International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 9 5519 - 5523

Avoid duplicate key overheads for same data in storage

Reshma S.Dhake
Dept.of Computer Engineering
BVCOE & RI ,Anjaneri
Trimbakeshwar, Nashik 422213
reshmadhake24@gmail.com

Swati B.Mali
Dept.of Computer Engineering
BVCOE & RI ,Anjaneri
Trimbakeshwar, Nashik 422213
swatimali0909@rediffmail.com

Shweta B. Pawar
Dept.of Computer Engineering
BVCOE &RI,Anjaneri
Trimbakeshwar, Nashik 422213
pawar.shweta999@gmail.com

Roshani B.Torawane
Dept.of Computer Engineering
BVCOE & RI ,Anjaneri
Trimbakeshwar, Nashik 422213
roshani.torawane@gmail.com

Prof. V. D. Badgujar
Dept. of Computer Engineering
BVCOE & RI, Anjeneri,
Trimbakeshwar, Nashik 422 213
badgujarvivek83@gmail.com

***Abstract*—** De-duplication is a technique used to weaken the amount of storage needed by service providers. Now a day the most originating challenge is to perform secure de-duplication in cloud storage. Although convergent encryption has been extensively adopted for secure de-duplication, a demanding issue of making convergent encryption practical is to efficiently and reliably managea massive number of convergent keys. We first introduce a baseline approach in which each user holds an autonomous master key forencrypting the convergent keys and outsourcing them to the server. As a proof of concept, encompass the implementation framework of proposed authorized duplicate check scheme and conduct experiments using these prototype. In proposed system involve authorized duplicate checkscheme sustain minimal overhead compared to normal operations.De-duplication is one of important data compression techniques for eliminating duplicate copies of repeating data. For that purpose Authorized duplication check system is used. This paper addresses problem of privacy preserving de-duplication in cloud computing and introduce a new de-duplication system supporting for Differential Authorization, Authorized Duplicate Check, Unfeasiblity of file token/duplicate-check token, In distinguishability of file token/duplicate-check token, Data affinity.In this project we are presenting the certified data de-duplication toprotect the data security by counting differential privileges of users in the duplicate check.Different new de-duplication constructions presented for supporting authorized duplicate check.

Keywords*-deduplication, distributed storage system ,convergent encryption, key management .*

*****

I. INTRODUCTION

Deduplication techniques are greatly employed to backup data and minimize network and storage overhead by detecting and removing redundancy among data, with the sudden growth of digital data. Instead of keeping more than one data copies with the identical content, deduplication removes redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication has taken much attention from both academia and industry because it can greatly make better storage utilization and save storage space, especially for the applications with high deduplication ratio such as valuable records storage systems. A number of deduplication systems have been pro-posed based on various deduplication carrying out plans such as client-side or server-side deduplications, file-level or block-level deduplications .

The arrival of important of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which stimulate interest of enterprises and organizations to outsource data storage to third-party cloud suppliers , as clearly proofed by many existing-life case studies . According to the particular structure of written account of IDC, the volume of data in the world is demanded to reach 40 trillion gigabytes in 2020 .Today's commercial cloud storage services are Drop box, Google Drive and Mozy etc. They have been applying deduplication to save the network bandwidth and the data storing cost with client-side deduplication. There are two types of deduplication in form of the size, first one is file-level deduplication, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and second one is block-level deduplication, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. The computations of block boundaries simplifies using fixed-size blocks or variable-size blocks (e.g., based on Rabin finger printing). It provides better deduplication efficiency. Data reliability is actually a very captious dispute in a deduplication storage system because there is only one copy for each file stored in the server. These files are shared by all the owners. If such a shared file was lost, a disproportionately large amount of

data becomes remote because of the unavailability of all the files that share this file. In the additional, the challenge for data privacy also arises as more and users are being outsourced more sensitive data to cloud. Encryption mechanisms have usually been used before out sourcing data into cloud to protect the confidentiality. Most commercial storage service provider is reluctant to apply encryption over the data because deduplication is not possible by it. The reason is that the conventional encryption mechanisms require different users to encrypt their data with their own keys, containing public key encryption and symmetric key. However, at the cost of decreased error resilience, these systems accomplished confidentiality of outsourced data.

II. LITERATURE SURVEY

Data de-duplication techniques are very appealing techniques that are usually employed for data backup in enterprise environments to decry network and storage overhead by detecting and eliminating redundancy among data blocks. Works have not considered and accomplish the tag consistency and integrity in the constitution. Convergent encryption. Convergent encryption insures data privacy in de-duplication. Bellare et al. formalized this primitive as message-locked encryption, and analyze its application in space systematic secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption modification for secure de-duplication. It is known that some commercial cloud storage providers, such as Bitcasa, also locate convergent encryption addressed the key-management issue in block-level de-duplication by alloting these keys over multiple servers after encrypting the files. Bellare et al. showed how to protect data confidentiality by reconstructing the predicatable message into a unpredicatable message. In their system, another third party called the key server was initiated to expand the file tag for the duplicate check. Stanek et al. we presented a novel encryption scheme that provided distinctive security for popular and unpopular data. For popular data that are not particularly deplomatic, the traditional conventional encryption is performed. Another two-layered encryption scheme with robust security while supporting de-duplication was proposed for undesirable data. In this way, they achieved better trade off between the efficiency and security of the outsourced data.Proof of Ownership scheme that has all features of the state of the art solution while incurring only a fraction of the overhead qualified by the competitor. In second, the security of the current mechanisms relies on information rather than assumptions. The quality of our current system is supported by extensive benchmarking. This scheme is used by R. D. Pietro in 2012[1].System formalize a new cryptographic primitive, Message Locked Encryption. Under the key encryption & decryption are performed itself developed from message. MLE provides secure deduplication , a goal recently achieved by numerous cloud-storage suppliers in year 2013[2].Rabins figure printing scheme is depends on mathematical module not reducible polynomial with coefficient in z, M. O. Rabin in 1981[3].The Farsite distributed file system gives availability by replicating each file onto multiple desktop computers. This repetation consumes significant storage space, it is important to recycle the space where possible by J. R. Douceur in 2002[4].In DupLESS, user convert data into code under message-based keys obtained from a key-server via an oblivious PRF protocol.It enables users to store encrypted data with an existing system, and it have the service perform deduplication on their behalf, and yet achieves strong confidentiality guarantees,M. Bellare in 2013 [5].

III. PROBLEM DEFINATION

The distributed systems counsel aim is to accurately store data in the cloud while achieving confidentiality and integrity. Its main goal is to permit and distributed storage of the data across multiple storage servers. In place of encrypting the data to reserve the confidentiality of the data, our new constructions utilize the secret splitting technique to split data into shards. These share will then be distributed across multiple storage servers. Building Blocks Secret Sharing Scheme. Two algorithms are presented in a secret sharing scheme, which are Share and Recover. The secret is divided and shared by using Share. With ample shares, the secret can be extracted and rediscovered with the algorithm of Recover.

IV. SYSTEM DESIGN

A. Existing System

The discrete kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data security, the problem of authenticating correctness of data storage in the cloud becomes even more challenging. Cloud Computing is not just a third party data storehouse. The data stored in the cloud may be frequently updated by the users, along with insertion, deletion, modification, appending, reordering, etc. One critical challenge of today's cloud storage services is the administration of the ever-increasing volume of data. According to the analysis report of IDC, the volume of data in the wild is conventional to reach 40 trillion gigabytes in 2020. The baseline approach suffers two critical deployment issues. First, it is ineffective as it will generate an enormous number of keys with the increasing number of users. especially, each user must accomplice an encrypted convergent key with each block of

its outsourced encrypted data copies, so as to next restore the data copies. Although different users may share the same data copies, they must have their own set of concurrent keys so that no other users can access their files. Second, the baseline approach is inaccurate, as it requires each user to dedicatedly protect his own master key. If the master key is unintentionally lost, then the user data cannot be recovered; if it is negotiate by attackers, then the user data will be leaked.

B. Proposed System Architecture

When the user desire to upload & download the file from cloud storage at that time first user request to the web server for uploading file means one of only authorized user can upload the file to web server for that purpose it use the confirmation of ownership algorithm. User to prove their ownership of data copies to the storage server. When file is uploaded it divides into blocks i.e block size is 4KB by default. According to file size the block occurs. Each block consist of there own cipher text , token for the unique identification and private key.

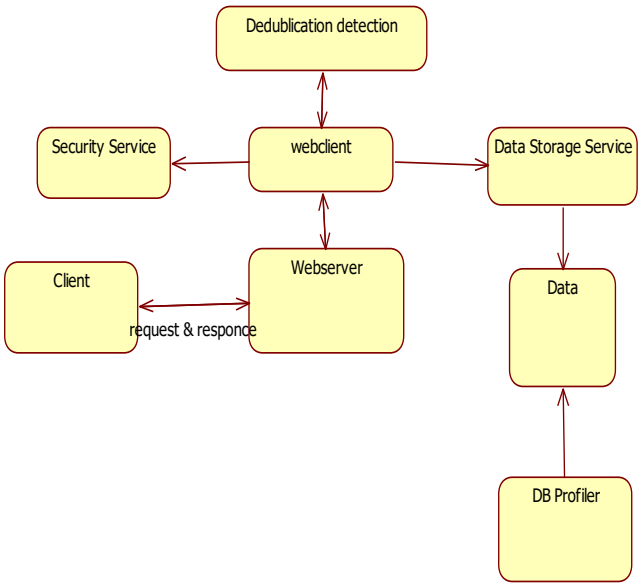

Fig- System Architecture

The data storage server comprise all the uploaded files and DB profiler store all the metadata of the file.

- Case 1:When file F1 & F2 are different the all the data will be store in the database in different blocks.
- Case 2: If the file F1 =F2 it stores only one file in the database avoid duplication of the data.
- Case 3: If F1€F2 then it compare the blocks with data storage and only different blocks of both file will be store in the data.

C. Activity Flow

- In fig a, data block A generate hash keyH0. Data block A encrypt with key H0 and H0 key encrypt with master key. Then data block A decrypt with H0 key. And also erH0 key decrypt with master key. Finally decrypted data goes to data block A.
- In fig b, data block A generate hash key H0 to Hm. Then first data block A encrypt with H0 key. In next step H0 encoded with H1 to Hm into n share. In download phase data block A decrypted with H0.After encoding process decode H0 from any S share and finally decode data & decrypted data goes to data block A.

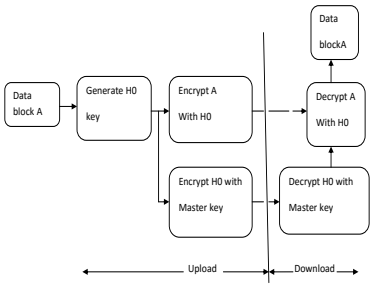

Fig. a

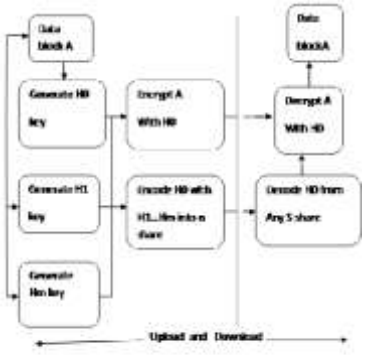

Fig. b

Fig – Flow of Secure Deduplication

D. Tag Generation Algorithm

In our constructions below, two kinds of tag generation algorithms are defined, that is, TagGen and TagGen'. TagGen is the tag generation algorithm that plans the original data copy F and outputs a tag T(F). This tag will be generated by the user and applied to perform the duplicate

check with the server. Another tag generation algorithm TagGen' takes as input a file F and an index j and outputs a tag. This tag, generated by users, is used for the proof of ownership for F.

E. RSA algorithm

Public key and private key involved in RSA. Everyone knows the public key which is used to encrypt messages. Private key decrypted this messages. RSA algorithm are generated the key in following way:

- Choose two different large random prime numbers p and q.
- Calculate η = pq.
- η is the modulus for the public key and the private keys
- Calculate the totient: ϕ = (p-1)(q-1)
- Choose an integer such that 1 <e<ϕ(η) , and e is co-prime to ϕ(η)i.e.: e and ϕ(η) share no factors other than 1; gcd (e,ϕ(η) = 1.
- eis released as the public key exponent.
- Computed to satisfy the congruence relation de ≡1(mod ϕ(η)) i.e. de = 1+ kϕ(η) for some integer k .
- d iskept as the private key exponent.

F. Mathematical Model

Suppose X be a system that find out duplicate copies of the file.

X={S,O,A,B,T,R,M} Where, S={S1,S2,S3,…Sn}

S1={A1,A2,A3,…An}

A1={BAi,TAi,Rki}

BAi=Set of cipher text block

T = Token [16-Bit, unique token for Block]

R= Private Key (RKi) used for encryption & descryption mechanism

M=Metadata of fig

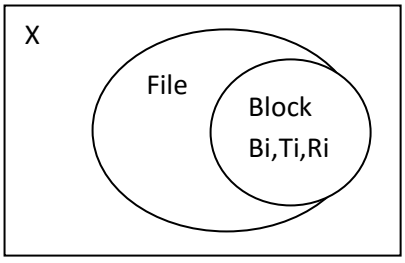

Fig.2 S∩(AU(B,T,R))

V. MODULES

A.Virtual Server Provider

- In this module, we establish Virtual Service Provider module. This is an entity that provides a data storage service in public cloud.
- The S-CSP contributes the data outsourcing service and stores data on behalf of the users.
- To reduce the storage cost, the S-CSP exterminate the storage of redundant data via deduplication and keeps only unique data.

B. Data Users Module

- A user is an entity that demand to outsource data storage to the S-CSP and access the data later.
- In a storage system supporting deduplication, the user one and only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be retained by the same user or different users.
- In the authorized deduplication system, each user is announced a set of privileges in the setup of the system. Each file is protected with the convergent encryption.
- key and privilege keys to recognize the authorized deduplication with differential privileges.

C. Private Virtual Server Module

- Correlational with the conventional de duplication architecture in cloud computing, this is a new entity introduced for facilitating user's protected usage of cloud service.
- Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully intimate in practice, private cloud is able to provide data user/owner with an execution environment and framework working as an interface between user and the public cloud.
- The private keys for the privileges are handled by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud grant user to submit files and queries to be securely stored and computed respectively.

D. Secure Deduplication System

 ISSN: 2321-8169
5519 - 5523

- We look at several types of privacy we need protect, that is, i) unforgeability of duplicate-check token: There are two form of adversaries, that is, external adversary and internal adversary.
- Secure Method For Authorized Deduplication And Data Dynamics In Cloud Computing
- As shown below, the external adversary can be viewed as an internal adversary without any privilege.
- If a user has privilege p, it desire that the adversary cannot forge and output a valid duplicate token with any more privilege p′ on any file F, where p does not match p′.

V. CONCLUSION

We proposed the distributed de-duplication systems to amend the reliability of data while achieving the confidentiality of the users outsourced data not with an encryption mechanism. Four constructions were proposed to support post-level and data de-duplication. Integrity and the security of tag consistency were achieved. We implemented our de-duplication systems using the Ramp secret sharing scheme and confirmed that it incurs small encoding/decoding overhead compared to the network transmission overhead in formal upload/download operations.

REFERENCES

[1] R. D. Pietro and A. Sorniotti, "Boosting efficiency and securityin proof of ownership for deduplication." in ACM Symposiumon Information, Computer and Communications Security, H. Y. Youmand Y. Won, Eds. ACM, 2012, pp. 81–82.

[2] 25(6), pp. 1 ——, "Message-locked encryption and secure deduplication," in EUROCRYPT, 2013, pp. 296–312.615–1625.

[3] M. O. Rabin, "Fingerprinting by random polynomials," Center for Research in Computing Technology, Harvard University, Tech. Rep. Tech. Report TR-CSE-03-01, 1981.

[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicatefiles in a serverless distributed file system." in ICDCS, 2002, pp. 617–624.

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

	**Reshma S.Dhake** is currently a student of Brahma Valley College of Engineering And Research Institute From the University of SavitribaiPhule Pune. His main research interests include: a)C++, b)Web designing
	**Swati B. Mali** is currently a student of Brahma Valley College of Engineering And Research Institute From the Universityof SavitribaiPhule Pune. His main research interests include: a) Database, b)Software Testing.
	**Shweta B. Pawar** is currently a student of Brahma Valley College of Engineering And Research Institute From the University of Savitribai Phule Pune. His main research interests include: a)Innovation in Computer Networking and Technology.
	**Roshani B. Torawane** currently a student of Brahma Valley College of Engineering And Research Institute From the University of SavitribaiPhule Pune. His main research interests include: a) software developer.
	**Prof.Vivek D.Badgujar**is working as a professor at Brahma Valley College of Engineering and Research Institute,Nashik.Under the University of SavitribaiPhulePune.His areas of interest includes Software Engineering and Advanced Database.