file (or bit string) F. To overcome attacks, they introduce the notion of proofs-of ownership (POWs), which makes clientto efficiently prove to a server that the client holds a file, rather than just some short information. They use the concept of proof-of-ownership, which consists ofrigorous security definitions, and rigorous efficiency requirements of Petabyte scale storage systems.

E. Dekey technology

In this paper [4]they discuss the problem of achieving key management in secure deduplication. Firstly they use a primary approach in which each of the users has an independent master key for document encryption using convergent keys and outsource it. However, using such management scheme generates number of keys with the increasing number of users protect the master keys.

In this scheme users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, they implement Dekey using Ramp secret sharing scheme and demonstrate Dekey incurs limited overhead in realistic environments. Specifically, the (n, k, r) RSSS (where n > k > r>= 0) generates and shares from an secret key with which First secret key can be recovered from any key K shares but cannot be recovered from fewer than key k shares, and second no information about the secret can be got from the r shares.When r = 0, the (n, k, O) RSSS becomes the (n, k) Rabin's Information Dispersal Algorithm (IDA) when r =k-1, the (n, k, k – 1)-RSSS becomes the (n, k) Shamir's Secret Sharing Scheme.

III. CONCLUSION

In this paper we have surveyed the various techniques available for secure data deduplication. We have presented the various techniques for deduplication using hybrid cloud.

REFERENCES

[1] Gaurav Kakariya,Prof. Sonali Rangdale A Hybrid Cloud Approach for Secure Authorized Deduplication. In International Journal of Computer Engineering and Applications, October 14

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management, In IEEE transactionson parallel and Distributed Systems2013.

[3] 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, 201

[4] A. Yun, C. Shi, and Y. Kim, On Protecting Integrity and Confidentiality of Cryptographic File System for Outsourced Storage, in Proc. ACM CCSW, Nov. 2009, pp. 67-76.

[5] M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, Secure Data Deduplication, in Proc. Storage SS, 2008, pp. 1-10.

[6] M.Bellare, Keelveedhi, and Ristenpart, Message-Locked Encryption and Secure Deduplication, in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631.

[7] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou A Hybrid Cloud Approach for Secure Authorized Deduplication, In IEEE Transactions on Parallel and Distributed Systems, 2014.

[8] 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.

