IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Study of Authorized Data Deduplication Methodology to Achieve Security in Hybrid Cloud

Priyanka Bhopale[1], A.M. Kanthe[2]

Department of Computer Science, Sinhgad Institute of Technology, Pune, India[1, 2]

Abstract: Number of files created per day is increased due to rapid rise in number of users. In cloud computing environment most of the communication is done through the file processing, and hence it becomes very crucial and significant to provide efficient approach for data security. In this research we are concentrating on data deduplication to provide efficient security service in cloud computing. Data deduplication is nothing but data compression technique which is used to remove the duplicate copies of echoing data. This methodology is regularly used for dropping the storage space and save bandwidth under cloud. Also, along with deduplication for data protection and privacy the encryption methods are used. In this paper we are studying the authorized data deduplication to provide the data security by using differential privileges of users in the cloud architecture. Different new deduplication constructions offered for sustaining authorized duplicate check. This paper shows how the security is obtained in hybrid cloud during the process of data deduplication.

Keywords: Data Deduplication, Data Security, Privilege authorization, Hybrid Cloud.

I. INTRODUCTION

Cloud computing is one of today's supreme appealing skill areas due to its cost-efficiency and flexibility. Storing data on CSP (Cloud Server Provider) [1] instead of at local system enables the organization to store large amount of data. This issue of saving the data at server level gives the Organization more flexibility and a quick access to the reconstructed data. As all data is managed at only one place, makes the users to use it sitting at any corner in the world. The main problem of cloud is the amount of data which is growing gradually. Everyday new users are added and the users wanted to accumulate their files on the cloud. So daily huge numbers of files get added to the storage which makes tough for the organization. One key security challenge is to show the property of assured deletion. Due to cloud storage this issue is resolved and also the security algorithm manages the authentication of users, which leads to privacy to the information.

Generally when cloud encrypts the information they use their own encryption key. With convergent encryption, the encryption key is obtained from the file itself. So, it produces identical cipher text from identical plaintext. Convergent encryption permits cloud storage to deduplicate data, without the service having access to the encryption keys used for protection of user files. It offers better privacy than traditional cloud storage.

The new technique which arises is data deduplication; it is an intelligent data compression technique, in which only one copy of a file is stored at the server level, no matter how many users want to store it. Only single instance of the file is available on the cloud storage and so reduces the storage wastage. Data deduplication is a usually used for eliminating duplicate copies of data, and it reduces storage space and upload bandwidth.

Data deduplication can operate at two level 1.The file level 2. The block level. In file-level deduplication, if more than one file is exactly alike, one copy of the file is stored and other repetitions receive pointers to the saved file. In block deduplication it looks within a file and saves unique copy of respectively block. If a file is updated, only the changed data is saved. Eliminating redundant data can knowingly shrink storage requirements and improve bandwidth competence. Since primary storage has gotten inexpensive over time, enterprises usually store many versions of the same information so that new personnel can reuse previously done work. Storage cost is reduced because less numbers of disks are required for storage.

The data which is stored on the cloud shared among the users under some predefined privileges, which outlines the access rights to the stored data. While doing the registration process, privileges are assigned to each user which provide the security to data.

This paper is organised as follows. In section II we have given Literature survey. Section III discusses authorized data deduplicate system. Section IV gives a detail about mathematical model and section V gives the conclusion followed by references.

II. LITERATURE SURVEY

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou [1] proposed authorized data deduplication to protect the data safety by counting differential privileges of users in the duplicate check. For data deduplication different techniques are used in hybrid cloud architecture which checks duplicate files. Tokens of files are produced by the private cloud server with private keys. Test experiment shows that authorized duplicate check

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

incurs minimal overhead equated to convergent encryption and network transfer.

Jin Li, Xiaofeng Chen, M. Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou [2] proposed a system for Data deduplication. Though convergent encryption has been broadly adopted secure deduplication, an important issue of making convergent encryption practical is to proficiently and reliably handle a vast number of convergent keys. This paper makes the first try to address the issue of achieving efficient and reliable key management in secure deduplication. They first bring in a baseline style in which each user holds an independent master key to encrypt the convergent keys after outsourcing them to the cloud. However, managing such a baseline key management system generates an enormous number of keys with the growing amount of users and requires users to dedicatedly protect the master keys. To this end, They propose Dekey , a fresh structure in which users do not require to manage any keys on their own but instead securely allocate the convergent key shares transversely multiple servers. Security analysis shows that Dekey is secure in terms of the definitions precisely given in the proposed security system. As a proof of concept, we implement Dekey using the Ramp secret allotment scheme and shows that Dekey incurs limited overhead in realistic environments.

Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang, Yang Xiang, and Mohammad Mehedi Hassan [3] This paper targets at security trials; creates the first attempt to validate the idea of distributed reliable deduplication system, authors propose novel distributed deduplication systems with developed reliability in which the blocks of data are distributed across multiple cloud servers. The security necessities of data confidentiality and tag reliability are also achieved by announcing a deterministic secret sharing method in distributed storage systems, instead with convergent encryption as in earlier deduplication systems. Security analysis demonstrates that these deduplication models are secure in terms of the definitions specified in the planned security model. As a proof of concept, they execute the planned systems and prove that the incurred overhead is very limited in realistic environments.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg [4] they proposed deduplicate is one of critical information packing plans to wipe out copy duplicates of rehashing information, and has been mostly utilized as a part of Cloud storage to reduce the measure of storage room and spare data transfer ability. To secure the secrecy of gentle information while supporting de duplication, the blend of encryption system has been planned to encode the information before outsourcing. To superior secure information security, this paper makes the effort to formally address the issue of accepted information de duplication. Not the identical as normal de duplication frameworks, the differential profits of clients are further taken in consideration in copy check other than the information itself. They additionally show a few new de duplication developments supporting standard copy weigh in a half and half cloud building design.

Chun-Ho Ng and Patrick P. C. Lee [5] offer RevDedup, a deduplication system that improves reads to latest VM image backups using a plan called reverse deduplication. In difference with conventional deduplication that eliminates duplicates from new data, RevDedup removes duplicates from aged data, thereby shifting fragmentation to old data while keeping the arrangement of new data as sequential as possible.

III. AUTHORIZED DATA DEDUPLICATION

In this work of authorized data de duplication offered to preserve the data security by including differential rights of users in the duplicate check. Similarly we have considered several new deduplication structures which supports authorized duplicate check in hybrid cloud design, in which the identical-check tokens of files are invented by the private cloud server with private keys. In this architecture the private keys are not directly given to the user instead it would be managed by the private cloud server. One more important aspect of deduplicate process is token generation [1]. Using this token which is generated after every request hash values of the files are matched. If the file is previously present in the server then only reference to the file is created for the new user. Which means that this file is now belongs to that user also and the server keeps only one instance of the file for all those users holding same file. At the time of download the token is matched with the previously stored token to restrict the access of the unauthorized user.

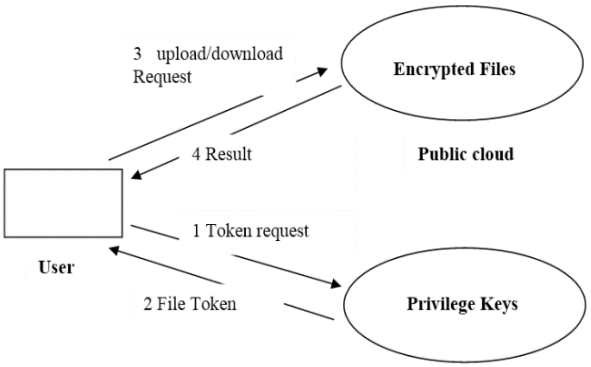

Fig I Authorized Data Deduplication

A. Design Goals

Following constraints are considered in the construction of the system by different researchers.

i. Differential authorization

Each official user is able to get his/her individual token of his/her file to perform duplicate check based on his/her privileges. Under this supposition, any user cannot generate a token for duplicate check which is not of his privileges or without the help from the private cloud server.

ii. Authorized duplicate check

Authorized user is capable to use his/her individual private keys to produce query for certain file and the privileges he/she maintained with the help of private cloud, whereas the public cloud performs duplicate check directly and

DOI 10.17148/IJARCCE.2016.5102

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

states the user if there is any duplicate. Convergent encryption [2] technique is used in order to receive the deduplication. Same file for any user can have the same key and so we can say that same data is present in the file. File is encrypted with the derived convergent key. Hence same file have the same key and hence the encrypted file is also same. Private cloud is responsible for managing all the convergent keys for all file.

iii. Data confidentiality
Unauthorized users without suitable privileges or files, which includes the S-CSP and the private cloud server, should be prohibited from access to the primary plaintext stored at S-CSP [1]. We can say that, the goal of the adversary is to regain and recover the files that do not belong to them. In this structure, compared to the previous definition of data confidentiality which based on convergent encryption, a advanced level confidentiality is defined and achieved.

B. Operations performed

i. To upload a file
User needs to register first. After registration the user can login to the system with the help of user id and password. As the user chooses the file to be uploaded on the server hash values of the files will be generated, also the hash values of the files already present on the storage are also generated then it finds a match. If a match is found that means file is already present on the server. So a reference is added to the old file instead of saving the new file. If match will not generate then a file is stored as a new file.

ii. To download a file
When any registered user wants to download the file then in that case a token will be generated which will check the privilege and reference is present or not. If match is found then user is able to download the file else server will reject the request.

C. Algorithm of System
Let File 1 is the file for which we want to square deduplicate and want to upload or download and is compared with File 2 at run time.

i. Algorithm to upload or Download
Step 1: Start (send request for login)
Step 2: Select the appropriate option
If upload go to step 3
If download go to step 4
Step 3: add the details for upload
If values of file 1 and file 2 are different
Then hash of the new file h' does not match with the hash h then
Then save file 1 and file 2 and create different references for both
r1=1 and r2=1
and also save the created convergent keys K.
If file 1 and file 2 are same
Then hash of the new file h' match with the hash h then
Then save only file 1 and
Add a reference of file 1 to the existing reference r of file 2
r2 = r2 + 1
In this case convergent key K will be same
Step 4: check if requested file is in the list
Then generate a token
Match the token
If yes then file is downloaded
If no transaction not complete.

D. Flow chart of the system

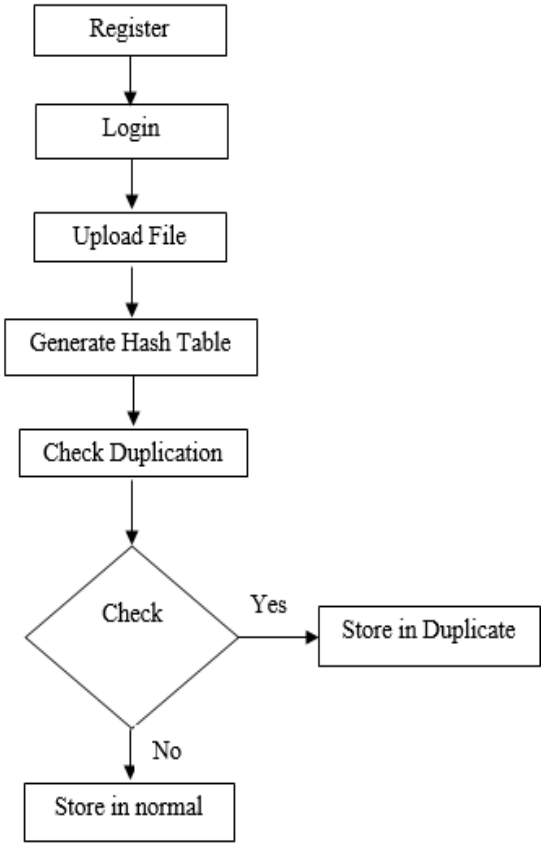

Fig II Flowchart of system

IV. MATHEMATICAL MODEL

Input = {d}
Output = {enc}
Where,
S: System which represented flow of project
Sign: - Signature for each file
d = Data file
enc = Encrypted File

FileTag (File) – here it finds the file tag, using SHA 1 hash method.

TokenReq (Tag, UserID) – Token is created by requesting to the Private Server, using the file tag which is obtained in previous step and user ID;

DupCheckReq (Token) – receives the token from Private server and requests the Storage Server check Duplicate of the File using the file token;

ShareTokenReq (Tag, {Priv.}) - It sends a requests to the Private Server for generating the Share Token by using the File Tag and Privilege Set;

FileEncrypt (File) - This encrypts the File with Convergent Encryption method which uses AES algorithm for 256-bit

FileUploadReq (FileID, File, Token) – It uploads the File Data to the Storage Server if the file is not duplicate and updates the File Token stored.

TokenGen(Tag, UserID) - It puts the respective tag of file and user ID to generate the token using HMAC-SHA-1 algorithm;

V. CONCLUSION

Security problems can be decrease by adding different constraints such as differential authorization, token generation and data confidentiality. Data deduplication is achieved by convergent encryption. The system gives us high performance if the time required for the upload and download will be minimum also the time taken by the system for deduplication check will get lower. Further we have studied several new deduplication constructions supporting permitted duplicate sign on hybrid cloud style, at intervals that the duplicate-check tokens of files unit generated by the non-public cloud server with personal keys. Security study and mathematical model demonstrates that our Schemes Square live secure in terms of executive and outsider attacks set enter the planned security model.

ACKNOWLEDGMENT

We would like to thank the editor and anonymous reviewers for their valuable suggestions that significantly improved the quality of this paper.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, Hybrid Cloud Approach for Secure Authorized Deduplication, Parallel and Distributed Systems, IEEE Transactions on Volume: 26, Issue: 5, pages 1206-1216, 2014

[2] Jin Li, Xiaofeng Chen, M Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, Secure Deduplication with Efficient and Reliable Convergent Key Management, IEEE transactions on parallel and distributed systems, vol. 25, no. 6, pages 1615-1625,2014

[3] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang, Mohammad Mehedi Hassan, Secure Distributed Deduplication Systems with Improved Reliability, Computers, and IEEE Transactions on Computers (accepted) pages 1-11, 2015

[4] 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.

[5] Chun-Ho Ng, Patrick P. C. Lee, RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups, Proc. of APSYS, 2013

[6] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicate storage. In USENIX Security Symposium, 2013.

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