ISSN (ONLINE): 2349-7084
GLOBAL IMPACT FACTOR 0.238
I2OR PIF 2.125

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 10, OCTOBER 2015, PP 687-693

Secure Data Deduplication and Data accessing among Multi-cloud Framework

[1]K.Naga Maha Lakshmi , [2] A.Shiva Kumar

[1]Asst Professor, Keshav Memorial Institute of Technology and Science, Narayanguda, Telangana, India,
[2]Asst Professor, Mahaveer Institute of Technology and Science, Bandlaguda, Telangana, India.

ABSTRACT— In this paper, we have reviewed on Secure Data deduplication mechanism over multi-cloud architecture, where Private Cloud is integrated as an intermediary to permit privileges' to securely perform the data duplications check with differential benefits. In our presented system data owners are just outsource their data to the public cloud for the sake of data storage and data sharing, while outsourcing the data to the public cloud it will not assure among secured data deduplication as well secured accessing due to follows traditional schemes .(i.e.) Data and privileged keys get stored in the same cloud (Public Cloud), in order to address the above issues as our proposed system proposed a novel secured convergent key mechanism to achieve the secured data deduplication among multicloud architecture .Finally our system proves that our system has secured data deduplication mechanism with improved storage space and bandwidth .

KEYWORDS—Public Cloud, Convergent Key Encryption, File level Check, Block Level Check, Convergent key.

****

I. INTRODUTION

Cloud computing is a model for delivering information technology services in which resources are retrieved from the internet through web-based interface and application, instead of direct connection to a server. Cloud storage provides a service for the evergreen management of vast amount of data in order to reduce the space and bandwidth. To make

Reliable and scalable management of data in the cloud computing, deduplication plays a vital role as a conventional technique. De-duplication is a data compression technique which is most commonly used for eliminating repeated copies of data/files in cloud storage to reduce space and bandwidth. This technique is used for reliable storage utilization and to provide scalable network data transfers to reduce number of bytes that must be sent. Data deduplication may occur as file level as well as block level data de-duplication. Keeping multiple duplicate copies of file/data with similar content, de-duplication detects and eliminates the redundant data by keeping original physical copy. The system recovers the storage consumption and it can be applicable to network data transfer to reduce the number of bytes that must be sent. Data de-duplication stretches lot of reimbursements, refuge and confidentiality anxieties ascend as the users' subtle data is liable to both inside and outside spasms. Profitable cloud storage services such as Dropbox, Mozy and Memopal, have been applying de-duplication for user data to bar preservation cost. Data outsourcing raises security and privacy concerns [3]. Deduplication improves storage and bandwidth competence and is attuned with convergent key management. Traditional encryption requires dissimilar users to encrypt their data with their own key. To stop unauthorized access; a secure proof of possession protocol is additionally required to provide the proof that the user indeed owns the similar file when a duplicate is found. Once the proof of consequent users with the

Similar file are going to be provided a pointer from the server while not having to transfer the similar file.

II .RELATED WORK

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

2.1 DupLESS Server Aided Encryption for Deduplicated Storage

DupLess: Server aided encryption for deduplicated storage for cloud storage service provider like Mozy ,Dropbox, and others perform deduplication to save space

by only storing one copy of each file uploaded .Message lock encryption is used to resolve the problem of clients encrypt their file however the saving are lock. Dupless is used to provide secure deduplicated storage as well as storage resisting brute -force tacks. Clients encrypt under message -based keys obtained from a key-server via an oblivious PRF protocol in dupless server. It allow clients to store encrypted data with an existing service, have the service occurs deduplication on their on the part , and yet achieves strong confidentiality guarantees. It show that encryption for deduplicated storage can successfully reach desired performance and space savings close to that of using the storage service with plaintext data [2].

Characteristic:

1. More Security.
2. Easily -deployed solution for encryption that supports deduplication
3. User Friendly: Use command-line client that supports both Dropbox and Google Drive.
4. Resolve the problem of message lock Encryption.

2.2 Proofs of Ownership in Remote Storage Systems

It stores only the single copy of the duplicate data. Client-side deduplication tries to identify deduplication chance already at the client and save the bandwidth of uploading copies of existing files to the server[11].To overcome the attacks Shai Halevi1, Danny Harnik, Benny Pinkas, and Alexandra Shulman -Peleg proposes the Proof of ownership which lets a client efficiently prove to a server that that the client keep a file, rather than just some short information about it present solutions based on Merkle trees and specific encodings, and analyses their security.[9]

Characteristic:

1. To identity the attacks that exploit client -side deduplication..
2. Proofs of ownership provide the rigorous security.
3. Rigorous efficiency requirements of Peta-byte scale storage systems

.

2.3 A Secure Deduplication with Efficient and Reliable Convergent Key Management

Data deduplication is a used for removing duplicate copies of data, and has been widely applied in cloud storage to reduce not only storage space but also upload band width. Promising as it is, an appearing challenge is to accomplish secure deduplication in cloud storage. Although convergent encrypt ion has been extensively acquired for secure deduplication, a uncertain issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys.

Techniques:

1. Key management
2. Convergent Encryption [4]

2.4 Twin Clouds: An Architecture for Secure Cloud Computing

S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider proposed architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. Income towards, the user communicates with a trusted cloud. Which encrypts as well as verifies the data stored and operations occurred in the untrusted cloud .It divide the computations such that the rusted cloud is used for security -critical operations in the less time-critical setup phase, whereas queries to the outsourced data are processed in parallel by the fast cloud on encrypted data [10]

Security and Privacy issues in the cloud:

Only the authorized persons need to access the data from the cloud. In order to ensure the integrity of user authentication, the 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 too 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 reduplicated 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 deduplicating 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

Convergent encryption [5], 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].

IV.SYSTEM STUDY

4.1. Presented System:

In our presented 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 data deduplication with secured manner is the foremost objective of the proposed system, in this connection we proposed secure data deduplication mechanism by distinguish sensitive and non-sensitive data at data uploading into cloud level and apply the crypto algorithm for sensitive data by applying this data get secured and authorized

ISSN (ONLINE): 2349-7084
INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 10, OCTOBER 2015, PP 687-693

System Architecture

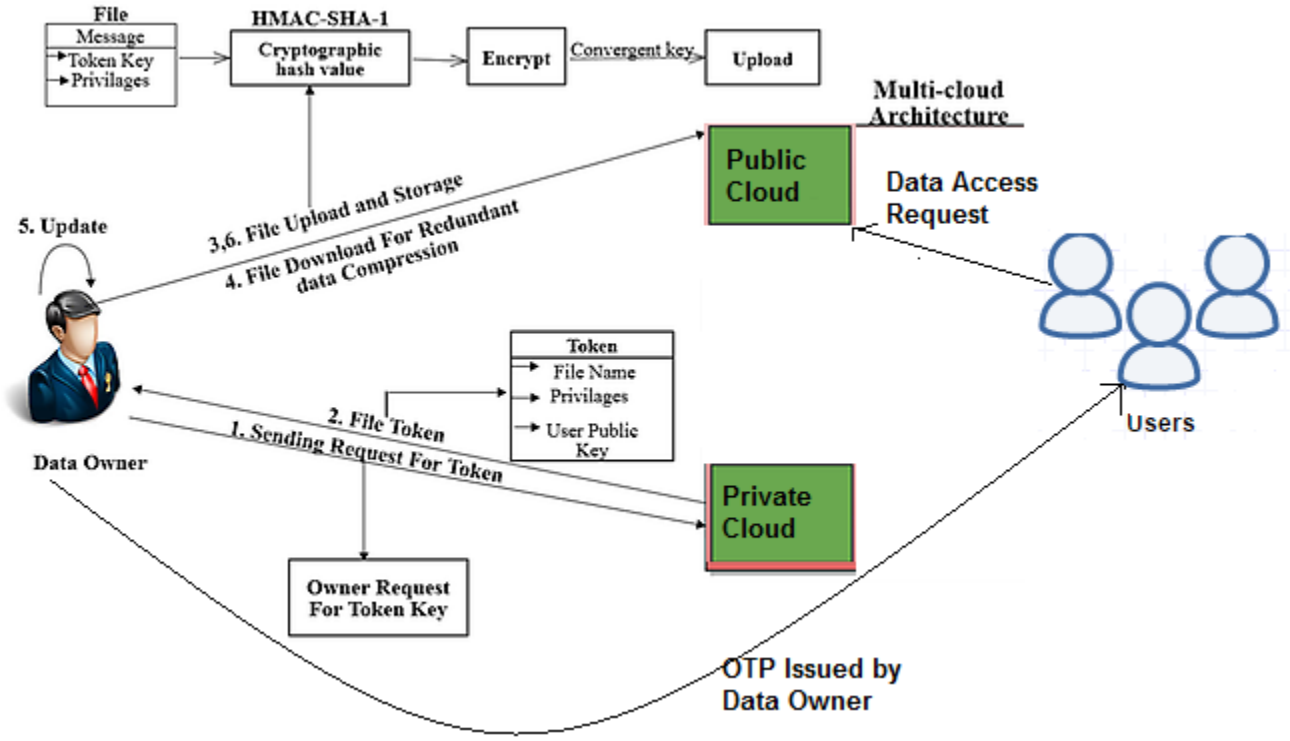

Fig 1. Proposed System Architecture

V. SYSTEM IMPLEMENTATIONS

User: User must be registered for to upload data into clouds by providing required info... like name, password, and email, mobile.

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 by encrypting the data in the files and giving privileges to other data users according to user requests and that given privileges information will be send to users registered e-mail.

Data deduplication with secured manner: while data uploading by user into public cloud , the identification of duplicate data will be notified by showing the warning pop message to users if the user wan to upload existing file again ,still user wan to upload file the new file need to update with existing file. while user uploading data into public cloud user can distinguish sensitive and non-sensitive data and can provide encryption for only sensitive data .If any unauthorized user wan to access or the user didn't have particular privileges (like read write, if user is having read privileges but they want to access file (downloading like that)) immediately message alert need to send to for a particular data owner

5.1. 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.

ISSN (ONLINE): 2349-7084
INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 10, OCTOBER 2015, PP 687-693

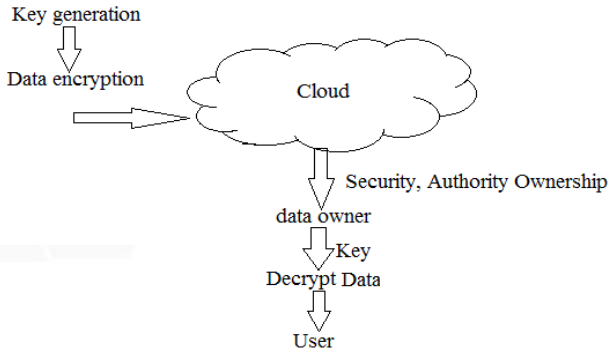

Fig 2. Confidential Encryption

(b) Proof of Data

The notion of proof of ownership (PoW) [11] 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.

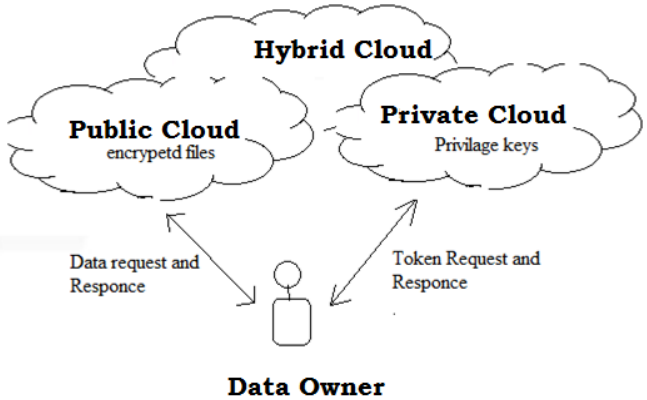

Fig 3. Proof of Data Owner

(c)Identification Protocol

The identification of protocol having two phases as follows:

1. Proof: The user can demonstrate his identity to a verifier by performing some identification proof related to his identity.

2. Verify: The verifier occurs verification with input of public information.

5.2. Data De-Duplication Types

A. File-level de-duplication

It is commonly known as single -instance storage, file-level data de-duplication compares a file that has to be archived or backup that has already been stored by checking all its attributes against the index. The index is updated and stored only if the file is unique, if not than only a pointer to the existing file that is stored References. Only the single instance of file is saved in the result and relevant copies are replaced by "stub" which points to the original file.

B.Block-level de-duplication

Block-level data de-duplication operates on the basis of sub-file level. As the name implies, that the file is being broken into segments blocks or chunks that will be examined for previously stored information vs redundancy. The popular approach to determine redundant data is by assigning identifier to chunk of data, by using hash algorithm for example –it generates a unique ID to that particular block. The particular unique Id will be compared with the central index. In case the ID is already present, then it represents that before only the data is processed and stored before .Therefore only a pointer reference is saved to the previously stored data. If the ID is new and does not exist, then that block is unique. The unique chunk is stored and the unique ID is updated in the Index.

The size of the chunk which needs to be checked varies from vendor to vendor. Some will have fixed block sizes, while some others use variable block sizes likewise few may also change the size of fixed block size for sake of confusing. Block sizes of fixed size may vary from 8KB to 64KB but the main difference with it is the smaller the chunk, than it will be likely to have opportunity to identify it as the duplicate data. If less data is stored than it obvious ly means greater reductions in the data that is stored. The only major issue by using fixed size blocks is that in case if the file is modified and the de-duplication result uses the same previously inspected result than there will be chance of not identifying the same redundant data segment, as the blocks in the file would be moved or changed, than they will shift downstream from change, by offsetting the rest of comparisons.

5.3. Restrict From Unauthorized Access

When user want to access the data from public cloud, that user need to authorized by the data owner by having privilege keys taken from data owner through any one of the secured communication system i.e. E-mail, unless and until get the access key from the data owner. Authorization system does not allow any access rights to words protecting access from unauthorized users.

Algorithm Used

Here in this section in order to provide secure data accessing from public cloud, while uploading the data into public cloud by the data owner, data need to be encrypted using secure cryptographic algorithm i.e. (HMAC-OTP)

It's a symmetric cryptographic algorithm, which performs secured data encryption and decryption by using same key, which leads easy key management along with high performance. In this concern encrypted data will be protected from cloud provider as well adversaries.

HMAC-BASED ONE-TIME PASSWORD ALGORITHM

K be a secret key, C be a counter
HMAC(K,C) = SHA1(K ⊕ 0x5c5c… ‖ SHA1(K ⊕ 0x3636… ‖ C)) be an HMAC calculated with the SHA-1 cryptographic hash algorithm Truncate be a function that selects 4 bytes from the result of the HMAC in a defined manner Then HOTP(K,C) is mathematically defined by HOTP(K,C) = Truncate(HMAC(K,C)) & 0x7FFFFFFF The mask 0x7FFFFFFF sets the result's most significant bit to zero. This avoids problems if the result is interpreted as a signed number as some processors do.[1]For HOTP to be useful for an individual to input to a system, the result must be converted into a HOTP value, a 6–8 digits number that is implementation dependent.
HOTP-Value = HOTP(K,C) mod 10d, where d is the desired number of digits HOTP can be used to authenticate a user in a system via an authentication server. Also, if some more steps are carried out (the server calculates subsequent OTP value and sends/displays it to the user who checks it against subsequent OTP value calculated by his token), the user can also authenticate the validation server

VI.CONCLUSION

Data Deduplication eradicates the redundant data by storing only the single copies of data. It uses the convergent encryption technique to encrypt the data with Authorized duplicate check, so that only authorized user with specified privileges can perform the duplicate check. The concept de-duplications save the bandwidth and reduce the storage space. It also eradicates the duplicates of data in the cloud storage. It does not allow the unauthorized user to steal information. Thus it provides lots of benefits based on the confidentiality, authorized duplicate check also the cloud storage space as well as the healing information is prevented.

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.