ISSN (Online): 2349-7084
GLOBAL IMPACT FACTOR 0.238
DIIF 0.876

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 470-474

Secure Redundant Data Avoidance over Multi-Cloud Architecture

[1]Kamarthi Rekha,[2] G.Somasekhar, [3]Dr S.Prem Kumar

[1](M.Tech), CSE, Assistant professor Department of Computer Science and Engineering

[2]Assistant Professor, Department of Computer Science and Engineering

[3]Professor & HOD, Department of computer science and engineering,

G.Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India.

Abstract— Redundant data avoidance systems, the Private Cloud are involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud, in this connection our presented system has follows traditional encryption while providing data confidentiality, is incompatible with redundant data avoidance. Identical data copies of different users will lead to different ciphertexts, making data avoidance impossible. To address above issues convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized redundant data avoidance. Different from traditional redundant data avoidance systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new redundant data avoidance constructions supporting authorized duplicate check in a multi-cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. In order to perform secure access controlling scheme user may satisfy fine-grained approach at cloud level towards access restricting from unauthorized users or adversaries.

KEYWORDS—Deduplication, authorized duplicate check, confidentiality, File level Check, Block Level Check, Convergent key, Metadata Supervisor.

———————————— ◆ ————————————

I. INTRODUTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services cross the entire Web, while concealing stage and usage points of interest. With the possibly unending storage room offered by cloud suppliers, clients tend to use as much space as they can and vendors always search for strategies meant to minimize repetitive information and amplify space investment funds. A system which has been broadly received is cross-client deduplication. The straightforward thought behind deduplication is to store copy information (either records or pieces) just once. Along these lines, if a client needs to transfer a document (square) which is as of now put away, the cloud supplier will add the client to the proprietor rundown of that record (piece). Deduplication has demonstrated to accomplish high space and expense investment funds and numerous distributed storage suppliers are right now embracing it. Deduplication can decrease stockpiling needs by up to 90-95% for reinforcement applications [11] and up to 68% in standard record frameworks [23]. Alongside low possession require the insurance of their information and secrecy ensures through encryption. Lamentably, deduplication and encryption are two clashing innovations. While the point of deduplication is to recognize indistinguishable information portions and store them just once, the aftereffect of encryption is to make two indistinguishable information sections indistinct in the wake of being scrambled. This implies that if information is encoded by clients in a standard manner, the distributed storage supplier can't make a difference deduplication since two indistinguishable information portions will be distinctive after Encryption. Then again, if information is not scrambled by Information proprietors, confidentiality can't be ensured and information is not secured against inquisitive distributed storage suppliers. A procedure which has been proposed to meet these two clashing prerequisites is Convergent encryption whereby the encryption key is normally the consequence of the hash of the information portion. Albeit Convergent encryption is by all accounts a decent possibility to accomplish confidentiality and deduplication in the meantime, it sadly experiences different surely understood shortcomings [15], [24] word reference assaults: an aggressor why should capable figure or foresee a document can without much of a stretch infer the potential encryption key and check whether the record is officially put away at the distributed storage supplier or not. In this paper, we adapt to the inborn security exposures of focalized encryption and propose secure information deduplication instrument, which safeguards the consolidated focal points of deduplication and Convergent encryption.

ISSN (Online): 2349-7084

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 470-474

. Data owner can restrict unauthorized access rights by performing fine-grained access controlling scheme where data owner defined set of access attribute sets before outsourcing to CloudSever1, if any user wants to access that file user need to satisfy the data owner access attribute sets, if its matched then data owner allow him to access that data by sending set of access privileges. Data deduplication will be done on secured manner by proving proof of the ownership.

II RELATED WORK

In this section we review some related works concerned with security and privacy issues in cloud. Also, we discuss the work which adopt similar techniques as our approach but serve for different purposes.

SECURITY AND PRIVACY ISSUES IN CLOUD:

Only the authorized persons need to access the data from the cloud. In order to ensure the integrity of user authentication, need of security mechanism which will keep track usage of data in the cloud? As with all cloud computing security challenges, it's the responsibility of the user to ensure that the cloud provider has taken all necessary security measures to protect the user's data and the access to that data.

De-duplication is the technique that is most effective most widely used but when it is applied across the multiple users the cross-user deduplication tend to have to many serious privacy implications. Simple mechanisms can be used which can enable the cross-user deduplication which will reduce the risks of the data leakage.

III.BACK GROUND

In previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization. The overview of the cloud deduplication is as follow:

DEDUPLICATION

According to the data granularity, deduplication strategies can be categorized into two main categories: file-level deduplication [29] and block-level deduplication [17], which is nowadays the most common strategy. In block-level deduplication, the block size can either be fixed or variable [27]. Another categorization criterion is the location at which deduplication is performed: if data are deduplicated at the client, then it is called source-based deduplication, otherwise target-based. In source-based deduplication, the client first hashes each data segment he wishes to upload and sends these results to the storage provider to check whether such data are already stored: thus only ”undeduplicated” data segments will be actually uploaded by the user. While deduplication at the client side can achieve bandwidth savings, it unfortunately can make the system vulnerable to side-channel attacks [19] whereby attackers can immediately discover whether a certain data is stored or not. On the other hand, by deduplication data at the storage provider, the system is protected against side-channel attacks but such solution does not decrease the communication overhead.

CONVERGENT KEY ENCRYPTION

The basic idea of convergent key encryption (CKE) is to derive the encryption key from the hash of the plaintext. The simplest implementation of convergent encryption can be defined as follows: Data owner derives the encryption key from his/her message M such that K = H(M), where H is a cryptographic hash function; Data owner can encrypt the message with this key, hence: C = E(K, M) = E(H(M), M), where E is a block cipher. By applying this technique, two users with two identical plaintexts will obtain two identical ciphertexts since the encryption key is the same; hence the cloud storage provider will be able to perform deduplication on such ciphertexts. Furthermore, encryption keys are generated, retained and protected by users. As the encryption key is deterministically generated from the plaintext, users do not have to interact with each other for establishing an agreement on the key to encrypt a given plaintext. Therefore, convergent encryption seems to be a good candidate for the adoption of encryption and deduplication in the cloud storage domain.

IV.SYSTEM STUDY

4.1. EXISTING SYSTEM:

In our existing system, data deduplication performed at service provider level without considering user privileges, data get stored at cloud server level with related privileges keys. More over there is a lack of security while accessing from cloud servers due to weak access controlling schemes like coarse-grained approach was performed at client level.

There might be possibilities are there to access the data by adversaries. If data duplication occur at block level i.e. if the context of the file is same or File level i.e. name of the file is same then duplication functioning will be executed, in order to function data deduplication mechanism system has verify POW (Proof of the ownership), and then verify the label tags which are maintained by the cloud service provider.

DISADVANTAGES:

- Lack of user privacy
- Lack of data confidentiality
- Lack of data integrity
- Unsecured data duplication mechanism performed
- Redundant data avoidance systems cannot support differential authorization duplicate check

4.2. THE PROPOSED SYSTEM

The idea of a hybrid cloud is aimed to bridge the gap between high controls, high cost “CloudSever2” and highly callable, flexible, low cost “CloudSever1”. “CloudSever2” is generally used to illustrate a VMware deployment in which the hardware and software of the environment is used and manage by a single entity.

The concept of a “CloudSever1” regularly involves some form of elastic/subscription based resource pools in a hosting provider data center that utilizes multi-tenancy.

Previously we have file level deduplication for plainness. Simply we can say that deduplication is a process which eliminates the storage of any redundant files. Actually, block level deduplication can be easily deduced from file-level deduplication, explicitly, to upload a file, a user first performs the file-level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well, if not, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. Each data copy it may be a file or may be a block is associated with a token for the duplicate check.

4.3. System Implementations

Data Owner: - Data owner will make account in our application by using the registration form and by using the his/her user name and password he can login in to our application they can upload and download data from our cloud server the data will be provide security

ISSN (Online): 2349-7084

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 470-474

by encrypting the data in the files and a key will be generated for every file that upload in the cloud data base.

TAG GENERATION:

Data owner the moment to upload the file, owner will provide his identity with uploading files details like name of the file and pass his own privileges to the CloudSever2, encrypted formatted request will be send to the CloudSever2, then CloudSever2 will processed that request and response as tag file to the data owner, now data owner can send that tag to CloudSever1 for verification for the sake of identifying duplicate occurrences.

CloudServer1: - is the data storage server i.e. CloudSever1 service provides the data outsourcing service and stores the data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via redundant data avoidance and keeps only unique data.

Here CloudSever1 verifies data owner tag request with existing log tag details and send back response to the data owner with results, if duplicate occurs it will not allow to owner to upload same file, if duplicate not occurs then CloudSever1 allow to data owner to upload file ,while uploading data owner encrypt owner file with HMAC-SHA-1 Algorithm with users privilege keys with set of access attributes then apply convergent key encryption and upload into CloudSever1.

Cloudserver2: - CloudSever2 service provider. Compared with the traditional architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the CloudSever1 is not fully trusted in practice, it is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the CloudSever1. The private keys for the privileges are managed by the CloudSever2, who answers the file token requests from the users. The interface offered by the CloudSever2 allows user to submit files and queries to be securely stored and computed respectively.

V. SYSTEM ARCHITECTURE

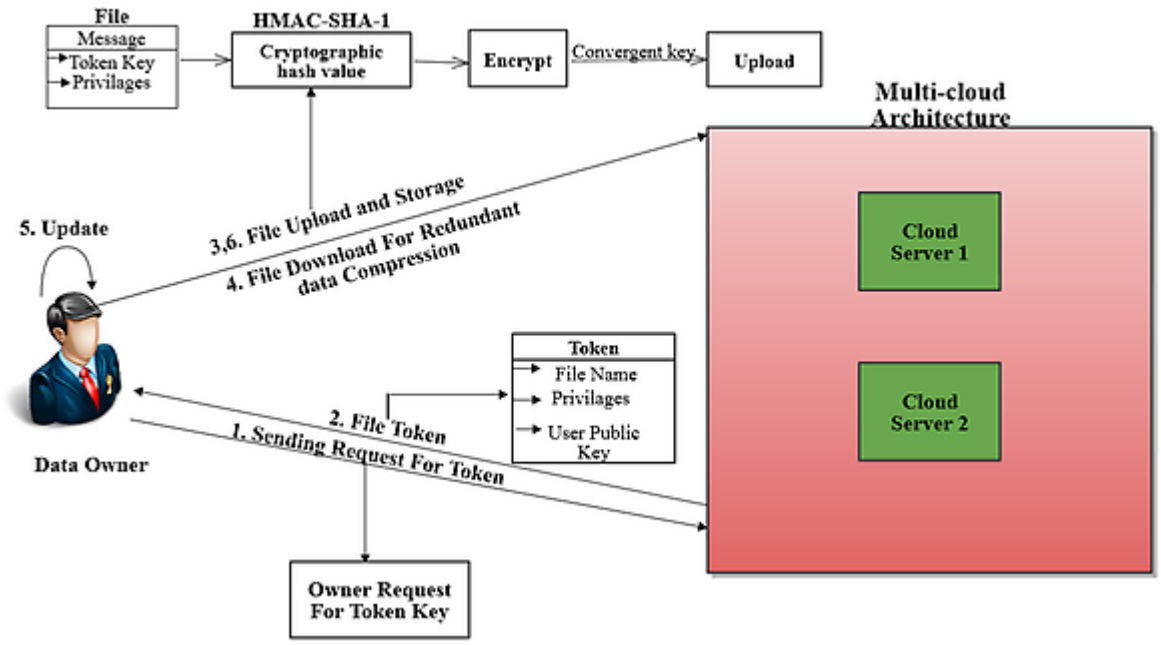

Fig.1 Architecture for Redundant Data Avoidance

In previous redundant data avoidance systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized redundant data avoidance system, each user is issued a set of privileges during system initialization. The overview of the cloud redundant data avoidance is as follow:

5.1 Post-Process Redundant Data Avoidance

With post-process redundant data avoidance, new information is initially put away on the capacity gadget and afterward a procedure at a later time will investigate the information searching for duplication. The advantage is that there is no compelling reason to sit tight for the hash computations and lookup to be finished before putting away the information in this manner guaranteeing that store execution is not corrupted. Executions offering strategy based operation can give clients the capacity to concede advancement on "dynamic" records, or to process documents taking into account sort and area. One potential disadvantage is that you might superfluously store duplicate

ISSN (Online): 2349-7084

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 8, AUGUST 2015, PP 470-474

information for a brief while which is an issue if the capacity framework is close full limit

5.2 In-Line Redundant Data Avoidance

This is the procedure where the excess information evasion hash counts are made on the objective gadget as the information enters the gadget progressively In the event that the gadget spots a piece that it effectively put away on the framework it doesn't store the new square, just references to the current piece. The advantage of in-line excess information evasion over post-process repetitive information shirking is that it requires less capacity as information is not copied. On the negative side, it is every now and again contended that in light of the fact that hash estimations and lookups takes so long, it can imply that the information ingestion can be slower subsequently diminishing the reinforcement throughput of the gadget. Nonetheless, certain sellers with in-line repetitive information evasion have exhibited gear with comparable execution to their post-process excess information evasion partners. Post-process and in-line excess information evasion strategies are regularly intensely bantered about

5.3 Encryption of Files

Here we are using the common secret key k to encrypt as well as decrypt data. This will use to convert the plain text to cipher text and again cipher text to plain text. Here we have used three basic functions,

- KeyGenSE: k is the key generation algorithm that generates κ using security parameter 1.
- EncSE (k, M): C is the symmetric encryption algorithm that takes the secret κ and message M and then outputs the ciphertext C;
- DecSE (k, C): M is the symmetric decryption algorithm that takes the secret κ and ciphertext C and then outputs the original message M.

(a) Confidential Encryption

It provides data confidentiality in redundant data avoidance. 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 derives a tag for the data copy, such that the tag will be used to detect duplicates.

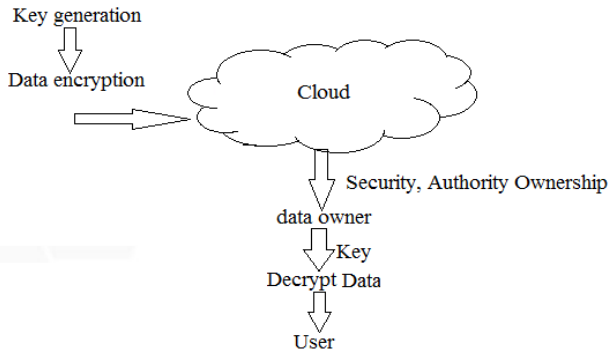

Fig 2. Confidential Encryption

(b) Proof of Data

The users have to prove that the data which he wants to upload or download is its own data. That means he/she has to provide the convergent key and verifying data to prove his ownership at server.

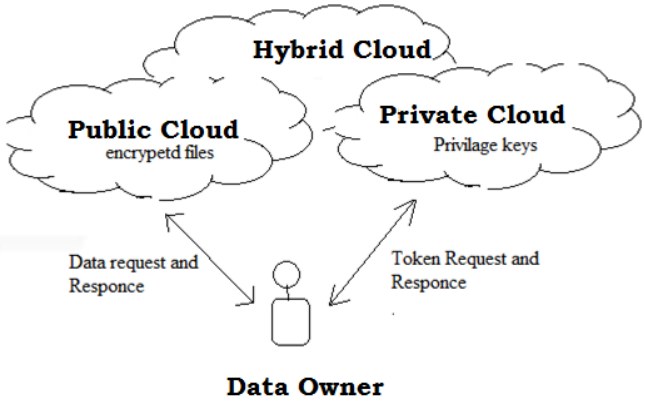

Fig 3. Proof of Data Owner

VI.CONCLUSION AND FUTURE SCOPE

Hybrid clouds offer a greater flexibility to businesses while offering choice in terms of keeping control and security. Hybrid clouds are usually deployed by the organizations willing to push part of their workloads to CloudSever1s either for cloud convulsive purposes or for project requiring faster implementation. Because hybrid clouds vary based on company needs and structure of implementation. In proposed system authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check system presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, the duplicate - check tokens of files are generated by the CloudSever2 server with private keys. Proposed system is secure in terms of insider and outsider attacks specified in the proposed security model. The proposed authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer. It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provides us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data. Hence it save the memory by deduplicating the data and thus provide us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data.

REFERENCES

[1] P. Anderson and L. Zhang. “Fast and secure laptop backups with encrypted de-duplication”. In Proc. of USENIX LISA, 2010.
[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. “Dupless: Server aided encryption for deduplicated storage”. In USENIX Security Symposium, 2013.
[3] Pasqualo Puzio, Refik Molva ,MelekOnen ,”CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage”, SecludIT and EURECOM,France.

[4] Iuon –Chang Lin, Po-ching Chien ,"Data Deduplication Scheme for Cloud Storage" International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)

[5] Shai Halevi, Danny Harnik, Benny Pinkas,"Proof of Ownership in Remote Storage System", IBM T.J.Watson Research Center, IBM Haifa Research Lab, Bar IIan University,2011.

[6] M. Shyamala Devi, V.Vimal Khanna,Naveen Balaji "Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup",IACSIT, August,2014.

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

[8] Tanupriya Chaudhari , Himanshu shrivastav, Vasudha Vashisht, "A Secure Decentralized Cloud Computing Environment over Peer to Peer",IJCSMC,April,2013

[9] Mihir Bellare, Sriram keelveedhi,Thomas Ristenart ,"DupLESS: Server Aided Encryption for Deduplicated storage" University of California, San Diego2013.

[10] Luna SA HSM. http://bit.ly/17CDPm1.

[11] Opendedup. http://opendedup.org/.

[12] Atul Adya, William J Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R Douceur, Jon Howell, Jacob R Lorch, Marvin Theimer, and Roger P Wattenhofer. Farsite: Federated, available, and reliable storage for an incompletely trusted environment. ACM SIGOPS Operating Systems Review, 36(SI):1–14, 2002.

[13] Mihir Bellare, Alexandra Boldyreva, and Adam ONeill. Deterministic and efficiently searchable encryption. In Advances in Cryptology-CRYPTO 2007, pages 535–552. Springer, 2007.

[14] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Dupless: Server-aided encryption for deduplicated storage. 2013.

[15] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Message-locked encryption and secure deduplication. In Advances in Cryptology–EUROCRYPT 2013, pages 296–312. Springer, 2013.

[16] Kevin D. Bowers, Ari Juels, and Alina Oprea. Hail: a high-availability and integrity layer for cloud storage. In Proceedings of the 16th ACM conference on Computer and communications security, CCS '09, pages 187–198, New York, NY, USA, 2009. ACM.

[17] Landon P Cox, Christopher D Murray, and Brian D Noble. Pastiche: Making backup cheap and easy. ACM SIGOPS Operating Systems Review, 36(SI):285–298, 2002.

[18] John R Douceur, Atul Adya, William J Bolosky, P Simon, and Marvin Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on, pages 617–624. IEEE, 2002.

[19] Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg. Side channels in cloud services: Deduplication in cloud storage. Security & Privacy, IEEE, 8(6):40–47, 2010.