INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 4, Issue 4 AUG 2015

Authorized Data to Protect Data Security by Using Hybrid Cloud Method

[1] N. Naveen Kumar, [2] Kiran Donupala

[1] Assistant Professor, Department of SE, School of Information Technology, JNTUH, Kukatpally, Hyderabad, Telangana state, India.

[2] M.Tech Student, Department of SE, School of Information Technology, JNTUH, Kukatpally, Hyderabad, Telangana state, India.

Abstract— Data deduplication is one among vital knowledge compression techniques for eliminating duplicate copies of repetition data, and has been wide utilized in cloud storage to cut back the number of space for storing and save information measure. to shield the confidentiality of sensitive knowledge whereas supporting deduplication, the oblique secret writing technique has been projected to encrypt the info before outsourcing. totally different from ancient deduplication systems, the differential privileges of users square measure further thought-about in duplicate check besides the info itself. we tend to additionally gift many new deduplication constructions supporting licensed duplicate sign in a hybrid cloud design. To better shield knowledge security, this paper makes the primary arrange to formally address the matter of licensed knowledge deduplication. Security analysis demonstrates that our theme is secure in terms of the definitions per the projected security model. As an indication of thought, we tend to implement a epitome of our projected licensed duplicate check theme and conduct test bed experiments mistreatment our epitome. we tend to show that our projected licensed duplicate check theme incurs minimal overhead compared to traditional operations.

Index Terms— Deduplication, authorized duplicate check, confidentiality, hybrid cloud.

I. INTRODUCTION

In computing, knowledge deduplication may be a specialised knowledge compression technique for eliminating duplicate copies of repetition knowledge. connected and somewhat synonymous terms area unit intelligent (data) compression and single-instance (data) storage. this system is employed to improve storage utilization and may even be applied to network knowledge transfers to scale back the number of bytes that has got to be sent. within the deduplication method, distinctive chunks of information, or byte patterns, area unit known and hold on throughout a process of study. since the revision persist, new hunk ingredient piece match up to the hold on copy and whenever a match happens, the redundant chunk is replaced with atiny low reference that points to the hold on chunk. only if identical computer memory unit pattern may crop up dozens , hundreds, or possibly thousands of times the measure of information that has got to be hold on or transferred is greatly reduced.

A Hybrid Cloud may be a combined variety of personal clouds and public clouds within which some grave familiarity resides within the enterprise's personal cloud whereas different knowledge is hold on in and accessible from a public cloud. Hybrid clouds obtain to deliver the benefits of quantifiability, responsibleness, rapid readying and potential value savings of public clouds with the protection and enhancedcontrol and management of personal clouds. As cloud computing becomes using, associate increasing amount of knowledge is being hold on within the cloud and employed by users with gave privileges, which define the access rights of the hold on knowledge. The important challenge of cloud storage or cloud computing is that the management of the continuously increasing volume of knowledge. info deduplication or Single Instancing primarily refers to the elimination of redundant knowledge. within the deduplication method, duplicate knowledge is deleted, going away just one copy (single instance) of the info to be hold on. However, categorisation of all knowledge remains maintained ought to that knowledge ever be needed. generally the info deduplication eliminates the duplicate copies of continuation knowledge.

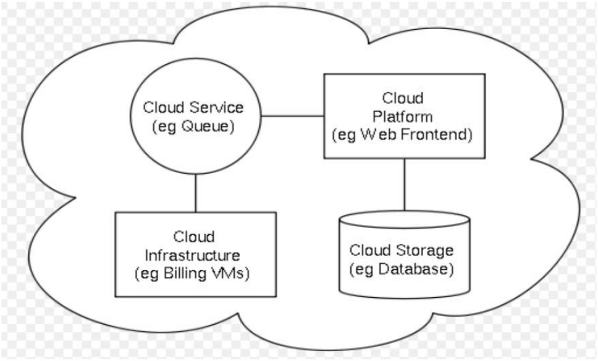


Figure: 1. Architecture of cloud computing.

The data is encrypted before outsourcing it on the cloud or network. This cryptography requires longer and area necessities to cypher knowledge. just in case of enormous knowledge storage the encryption becomes even additional complicated and demanding. By exploitation the info deduplication within a hybrid cloud, the crytography can become easier. As we have a tendency to all grasp that the network is contains exuberant quantity of knowledge, that is being shared by us rs and nodes within the network. several massive scale network uses the info cloud to store and share their knowledge on the network. The node or user, that is gift within the network have full rights to transfer or transfer knowledge over the network. however persistently completely different user uploads the same knowledge on the network. which is able to produce a duplication within the cloud. If the user desires to retrieve the info or transfer the info from cloud, anytime he must use the 2 encrypted files of same knowledge. The cloud can do same operation on the 2 copies of knowledge files. attributable to this the data confidentiality and therefore the security of the cloud get desecrated. It creates the burden on the operation of cloud.

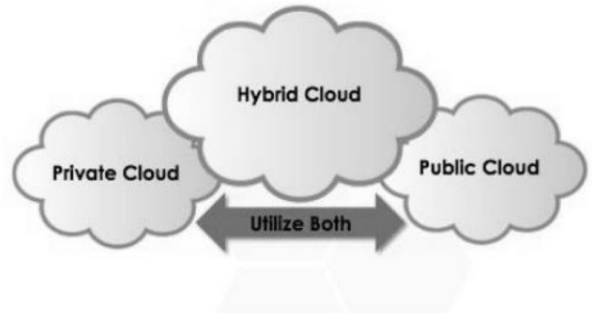


Figure: 2. Architecture of Hybrid cloud.

To avoid this duplication of knowledge and to take care of the confidentiality within the cloud we have a tendency to using the conception of Hybrid cloud. it's a mixture of public and personal cloud. Hybrid cloud storage combines the benefits of measurability, responsibleness, fast readying and potential price savings of public cloud storage with the safety and full management of personal cloud storage.

II. RELATED WORK

A However, previous deduplication systems cannot support differential authorization duplicate check, that is vital in many applications. In such a licensed deduplication system, every user is issued a group of privileges throughout system initialization. every file uploaded to the cloud is additionally delimited by a group of privileges to specify which type of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for a few file, the user has to take this file and his own privileges as inputs. The user is in a position to seek out a reproduction for this file if and provided that there's a duplicate of this file and a matched privilege keep in cloud. for instance, in an exceedingly company, many various privileges are going to be allotted to employees. so as to avoid wasting value and with efficiency management, the information are going to be affected to the storage server supplier (S-CSP) in the public cloud with nominal privileges and therefore the deduplication technique are going to be applied to store only 1 copy of constant file. attributable to privacy thought, some files are going to be encrypted and allowed the duplicate check by staff with specified privileges to understand the access management. ancient deduplication systems supported focused cryptography, although providing confidentiality to some extent, don't support the duplicate consult with differential privileges. In different words, no differential privileges are thought of within the deduplication supported focused cryptography technique. It appears to be contradicted if we would like to understand each deduplication and differential authorization duplicate check at constant time.

A. Symmetric Encryption

Symmetric encryption uses a common secret key κ to encrypt and decrypt information. A symmetric encryption scheme consists of three primitive functions:

KeyGenSE(1 λ)= κ is the key generation algorithm that generates κ using security parameter 1 λ.

EncSE (κ,M)= C is the symmetric encryption algorithm that takes the secret κ and message M and then outputs the ciphertext C.

DecSE(κ,C)= M is the symmetric decryption algorithm that takes the secret κ and ciphertext C and then outputs the original message M.

B. Convergent Encryption

Convergent encoding ,provides information confidentiality in deduplication. A user (or information owner) derives a confluent key from every original information copy and encrypts the info copy with the confluent key. additionally, the user conjointly derives a tag for the data copy, specified the tag are accustomed find duplicates. Here, we have a tendency to assume that the tag correctness property holds, i.e., if 2 information copies square measure a similar, then their tags square measure a similar. To find duplicates, the user initial sends the tag to the server side to envision if the identical copy has been already keep. Note that each the confluent key and also the tag square measure severally derived, and also the tag can\'t be accustomed deduce the confluent key and compromise information confidentiality. each the encrypted information copy and its corresponding tag are keep on the server facet.

C. Proof of Ownership

The notion of proof of possession (PoW) permits users to prove their possession of knowledge copies to the storage server. Specifically, prisoner is enforced as AN interactive algorithmic program (denoted by PoW) go by a prover (i.e., user) and a admirer (i.e.,storage server). The admirer derives a brief worth Φ(M) from a knowledge copy M. To prove the possession of the info copy M, the prover must send Φ' to the admirer specified $\Phi' = \Phi$ (M). The formal security definition for prisoner roughly follows the threat model in a very content distribution network, wherever AN assailant doesn't grasp the complete file, however has accomplices World Health Organization have the file. The accomplices follow the “bounded retrieval model”, specified they'll facilitate the assailant get the file, subject to theconstraint that they need to send fewer bits than the initial min-entropy of the file to the assailant .

D. Identification Protocol

An identification protocol $\prod$ will be delineated with 2 phases: Proof and Verify. within the stage of Proof, a prover/user U will demonstrate his identity to a admirer by acting some identification proof associated with his identity. The input of the prover/user is his non-public key skU that's sensitive data like non-public key of a public key in his certificate or mastercard range etc. that he wouldn't wish to share with the opposite users. The admirer performs the verification with input of public data phenylketonuria related to skU. At the conclusion of the protocol, the admirer outputs either settle for or reject to denote whether or not the proof is passed or not. There area unit several economical identification protocols in literature, together with certificate-based, identity-based identification etc. .

III. FRAME WORK

In the proposed system we are achieving the data deduplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. 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 some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

a. ENCRYPTION OF FILES

Here we have a tendency to area unit victimisation the common secret key k to write in code similarly as decode knowledge. This will use to convert the plain text to cipher text and once more cipher text to plain text. Here we’ve used 3 basic functions,

KeyGenSE: k is that the key generation formula that generates κ victimisation security parameter one.

EncSE (k, M): C is that the bilaterally symmetric encoding formula that takes the key κ and message M and then outputs the ciphertext C;

DecSE (k, C): M is that the bilaterally symmetric decipherment formula that takes the key κ and cipher text C and then outputs the initial message M.

b. CONFIDENTIAL ENCRYPTION

It provides knowledge confidentiality in deduplication. A user derives a focused key from each original knowledge copy and encrypts the info copy with the focused key. additionally, the user additionally derives a tag for the info copy, specified the tag are wont to notice duplicates.

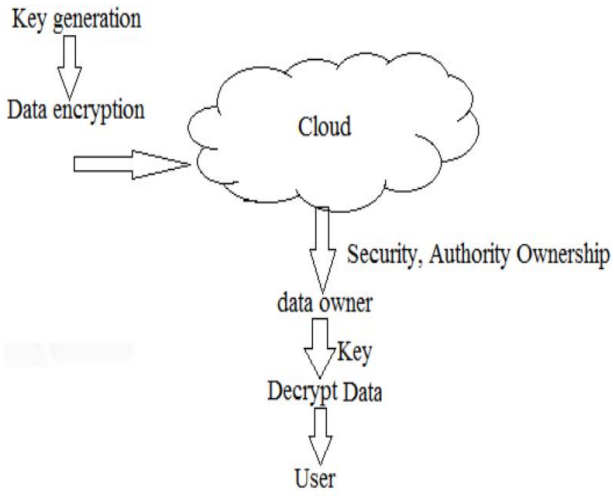


Figure: 3 confidential data encryption

c. PROOF OF DATA

The user got to prove that the information that he wish to transfer or transfer is its own data. meaning he got to offer the focused key and validating knowledge to prove his ownership at server.

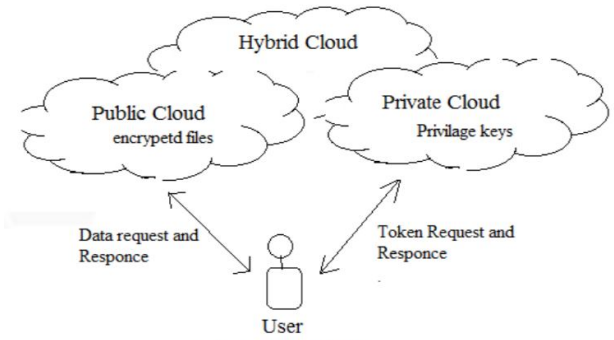
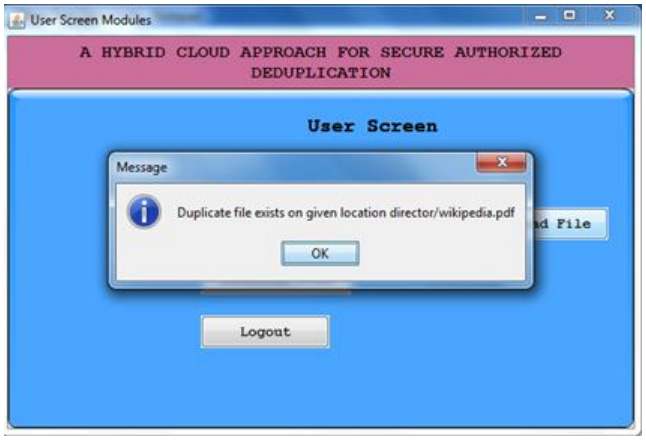


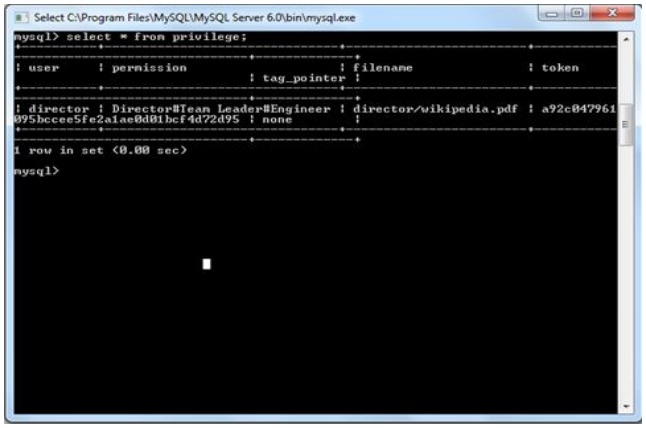
Figure: 4 System Architecture.

IV. EXPERIMENTAL RESULT

When you are trying to send something the duplicated data our proposed system will reject the process and it will shows the message as shown below.



When you are uploaded any data the data will be stores in the encrypted format as shown.



V. CONCLUSION

In this paper, the thought of approved data deduplication was projected to confirm the data security by counting differential advantages of purchasers within the copy check. we have a tendency to likewise displayed a number of new deduplication developments supporting approved copy weigh in cross breed cloud structural coming up with, during which the copy check tokens of records ar created by the non-public cloud server with non-public keys. Security investigation exhibits that our plans ar secure as way as corporate executive and outcast assaults determined within the projected security model. As a evidence of plan, we have a tendency to dead a model of our proposed approved copy check set up and behavior testbed analyses on our model. we have a tendency to incontestable that our approved copy check set up brings regarding insignificant overhead contrasted with united encoding and system exchange.

REFERENCES

[1] 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.

[2] 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.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.

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

[5] GNU Libmicrohttpd . http://www.gnu.org/software/libmicrohttpd/.

[6] 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.

[7] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

[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.

[9] libcurl. http://curl.haxx.se/libcurl/.

[10] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

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

[12] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan. Sedic: privacyaware data intensive computing on hybrid clouds. In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.