International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

Revocation based De-duplication Systems for Improving Reliability in Cloud Storage

Ketaki Ohale[1] , Aparna Junnarkar[2]

[1,2] *Department of Computer Engineering*
P. E. S. Modern College of Engineering, Pune-05

Abstract - *Due to the popularity of network and it's applications, users are willing to store and share their precious and confidential data. Cloud is the most prominent service which helps users to accomplish this task. Cloud as a service provides security as well as storage for user data. Due to some attacks or threats, these precious data may get duplicated illegally. This causes loss of revenue and violations of intellectual property rights of data owners. To avoid this, deduplication systems are deployed on the cloud. There are some raising concerns about these deduplication systems. In this paper, we mainly focus on the reliability issues of the deduplication systems. Here Shamir's secret sharing Scheme and Ramp Secret Sharing Scheme are implemented along with user revocation to improve reliability of the deduplication systems in cloud environment.*

Key Words: Deduplication, Key Server(KS), Message Authentication, Ramp Secret Sharing Scheme (RSSS), Revocation, Shamir's Secret Sharing Scheme(SSSS).

1.INTRODUCTION

Cloud and networking applications are getting more and more popular day by day. So, users trust the cloud storage services to store and protect their valuable data. If this data gets duplicated, the content holders and content creators have to suffer revenue loss as well as loss of Intellectual Property Rights. Also, when multiple copies of same data exist in the cloud storage, it is causes the waste of storage space as well as upload bandwidth. These problems are overcome by deduplication systems. A deduplication system eliminates redundancy of data and achieves single instance storage.

Deduplication systems can be deployed in standalone as well as distributed cloud environments. Out of them Convergent Encryption with Ramp Secret Sharing Scheme (CE-RSSS) is preferred on the basis of these parameters. Reliability and security are the raising issues in the deduplication systems. In these deduplication systems when data owners and users are revoked, they affect the reliability of the system. If User is removed then its file or block also get remove from server. So to maintain reliability deduplication has to be checked again. Shamirs Secret Sharing algorithm(SSSS)and Ramp Secret Sharing Scheme (RSSS) is used as encryption approach for secure deduplication. These de-duplication techniques save upload/download bandwidth and improve storage space utilization. These systems provide user privacy, data confidentiality and security against various attacks. Hence, deduplication satisfies the increasing need of cost efficient storage solutions.

2. REVIEW OF LITERATURE

Encryption helps to securely store this data on the cloud server. So, there are some variants of encryption like Message Locked Encryption and Convergent Encryption which can be used as approaches for secure deduplication in cloud environment. Out of these Message-Locked Encryption (MLE) is called as a cryptographic primitive, where message is encrypted and decrypted with the key which is derived from that message [1]. So, MLE was used for secure deduplication of small sized files. But in order to deduplicate large sized files there is need for maintenance of a large amount of metadata too. BL-MLE scheme can be called as enhancement over MLE for this purpose which can achieve file level as well as block level deduplication using small amount of metadata [2].

A secret sharing or secret splitting scheme is the method of distributing a secret i.e. a data copy among some participants (distributed cloud servers). The secret can be reconstructed from a sufficient number of shares. Some variants of Shamirs secret sharing Scheme (SSSS) [10] and Ramp secret sharing Scheme (RSSS) are studied where RSSS is proved to be the method to provide fast computation and storage efficiency [8].In the case of deduplication systems, security, data confidentiality, user privacy, efficiency and reliability are the main concerns which are taken into the consideration.

Deduplication is the process of removing redundant data which is getting significant attention from cloud industry. Owner encrypts the data with his public key and only authenticated users can access this data with their private keys. Convergent encryption is stated as proof of security in SALAD, a Self-Arranging, Lossy, Associative Database. This is an attempt for aggregating and analyzing file content information which is used for identification of identical encrypted files across a large number of machines. Usually these machines situated in robust and decentralized manner [13]. To reclaim spaces from duplicate files of a

scalable, server less, distributed file System Farasite convergent encryption is used.

DupLESS (Duplicate Less Encryption for Simple Storage) is an architecture which enables clients to encrypt message with keys obtained from a key-server (KS) via PRF Protocol and promises the message confidentiality. In the case of Dup-LESS, brute force attack is dealt by cipher text recovery[6].

DupLESS works on the principle that the brute-force attack can be dealt with by using a key server (KS) to derive keys, instead of setting keys to be hashes of messages. In the case of convergent encryption, a special kind of server called as key server (KS) is used to manage the convergent keys. So, Dekey is the construction which is proposed using Ramp Secret Sharing Scheme(RSSS) which causes a small encoding and decoding overhead to be achieved in upload/download operations[5].

In case of Convergent Encryption using Ramp Secret Sharing Scheme the original data copy is first encrypted with a convergent key derived by the data copy itself, and the convergent key is then encrypted by a master key that will be kept locally and securely by each user. The encrypted convergent keys are then stored, along with the corresponding encrypted data copies, in cloud storage. The master key can be used to recover the encrypted keys and hence the encrypted files. In this way, each user only needs to keep the master key and the metadata about the outsourced data.

File level and block level deduplications tend to improve the reliability of the deduplications systems in distributed environment. Ramp Secret Sharing Scheme which splits secret within distributed cloud servers provides better fault tolerance with data confidentiality[7].

A new way of public auditing mechanism for shared data with efficient user revocation in cloud is proposed. When a user is revoked, there is a semi-trusted cloud which re-signs blocks that were signed by the revoked user with the use of proxy re-signatures[16].

3.SYSTEM ARCHITECTURE / SYSTEM OVERVIEW

The major entities involved in the process are as follows:

1) Users and Owners:

– Owner: Data owner owns data in the form of multimedia and wishes to upload it on the cloud. If the contents get duplicated illegally, owners have to face the revenue losses for those contents. So, data owners wish to protect their data from duplication.

– User: A user wishes to access the data owned by an owner. A user is an entity who needs access to the resources and storage space offered by the cloud. They can access that data only when data owner allows them to do so. If users can access the owners data illegally, piracy has said to be happened. Data owners are also one of the data users, vice versa is not possible.

2) S-CSP: Storage Cloud Service Provider lets users access all applications and documents from anywhere in the world by providing a huge space for multimedia data. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data.

The owner owns a file which is uploaded on a distributed cloud. The file is broken into pieces called shares and its corresponding tag are derived. The deduplication is checked and duplicated copies are removed. The secure deduplication methods for cloud based systems are as shown in fig 3.1.

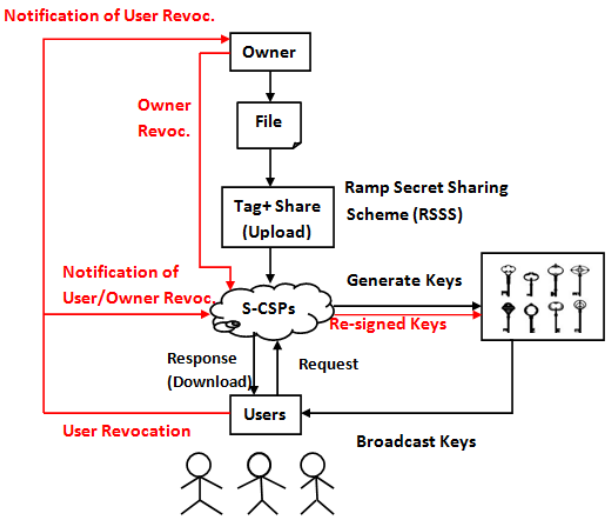

Fig -1: Block Diagram of secure Deduplication System

A user or data owner derives a key from each original data copy and encrypts the data copy with the same key. These keys are then stored in key Server. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. If two data copies are the same, then their tags are the same. Both the encrypted data copy and its corresponding tag will be stored on the server side. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Both the key and the tag are independently derived, and the tag cannot be used to deduce the convergent key and compromise data confidentiality.

A unique tag with index is generated by tag generation algorithm and is used for the Proof of Ownership (PoW) of the original data. Tags and shares are computed and are sent to the cloud service providers. The users then apply similar decryption schemes to download and access their data files.

While downloading the file the user sends request for a particular file to the S-CSP. S-CSP responds the user by providing the file by checking the user authentication and data integrity using message authentication. This is done to check integrity of the file downloaded by the user.

4. EXPRERIMENTAL SETUP

All types of multimedia files which are to be uploaded on cloud are considered as dataset. There is no restriction on the type of multimedia data uploaded on the cloud. It consists of images (JPEG, TIFF, GIF, PNG, BMP and many other raster and vector formats), audio/music clips (MPEG, AVI, MP3, WAV, WMA) and videos clips (AVI, MP4, FLV, WMV, MOV, WebM, MKV, FLV).

4.1 Software and Hardware Requirments

For the implementation of the system RAM of 1GB and above is preferred which is expected to process with speed of 1.1 GHz. The processor used is intel - CORE i3 and above with a Hard Disk of 20 GB. Java Version 1.8 installed on the Operating System Windows 7 with is used. Netbeans 8.0.2 is the Development Environment considered here. Cloud used for hosting is jelastic cloud. MySQL/SQLyog is used as database for implementation.

Shamir's secret sharing scheme (SSSS) and Ramp Secret Sharing Scheme(RSSS) are at the core schems which divide and manage this multimedia data in secure deduplication system. AES 128 bit encryption is used for key generation. SHA-256 encryption is done which produces output of 32 bytes.

4.2 Performance Parameters

Major performance considerations of the deduplication system consists of encoding and decoding times for RSSS parameters:

1) **Number of S-CSPs (n) :** n is number of S-CSPs used. The shares are divided and stored amongst n number of S-CSPs.

2) **Confidentiality Level(r):**Confidentiality level is denoted as r which is the minimum number of shares required to recover the secret in RSSS.

3) **Reliability Level(n-k):** Reliability analysis depends upon RSSS utilization and it is determined by parameters of n and k. According to RSSS data can be recovered from any k shares distributed amongst n S-CSPs. Here, k is the maximum number of shares required to recover the secret.

4) Encoding/Decoding time : The total time taken for hash generation and encryption is termed as Encoding time. On contrary to this decoding time is the time required for reconstructing the original data using the key.

5.RESULTS

Table-I gives the values of number of SCSPs and encoding and decoding times for existing deduplication system

Table -1: Impact of number of S-CSPS(n) on encoding and decoding times

S-CSP	Existing System		Implemented System	
	Encoding Time (usec)	Decoding Time (usec)	Encoding Time (usec)	Decoding Time (usec)
3	150	146	122	120
4	600	500	300	240
5	1020	800	600	400
6	1356	900	1000	850

and implemented deduplication (with revocation) for each block(4KB).

Here minimum number of S-CSPs is 3 because it denotes the threshold for Shamir's secret sharing scheme.

Fig.2 shows the bar graph of values of Encoding and Decoding times of existing deduplication system and implemented deduplication.

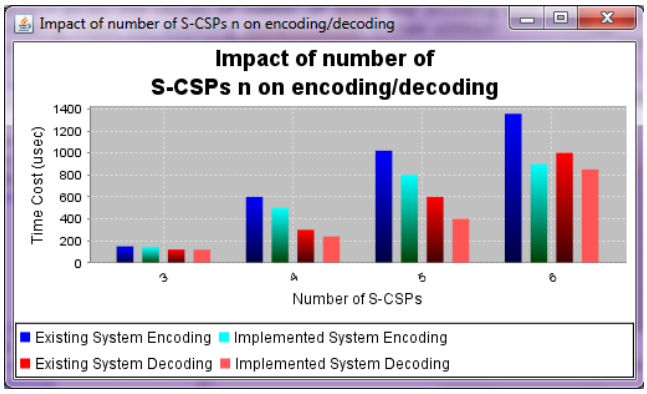

Fig -2: Impact of number of S-CSPs n on encoding and decoding Times

It is observed that encoding time is always higher than the decoding time because more number of shares are involved in encryption process.

Table-2 gives the values confidentiality level(r) v/s encoding and decoding times for existing deduplication system and implemented system with revocation.

Table -2: Impact of confidentiality level (r) on encoding and decoding times

Confidentiality Level	Existing System		Implemented System	
	Encoding Time (usec)	Decoding Time (usec)	Encoding Time (usec)	Decoding Time(usec)
0	500	540	390	410
1	580	610	425	450
2	610	660	500	560
3	800	900	620	680

Confidentiality level r denotes the minimum number of shares which can be kept confidential. No information about the secret cannot be deduced from r or less shares. When the minimum threshold for SSSS is 3, confidentiality level takes values less than equal to 3.

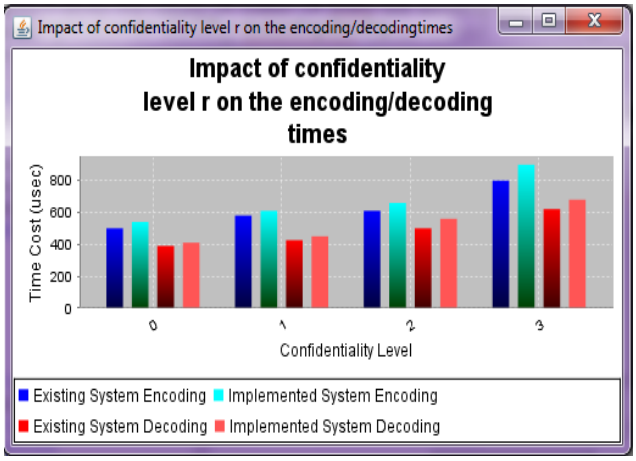

Fig. 3- Impact of confidentiality level r on Encoding and Decoding times

Reliability level is denoted with n - k. So, the difference between number of shares and minimum number of shares required to recover the secret.

Table-3 gives values on Impact of reliability level n - k on encoding and decoding times.

Table -1: Impact of Reliability level (n-k) on encoding and decoding times

Reliability Level	Existing System		Implemented System	
	Encoding Time (usec)	Decoding Time (usec)	Encoding Time (usec)	Decoding Time (usec)
1	710	610	640	580
2	600	510	550	460
3	540	400	400	320

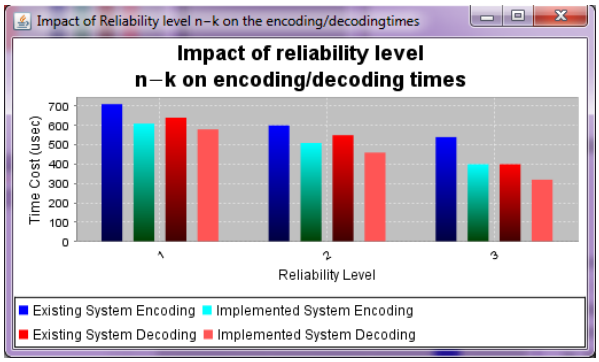

Fig. 4- Impact of reliability level n-k on encoding and decoding times

It is observed that the encoding time is higher than the decoding time because encoding operation always involves all n shares, while the decoding operation only involves a subset of n shares. Here, r is set to be 2 and n = 5 is also fixed.

Fig.4 shows impact of reliability level n-k on encoding and decoding times of deduplication system for each block (4KB). This shows that deduplication systems with user revocation shows decreased encoding and decoding times. Because instead of n shares, only k shares are involved in encoding.

5. CONCLUSION

Cloud computing is an emerging technology. Due to the increasing volumes of data in the digital world, data owners and data users are getting attracted to cloud storage services for protecting their data. These users trust cloud storage servers to store their precious and sensitive data. So, cloud servers apply deduplication methods on this data and store only unique copies. This helps to reduce upload bandwidth on user side and storage space on server side. To meet these increasing security and performance requirements, convergent encryption using Ramp Secret Sharing Scheme (CE-RSSS) was introduced. It uses convergent keys to identify duplicates. Due to its

vulnerability to online and offline attacks like brute force and collusion attacks, it is getting popular and is considered appropriate for different kinds of multimedia data. Also the deduplication system using RSSS encures lesser computational cost and encoding time for the same amount of data.

Data reliability is a raising issue in cloud deduplication systems because there is only one copy for each file stored in the server which is shared by all the owners or users. When this shared file gets lost, a very large amount of data becomes inaccessible. Also, when one or more users or owners are no longer available, they need to be revoked. The act of terminating and previously granted rights to authenticated users and owners is called as Revocation. Hence, reliability is improved by maintaining confidentiality of the users with Owner and User revocation mechanisms.

When a user gets revoked, only remaining shares (n-k) are re-computed and keys are re-assigned. Hence, encoding of n-k shares is done insted of n shares. Here the encoding and decoding times of deduplication system with revocation are reduced upto 18% than deduplication system without revocation.

Deduplication systems using convergent encryption along with Ramp Secret sharing Scheme (CE-RSSS) improves storage utilization, saves network bandwidth, storage cost and storage space efficiently.

ACKNOWLEDGMENT

Every orientation work has an imprint of many people and it becomes the duty of the author to express the deep gratitude for the same. I feel pleasure to express deep sense of gratitude and indebtedness to my guide Prof. (Ms) Aparna Junnarkar, for constant encouragement and noble guidance. I also express my sincere thanks to the Computer Department as well as Library of my college. Last but not the least, I am thankful to my friends and my parents whose best wishes are always with me.

REFERENCES

[1] Mihir Bellare, Sriram Keelveedhi, Thomas Ristenpart, "Message-locked encryption and secure deduplication", in EUROCRYPT , 2013, pp. 296-312.

[2] Rongmao Chen, Yi Mu, Guomin Yang, Fuchun Guo, "BL-MLE: Block- Level Message-Locked Encryption for Secure Large File Deduplication", IEEE Transactions on Information Forensics and Security, Volume: 10, Issue: 12, Dec 2015.

[3] Dharani P, Berlin M.A.,"Survey on Secret Sharing Scheme with Deduplication in Cloud Computing", IEEE 9th International Conference on Intelligent Systems and Control (ISCO),October 2015.

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

[5] J. Li, X. Chen, M. Li, J. Li, P. Lee and W. Lou, "Secure deduplication with effcient and reliable convergent key management", in IEEE Transactions on Parallel and Distributed Systems, 2014.

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

[7] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang ,"Secure Distributed Deduplication Systems with Improved Reliability", IEEE Transactions On Computers Volume, Year:2015.

[8] A. D. Santis and B. Masucci, "Multiple ramp schemes", IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 17201728, Jul. 1999.

[9] G. R. Blakley and C. Meadows, "Security of ramp schemes", in Advances in Cryptology: Proceedings of CRYPTO 84, ser. Lecture Notes in Computer Science, G. R. Blakley and D. Chaum, Eds. Springer-Verlag Berlin/Heidelberg,vol. 196, pp. 242268,1989.

[10] A. Shamir, How to share a secret, Commun. ACM, vol. 22, no. 11, pp. 612613, 1979.

[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, Proofs of ownership in remote storage systems. in ACM Conference on Computer and Communications Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds. ACM, 2011, pp. 491500.

[12] M. Li, C. Qin, P. P. C. Lee, and J. Li, Convergent dispersal: Toward storage-efficient security in a cloud-of-clouds, in The 6th USENIX Workshop on Hot Topics in Storage and File Systems, 2014.

[13] P. Anderson and L. Zhang, Fast and secure laptop backups with encrypted de-duplication, in Proc. of USENIX LISA, 2010.

[14] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, Secure data deduplication, in Proc. of StorageSS, 2008.

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

[16] Boyang Wang, Baochun Li, Hui Li, Panda: Public Auditing for Shared Data with Efficient User Revocation in the Cloud , IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 8,ISSUE.1, 2015.

BIOGRAPHIES

Ms. Ketaki Ohale received the B.E. (Computer Science &Engg) degree from Sipna College of Engineering and Technology ,Amravati, India in 2015 and pursuing M.E. (Computer) degree in P. S. E. Modern College of Engineering, Pune University, Pune, India. Her research interests include Cloud Computing and Networking. Published a paper in IJCA in December 2016.

Prof. Mrs. Aparna Junnarkar Prof. Mrs. Aparna A. Junnarkar is presently working as Assistant Professor at P. E. S. Modern College of Engineering, Pune, India. Her research areas include Network and network security. She is having 16 years of experience in teaching. She received B. E. (Computer Science and Engineering) degree in 1999 from Shivaji University, Kolhapur, M. E. (Computer Engineering) degree in 2009 from University of Pune, India. ”