International Journal For Technological Research In Engineering
Volume 4, Issue 4, December-2016 *ISSN (Online): 2347 - 4718*

ENHANCING THE CLOUD STORAGE MANAGEMENT BY USING AUTHORIZED DATA DE-DUPLICATION

Bantu Sneha[1], E. Krishnaveni Reddy[2]
[1]M. Tech Student, [2]Head of the Department & Associate Professor,
Department of CSE, Sridevi Women's Engineering College, Village Vattinagulapally, Mandal Rajendranager, District Ranga Reddy, Telangana, India.

Abstract: ***Data de-duplication may be a method for reducing the quantity of space for storing an organization needs to save its information. In most organizations, the storage systems contain copies of the various things of data. As an example, identical file is in addition saved in several altogether different places by different users, additional files that are not identical should still represent abundant of identical information. De-duplication eliminates these auxiliary duplicates by save only one original copy of the data and commutation the choice copies with pointers that lead back to the first copy. Companies frequently use de-duplication in backup and failure recovery applications; however it is used to unlock house in primary storage in addition. To avoid this duplication of data and to keep up the confidentiality at intervals the cloud we tend to tend to victimization the conception of Hybrid cloud. To safeguard the confidentiality of sensitive data whereas supporting de-duplication, the targeted secret writing technique has been planned to place in writing in code the information before outsourcing. To higher defend data security, this paper makes the primary plan to formally address the matter of approved data de-duplication.***

Index Terms: ***Hybrid cloud, Convergent Encryption, De-duplication, Proof of Ownership;***

I. INTRODUCTION

Cloud computing is obtaining more and more widespread because it will provide low-cost and on demand use of huge storage and method resources. As the volume of data grows, in addition increasing is the Total price of ownership that includes storage infrastructure worth, management worth and human administration price. therefore in cloud storage systems, reducing the number of data that need to be stored, transferred and managed becomes a crucial. As a result, data De-duplication is a necessary and widespread cost-saving feature for cloud storage. The term data de-duplication refers to techniques that store just one copy of unessential data, and provides links to that duplicate instead of storing alternative actual copies of this data. With the conversion of services from tape to hard disk or disk or floppy, data de-duplication has become a key component in the backup method. By storing and forwarding only one copy of duplicate data, de-duplication offers savings of every disk space and network metric. De-duplication could be a one among the necessary, technique to reduce storage space and transfer metric and has been used to make data management scalable. As a different of keeping varied data copies with the identical content, de-duplication eliminates unused data by keeping only one physical copy and referring alternative unused data to that copy. There are two types of de-duplication check one is file-level de-duplication and another is block-level de-duplication. Among that file-level de-duplication introduce the complete file where as block-level de-duplication refers to the fastened or variable size data block. to make de-duplication secure we have to use certain security mechanism like coding. Traditional coding wants whole different users to cipher their data with their own keys, therefore identical data copies of totally different users can cause different cipher text and for this reason de-duplication is incompatible with traditional cryptography. Convergent cryptography provides a possible chance to implement data confidentiality whereas produce the de-duplication. Convergent cryptography, a cryptosystem that produces in distinguishable cipher text files from identical plaintext files, regardless of their cryptography keys it encrypts or decrypts a data with a coding key, which is derived by computing the encoded hash value of the content of the data copy itself. Once key generation and data cryptography, users retain the keys and send the encoded-text to the cloud. Since coding is deterministic, identical data copies can generate constant convergent key and the same encoded-text. This permits the cloud to perform de-duplication on the encoded text. The encoded-texts can only be decrypted by the corresponding data owners with their merging keys. To prevent uncertified access, a secure proof of ownership (POW) protocol is in addition needed to provide t he proof that the user therefore owns an identical file once a duplicate is found. Once the proof, subsequent users with the same file are assign a pointer from the server without having to transfer a similar file. A user can download the encoded file with the pointer from the server, which can only be decoded by the consequent or corresponding data owners with their own convergent keys. Thus, convergent secret writing permits the cloud to perform de-duplication on the cipher texts and also the proof of owner-ship stop the uncertified user to access the file by using convergent encoding technique we are able to detect duplicate files moreover file compression. Through our implementation we can detect the duplicate files still as we are able to increase the storage space of cloud.

II. RELATED WORK

However, the traditional de-duplication systems cannot support differential authorization duplicate confirm is very

important in many applications. In such an authorized de-duplication system, every user is issued a bunch of privileges throughout system information. Every file uploaded to the cloud is in addition delimited by a bunch of privileges to specify which sort of users is allowable to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user should take this file and users own privileges as inputs. The user is in a very position to seek out a duplicate for this file if and providing there is a replica of this file and a matched privilege keeps in cloud. For example, in a passing company, many different privileges are about to be assigned to workers. therefore on avoid wasting value and with efficiency management, the data are planning to be affected to the storage server provider (S-CSP) inside the general public cloud with nominal privileges and therefore the de-duplication technique are about to be applied to store only one copy of constant file. As a result of privacy thought, some files are progressing to be encrypted and allowed the duplicate check by workers with fixed privileges to understand the access management. Ancient de-duplication systems supported targeted encryption, although providing confidentiality to some extent, do not support the duplicate refer to differential privileges. In many words, no differential privileges are thought of within the de-duplication supported targeted cryptography technique. It appears to be contradicted if we might prefer to understand each de-duplication and differential authorization duplicate check at constant time. Symmetric coding uses a typical secret key κ to encipher and decode data. A symmetric coding theme consists of three primitive functions: Key-GenSE(1 λ)= κ is that the key generation rule that generates κ using security parameter one λ . Enc-SE(κ,M)= C is that the symmetric coding algorithm that takes the key κ and message M thus outputs the cipher text C. Dec-SE(κ,C)= M is that the symmetric decipherment algorithm that takes the key κ and cipher text C thus outputs the initial message M. convergent encoding ,provides data confidentiality in de-duplication. A user (or data owner) derives a convergent key from every original data copy and encipher the data copy with the convergent key. To boot, the user together derives a tag for the data copy, fixed the tag are accustomed notice duplicates. Here, we have we tend to assume that the tag correctness property holds, if a combine of information copies unit of measure an equivalent, then their tags are an equivalent. To find duplicates, the user initial sends the tag to the server aspect to look at if the identical copy has been already kept. Note that each the convergent key and additionally the tag are severally derived, and together the tag cannot be accustomed deduce the convergent key and compromise data confidentiality. Each the encrypted data copy and its corresponding tag are unbroken on the server aspect. The notion of proof of possession permits users to prove their possession of data copies to the storage server. An identification protocol ∏ are going to be delineated with a try of phases: Proof and Verify. Inside the stage of Proof, a User U will demonstrate his individuality to an admirer by acting some identification proof associated to his identity. The input of the user is his non-public key that's sensitive data like non-public key of a public key in his certificate or master-card varies etc. that he would not prefer to share with the opposite users.

III. FRAME WORK

As the name suggests a Hybrid cloud is a mixture of every a public and private cloud as an example a corporation may prefer to place their in operation settings in very public cloud whereas the event surroundings is additionally placed throughout a very personal cloud. additionally several organizations prefer to run their sales and marketing operations throughout a public cloud, whereas keeping their cash operations among a private cloud another choice is two run twin systems. Personal cloud we tend to are storing encoded keys. Private cloud is high flexibility and high price and high secure than compare to public cloud. Public Cloud we tend to are storing the encoded text or unclear text. Public cloud is low flexibility and low price and less secure than compare to personal cloud. The Secure-Cloud Storage providers are an entity that provides the secure data storage service for the users. Within the de-duplication system, once users own and store identical content, the S-CSP will only store one copy of these files and retain only distinctive data.

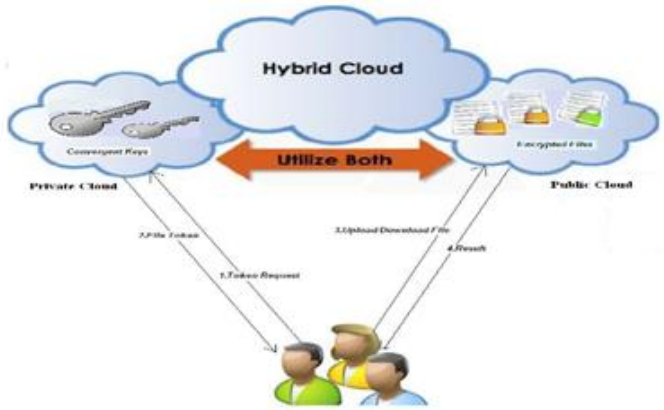


Figure 1: Architecture of Hybrid Cloud

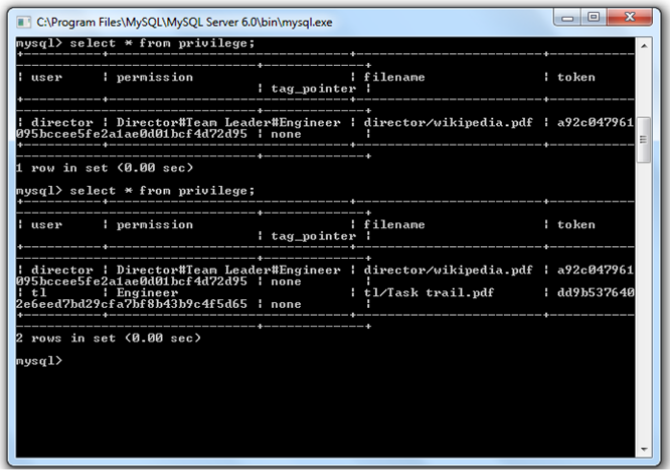
A de-duplication technique, on the other hand, can reduce the storage price at the server aspect and save the transfer bandwidth at the user aspect. For backup and confidentiality of data storage, we tend to consider a gathering of Secure-cloud storage providers. Data De-duplication involves finding and removing of duplicate data while not considering its fidelity .Here the goal is to store lots of data with less bandwidth and cost. Convergent coding provides data confidentiality in de-duplication. A user (or data owner) derives a convergent key from every original data copy and encrypts the data copy with the convergent key. The basic arrange of convergent encryption (CE) is to derive the cryptography key from the hash of the legible text. The only implementation of convergent cryptography are going to be defined as follows: Alice derives the cryptography key from her file M such that K = H (M), where H is a cryptographic hash perform he will encode the message with this key, hence: C = E (K; M) = E (H (M), where E is a block encoded. By applying this technique, two users with two identical plain texts can acquire two identical encoded texts since the cryptography key is the same; thus the cloud

storage :possession, is also able to perform de-duplication on such encoded texts. Moreover, encryption keys are generated, maintained and protected by users. As the encoded key is deterministically generated from the plain text, users do not need to act with each other for agreement on the key to cipher a given plaintext. Therefore, convergent cryptography looks to be a wise candidate for the adoption of cryptography and de-duplication in the cloud storage domain. In addition, the user verify a tag for the data copy, specified the tag are used to detect duplicates. A convergent cryptography theme are going to be defined with four primitive functions: Key Gen(M) → K is the key generation algorithm that maps a data copy M to a merging key K Encrypt(K,M) → C is the original cryptography formula that takes each the encoded key K and thus the data copy M as inputs then outputs a encoded text Decrypt(K,C)→M is the decoding algorithm that takes each the cipher text C and therefore the convergent key K as inputs then outputs the initial data copy M Tag-Gen(M) →T(M) is the tag generation algorithm that maps the original data copy M and outputs a tag T(M). We tend to enable Tag-Gen to generate a tag from the corresponding cipher text by using T (M) =Tag-Gen(C), where C=Encrypt (K, M). The notion of proof of rights enables users to prove their ownership of data copies to the storage server.

IV. EXPERIMENTAL RESULTS

In our experiments, any number of users can registered and login into the system. Who are authorized users they can upload the files into the cloud. Any user can give the access permission to other authorized users after upload the files. These uploaded files are stored in public cloud and keys are stored in private cloud. If any duplicate files are available in public cloud, then that file cannot uploaded in the cloud and to that particular file tag pointer will be assigned to the user. But that file can be downloaded by data owner as well as data users.

Below image shows that the tag pointers and access permission can view in the database;



Through our implementation we can store the data into the cloud in public cloud and encoded keys are stored in private cloud also detect the duplicate files as well as we can increase the storage space of cloud and also we can decrease the network bandwidth and cost.

V. CONCLUSION

Cloud computing has reached a majority, that leads it into a productive section. This suggests that the majority of the main issues with cloud computing are addressed to a degree that clouds have become interesting for full industrial exploitation. This however will not mean that all the issues listed above have truly been solved, only that the according risks will be tolerated to a certain degree. Cloud computing is so still as much a research topic, because it could be a market providing. For higher confidentiality and security in cloud computing we tend to have projected new de-duplication constructions supporting approved duplicate check also as file compression in hybrid cloud design, in this the duplicate-check tokens of files are generated by private cloud server with non-public keys. Planned system includes proof of data owner therefore it will facilitate to implement higher security issues in cloud computing.

REFERENCES

[1]. ANDERSON, P., ANDZHANG, L. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA (2010).

[2]. BELLARE, M., BOLDYREVA, A. Deterministic and efficiently searchable encryption. In CRYPTO 2007(Santa Barbara, CA, USA, Aug. 1923,2007), A. Menezes, Ed. ,vol. 4622 of LNC, Springer, Berlin, Germany, pp. 535–552.

[3]. MKEELVEEDHI, S.ANDRISTENPART, T. Message-locked encryption and secure de-duplication. In EU-ROCRYPT 2013, to appear. Cryptology ePrint Archive, Report 2012/631, November 2012.

[4]. Pasqualo Puzio, Refik Molva, MelekOnen ,"Cloud Dedup: Secure Deduplication with Encrypted Data for Cloud Storage", SecludIT and EURECOM

[5]. W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication proto cols in cloud storage," in Proc. 27th Annu. ACM Symp. Appl. Com-put., 2012, pp. 441–44.

[6]. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proc. 4th ACM Int. Workshop Storage Security Survivability, 2008, pp. 1–10.

[7]. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. ACM Conf. Com-put. Commun. Security, 2011.

[8]. M. Shyamala Devi, V.Vimal Khanna, Naveen Balaji "Enhanced Dynamic Whole File De-Duplication for Space Optimization in Private Cloud-Storage Backup",IACSITAugust,2014.

[9]. Weak Leakage-Resilient Client–Side deduplication of Encrypted Data in Cloud Storage" Institute for Info Comm. Research,Singapore,2013

[10]. KEELVEEDHI, S., ANDRISTENPART, T. Message-locked encryption and secure deduplication. In EU-OCRYPT 2013, to appear.

Cryptology ePrint Archive, Report2012/631, November 2012.
[11]. Open SSL Project. http://www.openssl.org/.
[12]. GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.
[13]. C. Ng and P. Lee, "Revdedup: A reverse deduplication storage system optimized for reads to latest backups," in Proc. 4th Asia-Pacific Workshop Syst., http://doi.acm.org/10.1145/2500727.2500731, Apr. 2013
[14]. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models, "IEEE Comput., vol. 29, no. 2,pp. 38–47, Feb. 1996
[15]. J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," IACR Cryptology ePrint Archive, 2013:149, 2013.

