ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

STORAGE MANGEMENT IN HYBRID CLOUDUSING DEDUPLICATION TECHNIQUE

[1] KALA.M, [2]ARIVU SELVI.P, [3]RAMYA.R, [4]SASIKALA.M M.E.,
[1,2,3]M.E. First year, Department of Computer Science and Engineering
srikalai6632@gmail.com,
[4]Assistant Professor, Department of Computer Science and Engineering
sasibtech91@gmail.com

BHARATHIYAR INSTITUTE OF ENGINEERING FOR WOMEN, DEVIYAKURICHI

Abstract—A special form of data compression technique is deduplication that eliminates the repeated copies of same data. Deduplication is very important to cloud storage even increase in the number of users and size of data. Deduplication is used in the cloud servers to reduce the volume of storage space and save bandwidth. For illustrate, the same file may be uploaded in several different places by different users. Deduplication eliminates the repeating copies by saving one copy of the original data and replacing the other copies with link that points to original copy. The convergent encryption technique is used to encrypt the user data before uploading, to provide the security of the data. The certified duplicate check is applied with the encryption technique to provide the secure data in hybrid cloud storage.

I. INTRODUCTION

Cloud computing is computing technique that relies on sharing*computing resources* rather than having local servers or personal devices to handle applications.Cloud computing undertakings several attractive benefits for businesses.Clouds can be classified as public, private, hybrid. Private cloud services are delivered from a business' data center to internal users. Public cloud services are sold on-demand, typically by the minute or the hour. Primary public cloud providers include Amazon Web Services (AWS), Microsoft Azure, IBM/Soft Layer and Google Compute Engine.The cloud computing is a newly evolved computing terminology or metaphor based on utility and consumption of resources. One energetic challenge of cloud storage services is the management of the always-increasing volume of data. In cloud computing, deduplication [1] has been a famous technique to conserve the cloud storage. Data deduplication is a specific form of data compression technique for eliminating duplicate copies of recapping data in storage. The deduplication technique is used to expand storage utilization and can also be applied to network data transfers to reduce the quantity of bytes that must be sent. Instead of keeping data copies with the same data, deduplication removes redundant data by keeping only one physical fake and provide pointer referring

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

other redundant data to that copy. The deduplication take performed in either block level or file level.In file level approach duplicate files are cancelled,and in block level approach duplicate blocks of data that occur in non-identical files.The deduplication has a important concerns the security to protect the data from insider or outsider attack.For data secrecy,encryption is used by dissimilar user to encrypt their files or data, usinga secrete key user perform encryption and decryption operation.The encryption operation is consequent from the data content,duplicatecopies generate the same convergent keys.To avoidunauthorized access proof of ownership protocol is used to offer proof that the user definitely owns the similar file when deduplication is found.After the proof, server provides a link to subsequent user for retrieving same file without needing to upload similar file.When user want to download file,they simply download encrypted file from cloud and decrypt this file using convergent keys.

II. RELATED WORK

Secure deduplicaton

Secure data deduplication [2] has attractedprofuseawareness recently from research community. Deduplication technique inthe cloud storage to reduce the storage size of thetags forintegrity check. To improve the security ofdeduplication andsafeguardthe data confidentiality.To safeguard the dataprivacy, by transforming the predictablemessageinto accidental message. In their system, additional third party called key server is introduced to generate the file tag for duplicate check. The encryption scheme [3] that provides differential security for popular data and unpopular data.Another two layered encryption scheme with tougher security while supporting deduplication is proposedfor unpopular data. In the way, theyachieved improved tradeoff betweenthe efficiency and security of the outsourced data. Addressed the keymanagement problemin block-level deduplication by distributing these keys across several servers after encrypting the files

Convergnt encryption

Convergent encryption ensure data privacy in deduplication. The message-locked encryption [4] and its application are discovered in space-efficient secure outsourced storage. A secure convergent encryption for efficient encryption, without considering problems of the key-management [5] and block-level deduplication. There are several executionsof different convergent encryption variants for safe deduplication. It is known that some profitable cloud storage providers and also organize convergent encryption

Proof of Ownership

The concept of “proofs of ownership” (PoW) [6] for deduplication systems, such that a client can professionallyconfirm to the cloud storage server that he/she owns a file without uploading the file itself. NumerousPoW constructions based on the Merkle-Hash Tree are planned to enable client-side deduplication, which include the enclosedoutflow setting. Another efficient PoW scheme by choosing the projection of a file onto some casually selected bit-positions as the file proof. Recently, PoW extended for encrypted files, but they do not address how to decrease the key management overhead.

Twin cloud

An architecture consisting of twin clouds [8] forsecure outsourcing of data and arbitrary computations to untrusted commodity.The hybrid cloud [9]techniquesisto support the privacy-aware data-intensive computing. The authorized deduplication difficult over data in public cloud.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

The security modelof our systems isparallel to those related work, where the private cloud is assume to be straightforward but interested.

III. SYSTEM MODEL

A. Symmetric encryption

Symmetric encryption offers a general secret key κ to encrypt and decrypt data. A symmetric encryption scheme suggests three primitive functions:

KeyGenSE(M)-κ is the key generation algorithmthat generates κ using security parameter K;

EncSE(κ, M)-C is the symmetric encryption algorithm that takes the secret κ and message *M* , then outputs the cipher text*C*; and*DecSE (κ, C)-M* is the symmetric decryption algorithmthat takes the secret κ and cipher text*C*, then outputs the unique message *M*.

B. Convergent encryption

Christo Ananth et al. [7] discussed about Reconstruction of Objects with VSN. By this object reconstruction with feature distribution scheme, efficient processing has to be done on the images received from nodes to reconstruct the image and respond to user query. Object matching methods form the foundation of many state- of-the-art algorithms. Therefore, this feature distribution scheme can be directly applied to several state-of-the-art matching methods with little or no adaptation. The future challenge lies in mapping state-of-the-art matching and reconstruction methods to such a distributed framework. The reconstructed scenes can be converted into a video file format to be displayed as a video, when the user submits the query. This work can be brought into real time by implementing the code on the server side/mobile phone and communicate with several nodes to collect images/objects. This work can be tested in real time with user query results.

Convergent encryption scheme can be separate with four primitive functions:

KeyGenCE (M) -K is the key generation algorithmthat maps a data copy *M* to a convergent key *K*;

EncCE (K,M)-C is the symmetric encryptionalgorithm that takes both the convergent key *K* and the data *M* as inputs and then outputs a cipher text*C*;

DecCE(K,C)-M is the decryption algorithm that takes together the cipher text*C*and the convergent key *K* as inputs and then outputs the original data copy *M*; and

TagGen (M)-T(M) is the tag generation algorithmthat maps the unique data copy *M* and outputs atag*T*(*M*).

IV. PROPOSED SYSTEM

The hybrid cloud is a combination of the both public cloud and private cloud.The private keys for privileges willnot be send to users directly, which will be kept and managed by the private cloud server. The users cannot distribute these private keys of rights, which means that it can prevent the privilege key sharing among users.The user wants to send a request to the private cloud server to get a file token. To achieve the duplicate check for some file, the user wants to get the filetoken from the private cloud server. The private cloud server will verify the user's identity before issue the related file token to the user. The authorized duplicate check for this file can be performed by the user with public cloud before uploading the file. The user either uploads the file or runs PoW based on the outcome of duplicate check. If we used the public cloud the user can't provide the security for our private data and the private data will be loss. User can upload and download the files from public cloud but private cloud provides the protection for that data. Only the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

authorized person can upload and download the files from the public cloud. For that, user creates the key and stored that key onto the private cloud. At the time of downloading user request to the private cloud for key and then the user gets right to use that particular file.

V. AUTHORISATION ANALYSIS

The security will be determined in terms of two aspects, one is the authorization of duplicate check and the another is privacy of data. Some tools have been used to construct the securededuplication, which provides secure to data. Such basic tools includes the convergent encryption, symmetricencryption , and the PoWscheme.

A. Security of Duplicate-Check Token

For protection, unforgeability of duplicate-check token and indistinguishability of duplicate check token. Inunforgeability of duplicate-check token there are two types of adversaries, that is, external adversary and internaladversary.The security of indistinguishability of token can be proved based on the assumption of the underlying message authentication code is secure. The security of message verification code requires that the adversary cannot distinguish if a code is produced from an unknown key. In our deduplication system, all the privilege keys are kept secret by the private cloud server.

B. Confidentiality of Data

The data will be encrypted in our deduplication system before uploading to the S-CSP. Two kinds of different encryption methods have been applied in our two constructions. Some new security notations of privacy against chosen-distribution attack have been defined for unpredictable message. Theprotection analysis for external adversaries and internal adversaries is nearly identical, except the internal adversaries are provided with some convergent encryption keys.The convergent encryption keys have no privacy impact on the data confidentiality because these convergent encryption keys are computed with different rights. The symmetric key k is randomly chosen, it is encrypted by convergent encryption key $k_{F,p}$. The confidentiality of file will be minimised to convergent encryption because the encryption key is deterministic.

VI. IMPLEMENTATION

A *Client* program is used for model the data users to carry out the file upload process. A *Private Server* program is used to ideal the private cloud which manages the secret keys and handles the file token computation. A *Storage Server* program is usedto model the S-CSP which storesand deduplicates files.

Theexecutionof the **Client** provides the followingfunction calls to support token generation and deduplication along the file upload process.

• FileTag (File) - It evaluates SHA-1 hash of theFile as File Tag;

•ShareTokenReq(Tag, {Priv.}) - It needs the Private Serverto generate the Share File Token with the File Tag and Target Sharing Privilege Set;

• File Encrypt(File) - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode, where the convergent key is from SHA-256 Hashing of the file;

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

• FileUploadReq (FileID, File, Token) – It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.

The execution of **Private Server** includes corresponding request handlers for the token generation and maintains a key storage with Hash Map.

• TokenGen (Tag, UserID) - It loads the associated privilege keys of the user and produce the token with HMAC-SHA-1 algorithm;

• ShareTokenGen (Tag, {Priv.}) - It producesthe share token with the related privilege keys of the sharing right set with HMAC-SHA-1 algorithm.

VII. RESULT ANALYSIS

The result estimation is focuses on comparing the overhead induced by authorization steps including file token generation and share token generation, against the convergent encryption and file upload steps. The upload process divided into 6 steps, 1) Tagging 2) Token Generation 3) Duplicate Check 4) Share Token Generation 5) Encryption 6) Transfer. For each step, we record the start and end time of it and finally obtain the breakdown of the total time spent.

A. File size

To estimate the effect of file size to the time spent on different steps, we upload 100 unique files (i.e., without any deduplication opportunity) of particular file size and record the time break down.The exclusivefiles enables us to choose the worst-case scenario where we have to upload all file data.The time spent on tagging, encryption, upload increases linearly with the file size, since these operations involve the definite file data and incur file. In contrast, other steps such as token generation and duplicate check only use the file metadata for calculation and therefore the time spent remains constant.

B. Number of Stored Files

To estimate the effect of number of stored files in the system,The user upload 10000- 10MB distinctive files to the system and record the breakdown for every file upload. Token check is done with a hash table and a linear search would be carried out in case of collision. In spite of the possibility of a linear search, the time taken in duplicate check residueconstant due to the low collisionprobability.

C. Deduplication ratio

The result of the deduplication ratio is evaluated, we prepare two unique data sets, each of which consists of 50- 100MB files. The user upload the first set as an initial upload. For the second upload, we pick a portion of 50 files, based on the given deduplication ratio from the first set as duplicate files and remaining files from the second set as unique files. As uploading and encryption would be skipped in case of duplicate files, the time spent on both of them decreases with increasing deduplication ratio. Total time spent on uploading the file with deduplication ratio at 100% is only 33.5% with unique files.

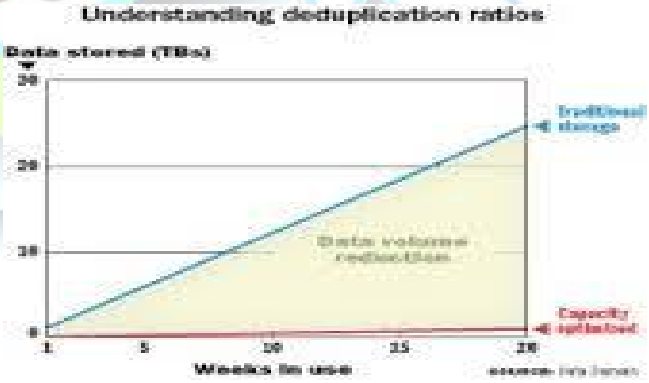

Figure 1.Deduplication Ratio.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016

VIII.CONCLUSION

The authorized data deduplication was implemented to protect the data security by including differential authority of users in the duplicate check. Our data are securely stored as encrypted files in public cloud, and key is store in private cloud with respective file. Without key no one can access our file or data from public cloudstorage.The deduplication constructions sustaining authorized duplicate check in hybrid cloud architecture, in that the duplicate-check tokens of files are created by the server in private cloud with private keys. Security analysis shows that our schemes are secure in terms of insider and outsider attacks particular in the expected security model. For that, we implemented a prototype of our proposed approved duplicate check scheme and conduct test bed experiments on our prototype. Our approved duplicate check scheme incurs nominaloverhead whencompared to convergent encryption and network transfer.

REFERENCES

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

[2] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical *Report*, 2013.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart.Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

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

[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEETransactions on Parallel and Distributed Systems*, 2013.

[6] ShaiHalevi, 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.

[7] Christo Ananth, M.Priscilla, B.Nandhini, S.Manju, S.Shafiqa Shalaysha, "Reconstruction of Objects with VSN", International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Vol. 1, Issue 1, April 2015, pp:17-20

[8] JiaXu, Ee-Chien Chang, and JianyingZhou.Weak leakage-resilientclient-side deduplication of encrypted data in cloud storage. InProceedings of the 8th ACM SIGSAC symposium on Information,computer and communications security, pages 195–206. ACM, 2013.

[9] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg.Proofsofownership in remote storage systems. In Y. Chen, G. Danezis, And V. Shmatikov, editors, *ACM Conference on Computer andCommunications Security*, pages 491–500. ACM, 2011.

[10] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication Protocols in cloud storage. In S.Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 2, March 2016