Deduplication Using SHA-1 and IBE with Modified AES

Renuka C. Deshpande[1], **S. S. Ponde**[2]

[1]Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2016-2017

[2]Associate Professor, Department of Computer Science and Engineering Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies, Aurangabad Maharashtra state, India 2016-2017

Abstract: ***Deduplication process is being mostly use in cloud server space to shrink the quantity of server space and accumulate network bandwidth. To eradicate duplicate pieces of repeat data, Data deduplication is exclusive data compression proposal used. To protect the confidentiality & isolation of receptive data while supporting deduplication, the convergent encryption method has been proposed to encrypt data before outsourcing. To healthier data defense, this manuscript takes the primary effort to properly deal with the difficulty of certified data deduplication. Apart from usual deduplication structure, the differential rights concept for users is further measured in replica check moreover the records itself. There are numerous novel deduplication construction supporting certified replica check in a fusion cloud architecture. Security investigations express that our proposal is secure in conditions of the characterization specified in the proposed security representation. As a evidence of conception, we implement a trial product of our future certified duplicate check scheme and conduct testbed experiments using our trial product. We will enhance security of cloud data by using symmetric algorithms with least operating cost evaluate to previous research operations.***

Keywords: Deduplication, Confidentiality, Hybrid cloud, differential privileges

1. Introduction

Cloud Computing offers high performance, shared resources, parallel supercomputing, effectively infinite size to users. Users can access cloud-based applications and services from anywhere. All he need is a device with an Internet connection. Simultaneously, Cloud providing massive amount of storage and feasible resources within low costs. Increasing extreme amount of data in military & research areas, financial portfolios, IT environments, needs large storage database, large network bandwidth & power sources creating the major challenge in future. Data deduplication will be the only solution to this critical problem. Data deduplication be the efficient technique to density technique for dropping duplicate replica of data. It reduces the extent of byte sent over network and improve capacity to store data.

Rather keeping many copies of data it saves only one. deduplication has been done on various level like file level, block level, byte level, bit level, document level, piece level. Even though applying powerful algorithm data is susceptible to both in-house and outdoor attacks. A traditional encryption strategy requires special users toward encrypt their records with their personal keys. Convergent key is produced after computing hash price of the content of data. convergent keys are used to encrypt/decrypt the data, but in several papers, separate encryption algorithms be use to encrypt/decrypt data like symmetric or asymmetric algorithms. To prevent unofficial attackers to get to up to sensitive data (POW) evidence of ownership protocol is needed to check authority of users. If the identical files are found in S-CSP then pointer will be provided to the user for that file. Only data owners can decrypt that file using decryption keys. Convergent keys play the role of deduplication. The every file is having particular rights to specify different category of users & data owners are having differential privileges access. In subsequent systems they cannot hold up differential authorization replica check. The only user is capable to perform duplicate check which has found matching privileges and copy of same file in cloud storage.

2. Literature Survey

In [1] Guljar P. shaikh has proposed in his paper new deduplication technique used hybrid cloud structure to remove previously created issues. Authorised Duplicate Check of data is performed to identify replicas from stored files. Here two main ways are present, deduplication on file level and block/chunk level. The SHA-1 And MD-5 type of cryptography hashing algorithm were used for the purpose of finding tokens of duplicates. MD-5 is faster than SHA-1on 32 bit machines. Key length of SHA-1 is 160 bits. Deduplication is performed for keeping a backup of the data and for the purpose of disaster recovery. Deduplication only store distinct form of data. Time performance is measured on a variety of files and their sizes. Hybrid cloud approach offers enhanced security and less utilization of network bandwidth. Disadvantages of this paper are, it considers only single cloud, in future it may use for sky computing and can provide security strategies for multi users.

In [2] Distinguish sensitive data at data uploading into cloud level and apply the crypt algorithm for receptive data by applying this data get secured and authorized. It is restricted from unauthorised access. User needs B. Aparna has proposed In her paper, secured data deduplication mechanism by to authorised by data owner by having privilege keys taken from the data owner through any one of secured communication system i.e. E-mail, unless and until get the access key from data owner. Encryption of data is done using symmetric cryptography- algorithm. The user must be registered for to upload data into clouds by providing required info... like name, password, and email, mobile. Data owner will make an account in our application by the registration form and by his/her signature he can login into our application they can upload and download data from our cloud server provider the data will be provided security by encrypting the data contents in the files and giving privileges to other data users according.

While data uploading using user into public cloud, the identification of duplicate data will be notified by showing the warning pop MSG to users if the user want to upload existing files again, still user want to upload file the new file need to update with presented file. While user uploading data into public cloud user can distinguish susceptible and non-susceptible data and can provide encryption for only sensitive data.

If any unauthorized user wants to access or the user didn't have particular privileges (like read writes, if the user is, having read privileges, but they want to access file (downloading like that)) immediately message alert needs to send to for a particular data owner Here addressed for every uploaded data into public cloud by separate the procedure of susceptible data and non susceptible data. To user requests and that given privilege information will be sent to users registered e-mail. The OTP will create and send to user registered personal mobile number. In this paper they only eliminate repeating copies of data on the basis of the file name.

In [3] the proposed project is concentrated on, only file level deduplication, including degree of difference privileges of users in the duplicate check. Here also used hybrid cloud scheme, in which personal cloud server generates duplicate check sign of the files. Consider the Example, a project can use an open cloud service, such as Amazon S3. In future may increase the national security. It could save memory by deduplicating the data and thus provide us with strong memory. It could support authorization to private firms and protects confidentiality of data.

In [4] proposed work shows that hashing algorithm is used for encryption/decryption data with a convergent key, which is found after cryptography hash of the content of data. Here implement play fair modified version to generate authentication sign as 2FA–OTP, communication to user's registered e-mail id and mobile no, using SMTP protocol and free SMS service for authorised logging purpose. The byte level deduplication check with file and variable size block –level has been implemented. With increasing file size, time spent for checking duplication, Encryption and transfer process is also increases. It analyses deduplication time factor at file level, at byte level and uploading unique files at different types.

In [5] document level deduplication, it takes out copy duplicates of same record, at the piece level, which takes out copy squares of information that happens in non distinguishable records. The objective of the work to improve integrity, to increase storage utilisation, to remove the duplicates copies of data and improve the reliability. In proposed architecture secrete sharing scheme performs two operations namely share and Recover. His secrete has divided and shared by via share. With enough shares, the secret can be extracting and improved with algorithm of recover. The input to this module is file. It performs dividing of file into fixed size blocks or share. These blocks are then encoded and allocated on cloud server at different nodes.

3. System Description

In the Hybrid cloud mixture of public cloud and private cloud is present. In deduplication process the cloud service provider is present in the public cloud, private cloud is separate cloud to save sensitive data. The tokens are provided to users for accessing the files from private cloud.

3.1 Public Cloud

Public cloud architecture is designed with the view to create an accessible business environment that can be shared and accessed from anywhere and at any time of the hour in less expensive costs. They are hosted at the vendor's premises. It provides the benefits like, supports, low investment, multiple customers, cheaper than private cloud, shared infrastructure, support connectivity on internet, highly scalable, cost effective, flexibility, Location independence.

Limitations:

1) It gives less security because of its openness.
2) Security concerns: multi-tenancy and transfers on Internet.

Examples of public cloud companies are sales force, engine yard, Gcloud3, Google, Microsoft, Rackspace, etc.

3.2 Private Cloud

1) High investment problem in private cloud implementation, along with purchases of new hardware and software.
2) New set of processes are essential, older processes not all appropriate for private cloud.
3) Computing architecture be enthusiastic to the customer and is not shared with other organizations.
4) Provides more control on data centers, hardware & resources, improves reliability.

Limitations

1) They are expensive and be considered highly secure than Public Clouds.
2) Private clouds may be externally hosted ones as sound as in premise hosted clouds.

Examples of private cloud companies are Amazon EC2, elastic computer cloud, IBM's Blue cloud, Sun cloud, Google app engine, windows Azure service platform.

3.3 Hybrid Cloud

Critical Activities are performed using private cloud while the non-critical activities are performed using public cloud.

1) One can manipulate and configure the applications online at any time.
2) It does not require to install a software to access or manipulate cloud application.

3) Scalability: run peak and bursty workloads on the public cloud.

Challenges:
1) Security7 privacy
2) Portability
3) Interoperability
4) Comuting performance
5) Reliability & availability.

Applications of hybrid cloud are in business Environment, infrastructure, social networking, management, education, art and global positioning.

In existing systems old techniques were used like coarse grained approach, HMAC, OTP because of that, lack of security. Lack of user privacy, less data confidentiality, lack of integrity, unsecured data deduplication occurs.

Table 2.1: Summary of History Papers

Sr.	*Author*	*Key points*	*Algorithm used*
1.	Guljar P. shaikh	--File & Block level --Time Performance on the basis of variety of files & their sizes. --Used single cloud	SHA-1
2.	B. Aparna	--Sensitive --Nonsensitive --Block/Name Of File -- Role based,Time based, Attribute access ---rights, access according to priviledges --Only on the basis of file name	HMAC based OTP/ symmetric encryption/ SHA-1
3.	Prof. N. B. kadu	--File Level --Restrics from unauthorised accesss --Differential priviledges of users in duplicate check --No security	Hashing Algorithm
4.	Sumedha A telkar	--File Level --Restrics from unauthorised accesss --Differential priviledges of users in duplicate check --No security	2FA-OTP
5.	Vutukuri Bharath	--file/block level --Defend the confidentialityusing differential priviledges --mention dismissed dataconvergent key encryption	SHA-1

To overcome these types of downsides, new authorized deduplication strategies used. When the deduplication has done at the client side it means data has been hashes first in client side and the results in the form of tokens, to check duplicate check it is called as source based deduplication. It achieves saving bandwidth but, unfortunately vulnerable to side channel attacks.

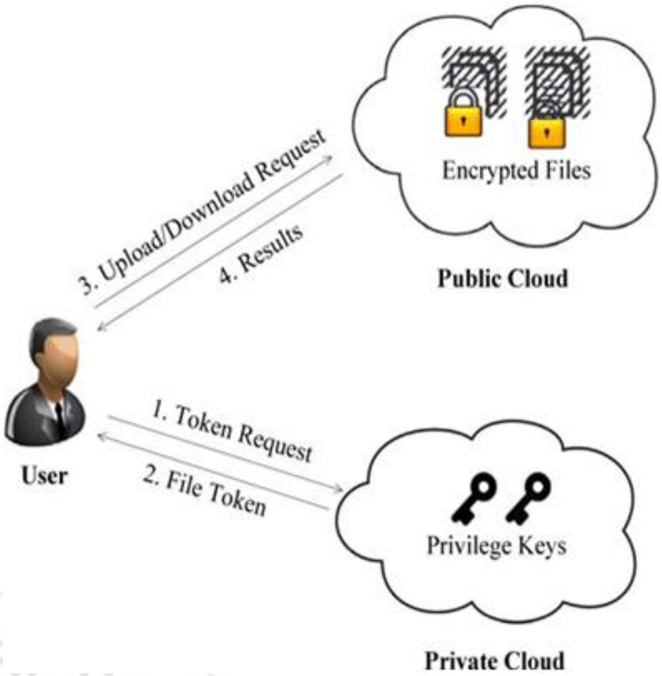

Figure 1: Hybrid Cloud Architecture

4. Conclusion

This paper says, we study different hashing & encryption algorithms proposed in history papers which were used to encrypt very sensitive data. In authorized deduplication process of work differential privileges are used to give access rights to data server and to data users for the purpose to keep confidentiality, maintain isolation, avoid collisions and keep high security aspects. Deduplication procedure could help to reduce network bandwidth, improve storage capacity and utilize space in cloud providers.

References

[1] Guljar P. Shaikh1, S. D. Chaudhary, Priyanka Paygude, Debnath Bhattacharyya "Achieving secure deduplication by using private cloud and public cloud"-International Journal of Security and Its Applications Vol. 10, No. 5 (2016) pp.17-26 2016

[2] B. Aparna, Prof K. S. M. V Kumar "Privacy preserving and Authorized data Deduplication in public cloud framework,"- International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 10, October-2015.

[3] Prof. N.B. Kadu, Mr. Amit Tickoo, Mr.Saurabh I. Patil, Mr. Nilesh B. Bhagat, Mr. Ganesh B. Divte "A hybrid Cloud approach for secure Authorized deduplication"- International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015.

[4] Sumedha A. Telkar (S. A. Maindakar) 1, Dr M Z Shaikh "secured and efficient cloud storage data deduplication system"- International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016.

[5] Rushikesh Naiknaware, Omkar Deshpande, Abhishek Kangle, Himalay Koli, Dipmala Salunke "A hybrid approach for secure authorised deduplication"- International Journal of Innovative Research in Science, Engineering and Technology Vol. 5, Issue 5, May 2016.

[6] Jin Li, jingwei Li,Xiaofeng Chen,Chunfu jia and wenjing Lou "identity-based Encryption with outsourced revocation in Cloud Computing"- IEEE TRANSACTIONS ON COMPUTERS.

[7] Rohan Rayalikar, Sanket upadhyay, Priyanka pimpale "SMS Encryption using AES Algorithm on Android"- International journal of Computer application(0975-8887) Volume 50-No.19, july 2012.