Indian J.Sci.Res. 17(2): 159-162, 2018 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

ENHANCING THE SECURE FILE-LEVEL AND BLOCK-LEVEL AUTHORIZED DATA DE-DUPLICATION WITH RELIABILITY IN CLOUD ENVIRONMENTS

[1]Pingili Sravya
[1]Technical Support, RKS Saboo Motors, Bangalore.

Abstract— To eliminate duplicate copies of data we tend to use data de-duplication process. Still because it is employed in cloud storage to reduce memory space and upload bandwidth only one copy for each file stored in cloud which will be utilized by a lot of number of users. De-duplication method helps to enhance storage space. Another challenge of privacy for sensitive data additionally arises. The aim of this paper is to make the primary attempt formalize the thought of distributed reliable de-duplication system. In our projected system we tend to are progressing to develop new distributed de-duplication systems that are highly reliable. In de-duplication methodology data chunks are distributed across multiple cloud servers instead of using convergent encryption as in previous de-duplication systems we tend to use deterministic secret sharing scheme in distributed storage systems. Thus we are able to reach the required concepts for security that are data privacy and tag reference within the projected security model, Security analysis express that our de-duplication systems are secure.

keywords-System model, Secure-Cloud Storage Provider, Data De-duplication, Secret Sharing, File-level De-duplication, Block-level De-duplication.

I. Introduction

In our planned system we tend to are aiming to use data de-duplication method. Initial we should know what is data de-duplication, it reduces the number of data that must be physically stored by eliminating extra data and replacing when repetition of it with a respect to the first. In data de-duplication we tend to remove unwanted copy of data and save the memory space. With the use of de-duplication methodology reliability is improved moreover as a result of it avoids wastage of memory area .Secure implies that in our system we tend to use encoding and decode or decryptionmethods. Encodingmeans that convert plane text into encoded text these techniques called an encoding. This encoded text is transfer to server CS1, CS2, CS3. Again this encoded text is converted into plane text mentioned as decoding techniques. Distributed means that we tend to create server CS2, CS2, CS3 from this server user select any one, from server choice data was distributed information de-duplication means that delete duplicate copy in which use two techniques initial one is file level de-duplication and alternative is block level de-duplication. In file level duplication authenticatetheduplicate copy file name wise that discover redundancy between file and block; remove this redundancy. In block level duplication checking duplicate copy blocks wise that discover redundancy between absolutelydissimilar block. File is split into smaller fix size or variable size block throughout this block contain ten number of linereliability means that maintain integrity. Today's industrial cloud storage service like Google drive, mostly we have been applying de-duplication to save lots of network bandwidth. With the insubstantialgrowth of digital information, de-duplication techniques are extensively used to store and back-up the data and minimize network and storage overhead by detection and eliminating excess among data. Instead of keeping multiple information copies with identical content, de-duplication eliminates duplicate information by storing just one physical copy and referring different unwanted data to that copy. De-duplication system is usually used in every business and tutorial as a result of it'll save storage space on memory and a lot of increase storage usage on memory, particularly for applications that have high de-duplication quotient like concurrence storage systems. Once use this techniques eliminate duplicate copy. For reducing storage space and uploading bandwidth in great operation it has used, in cloud storage. A very different type of de-duplication systems has planned those are supported be different of approaches those methods redolent of client-side or server-side de-duplications, block-level or file-level de-duplications. The most aim of our planned system is to explain the distributed and reliable de-duplication system with a lot of security. We tend to are described a new distributed de-duplication system, that has more and more reliability. In this chunks are distributed across various or multiple cloud servers. De-duplication method can used for to save the memory area on the memory for the cloud storage this is often reduces the reliability of the system. Security analysis indicates that our de-duplication systems unit of measurement secure in terms of the definitions specified in this security model. As a proof of concept, we tend to implement the projected systems that indicate the acquired aerial is incredibly limited in actual environments. De-duplication methodology principally improves storage utilization and it saves space for storing .That is why the de-duplication system is beneficial in period moreover as in tutorial. It is useful in such application that has high de-duplication ratio like as actual storage system. the foremost

[1]Corresponding Author

industrial storage to the number of service providers is opposing to use encoding over the data as a result of it's not possible to make de-duplication. The reason of that system is that the traditional encoding mechanism.

II. Related Work

A number of De-duplication systems are planned supported numerous de-duplication strategies such as client-side or server-side de-duplication method, and file-level or block-level de-duplication method. To formalized this ancient as Message Locked encoding, and explored its application in space efficient secure out-sourced storage. There are additionally several implementations of convergent implementations of completely different convergent encryption variants for secure de-duplication. the key-handling issues in block-level de-duplication by dispense these keys across multiple servers once encoding the files Baler has showed how to protect data confidentiality by transforming the predictable message into a unpredictable message. Data reliability is really a very crucial issue in a de-duplication storage system there is just one copy for every file stored in the server shared by all the owners. Most of the previous de-duplication systems have only been considered throughout a single-server setting. The normal de-duplication strategies cannot be directly applied in distributed and multi-server systems.

III. Frame Work

In this paper, we tend to show however to design secure de-duplication systems with higher reliability in cloud computing. We tend to introduce the secure distributed cloud storage servers into de-duplication systems to generatehigh fault tolerance. To further protect data privacy, the secret sharing technique is used, that is additionally compatible with the distributed storage systems. In additional details, a file is initial split and encoded into fragments by using the technique of secret sharing, rather than encryption mechanisms. These shares are distributed across multiple or various independent storage servers. Additionally, to support de-duplication, a quick cryptographic hash value of the content are computed and sent each cloud storage server as a result of the fingerprint of the fragment hold on at every server. only the data owner who initial uploads the data is required to reckon and distribute such secret shares, whereas all following users who own a similar data copy do not need to calculate and store these shares any more. To recover data copies, users ought to access a minimum sort of storage servers through authentication and get the secret shares to reconstruct the data. In several words, the secret shares of data will only be accessible by the approved users who own the corresponding data copy. The planned system will provide the four new secure de-duplication systems they are efficient de-duplication with high reliability for block-level and file-level de-duplication, respectively. The secret splitting technique, rather than traditional encoding ways, is used to protect data confidentiality. Specifically, data are split into fragments by using secure secret sharing schemes and hold on at different servers.

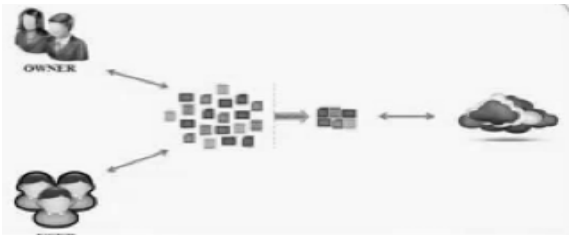

Figure 1: System Architecture

System Model:

In this initial module, we tend to develop two entities: User and Secure-Cloud Service provider. User: The user is an individual or entity that needs to source data storage to the Secure-Cloud Storage providers and access the data later. In a very storage system supporting de-duplication, the user only uploads distinctive data howeverdoes not upload any duplicate data to save lots of the upload bandwidth. moreover, the backup is needed by users in the system to provide higher reliability.

B. Secure-Cloud Storage Providers:

The S-CSP or Secure-Cloud Storage providers is an entity that provides the secure data storage service for the users. Within the de-duplication system, once users own and store identical content, the S-CSP or Secure-Cloud Storage providers will solely store one copy of those files and retain only distinctive data. A de-duplication technique, on the opposite hand, can reduce the storage value at the server aspect and save the transfer bandwidth at the user aspect. For backup and confidentiality of data storage, we tend to consider a gathering of Secure-cloud storage providers .The user data is distributed across multiple Secure-cloud storage providers.

C. Data De-duplication:

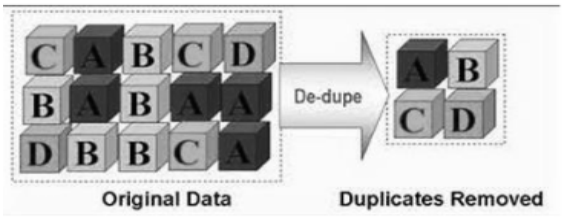

Figure 2: Data De-duplication

Data De-duplication involves finding or discovery and removing of duplicate data without considering its fidelity here the goal is to store lots of data with less bandwidth. Files are uploaded to the CSP and only the data owners can view and download it. The protection wants are also achieved by Secret Sharing theme. Secret Sharing proposal

uses two algorithms, share and recover. Data is uploaded every file and block level and the finding duplication is additionally in the same procedure.

This is created possible by finding duplicate chunks and maintaining one copy of chunks.

D. Secrete Sharing Scheme:

The Secrete sharing scheme having two strategies are used that are Share and Recover. Share technique is used for divided and shared secret. With enough shares, Extracted and retrieved the key with the assistance of Recover technique. Share divides secret S fragments of same size that produces r for random fragments of the equal size, and translates into the similar size. Then outputs the original secret message authentication code is a small section of data used to attest a message and to provide integrity and authenticity certainty on the message. The de-duplication check in these proposed system we have two ways they are first one is File level de-duplication system and second one is Block level de-duplication. The File level duplication means it support capable duplicate check, tags for each file will be calculated and send to storage cloud service provider. To stop balanced invasion organized by the Secure-cloud storage suppliers, tag collected at totally different storage servers and the second one Block level duplication means in this module we are going to show to comprehendsuperior grained block-level distributed de-duplication scheme. Inblock-level de-duplication design, the user additionally should first perform the file-level de-duplication before uploading his file if there is no duplicate file is found, the user divides this file into blocks and performs block-level de-duplication. The System setup is similar to the file level de-duplication except the parameter changes. To transfer a block the user gets the secret shares and downloads the blocks from CSP.

Different feature of our proposal is that data integrity, also as tag consistency, is achieved. To our data, no existing work on secure de-duplication can properly address the responsibility and tag consistency drawback in distributed storage systems. Our planned constructions support each file-level and block-level de-duplications. Security analysis describes that the planned de-duplication systems are secure in terms of the definitions specified in the planned security model. In additional details, responsibility and integrity and confidentiality, are going to be achieved in our planned system. Two varieties of complicity attacks are considered in our solutions. These are the complicity attack on the data and also the complicity attack against servers. Particularly, the data remains secure even if the attacker controls a restricted number of storage servers. Our analysis results describes that the new planned constructions are economical and also the redundancies are optimized and comparable the alternative storage system supporting a similar level of reliability.

IV. Experimental Results

In our experiments, any number of users can registered and login into the system. Who are authorized users they can upload the files into the cloud. Those uploaded files are stored in chunk format in cloud. Those upload files duplicate check in two ways file-level and block-level if any duplicate files are available in file-level and block-level , then that file cannot uploaded in the cloud and to that particular file tag consistency will be assigned to the user. But that file can be downloaded by data owner as well as data users.

In the below chart we can observe that difference between the length of both Total execution time and RSSS execution time.

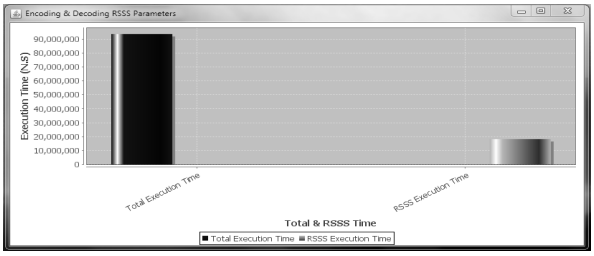

Figure:3

We can observe that Total execution time length is higher than RSSS execution time length. The difference will be shown in the sense of time length. So we can consider that the advantage of file compression.

Through our implementation we can store the big file in chunks format and detect the duplicate files as well as we can increase the storage space of cloud with file compression by using Ramp Secret sharing mechanism.

V.Conclusion

In this review paper, we tend to studied concerning some previously existed de-duplication system. In the study of literature survey, we have a tendency to examined that existed systems Have several problems such as, data reliability, data security, overheads of data storage etc. we tend to in addition detected that data de duplication reduces the data storage overheads by conserving single copy of data in server aspect. Within the study of base paper, we tend to analyze that S-CSP reduces storage price by storing distinctive copy data and bandwidth of transfer data at user side. Whereas, Ramp Secret sharing technique is economical for sharing data firmly in distributed system because it support to higher reliability and confidentiality levels of data. According to our analysis during this paper, Ramp secret sharing and S-CSP each are powerful to achieve improved reliability in decentralized data de-duplication.

References

[1] H. Shacham and B. Waters, "Compact proofs of retrievability," in ASIACRYPT, 2008, pp. 90–107.

[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East," reports/idcthe-digital-universe-in-2020.pdf, Dec 2012.

[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 duplicate files in a server less distributed file system." in ICDCS, 2002, pp. 617–624.

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Server aided encryption for de-duplicated storage," in USENIX Security Symposium, 2013.

[6] "Message-locked encryption and secure de-duplication," in EUROCRYPT, 2013, pp. 296–312.

[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Proceedings of CRYPTO at 84.

[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1720–1728, Jul. 1999.

[9] M.O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," Journal of the ACM, vol. 36, no. 2, Apr. 1989.

[10] P. Golle, I. Mironov Cryptographic primitives enforcing communication and storage complexity. In"Financial Cryptography ‟02", volume 2357 of LNCS, pages 120–135, 2003.

[11] A. Juels and B. S. Kaliski, Jr. Pors: proofs of retrievability for large files. In ACM CCS ‟07, pages 584–597. ACM, 2007

[12] H. Shacham and B. Waters. Compact proofs of retrievability. In ASIACRYPT ‟08, pages 90–107. Springer-Verlag, 2008.

[13] A.D. Santis and B. MasucciMultiple Ramp Schemes,"" IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1720-1728, July 1999.

[14] G.R. Blakley and C. Meadows, Security of Ramp Schemes‟ in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science, G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268.

[15] A. Sorniotti and E. Androulaki A secure data de-duplication scheme for cloud storage. 2014.