IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 12, December 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

Dynamic Ownership Management with Secure Data Deduplication in Cloud Storage

Priyankesh Sinha[#], **Er. Piyush Rai**[*]

[#]Computer Science & Engineering, I.E.T, Dr. Rammanohar Lohia Avadh University, Faizabad, Uttar Pradesh, India.

*Assistant Professor, Dept. of Computer Science & Engineering, I.E.T, Dr. R M L Avadh University Faizabad, Uttar Pradesh, India

Abstract— **In distributed storage administrations, deduplication innovation is generally used to lessen the space and transfer speed necessities of administrations by killing excess information and putting away just a solitary duplicate of them. Deduplication is best when various clients outsource similar information to the distributed storage; however it raises issues identifying with security and proprietorship. Evidence of-possession plans permit any proprietor of similar information to demonstrate to the distributed storage server that he claims the information powerfully. Nonetheless, numerous clients are probably going to encode their information before outsourcing them to the distributed storage to save security, yet this hampers deduplication due to the randomization property of encryption. As of late, a few deduplication plans have been proposed to tackle this issue by enabling every proprietor to have a similar encryption key for similar information. In any case, the vast majority of the plans experience the ill effects of security defects, since they don't consider the dynamic changes in the responsibility for information that happen as often as possible in a down to earth distributed storage benefit. In this paper, we propose a novel server-side deduplication conspire for scrambled information. It enables the cloud server to control access to outsourced information notwithstanding when the possession changes progressively by abusing randomized joined encryption and secure proprietorship amass key dispersion. This anticipates information spillage not exclusively to repudiated clients despite the fact that they beforehand possessed that information, additionally to a fair however inquisitive distributed storage server. Furthermore, the proposed plot ensures information uprightness against any label irregularity assault. In this manner, security is improved in the proposed conspire. The effectiveness examination comes about exhibit that the proposed conspire is practically as proficient as the past plans, while the extra computational overhead is immaterial.**

1. INTRODUCTION

Distributed computing gives versatile, minimal effort, and area free online administrations extending from basic reinforcement administrations to distributed storage foundations. The quick development of information volumes put away in the distributed storage has prompted an expanded interest for procedures for sparing circle space and system transfer speed.

To decrease asset utilization, many distributed storage administrations, for example, Dropbox, Wuala, Mozy, and Google Drive, utilize a deduplication method, where the cloud server stores just a solitary duplicate of excess information and gives connects to the duplicate as opposed to putting away other real duplicates of that information, paying little respect to what number of customers make a request to store the information.

The investment funds are critical, and allegedly, business applications can accomplish plate and transfer speed reserve funds of over 90. Be that as it may, from a security point of view, the mutual utilization of clients' information raises another test. As clients are worried about their private information, they may scramble their information before outsourcing with a specific end goal to shield information security from unapproved outside enemies, and in addition from the cloud specialist co-op.

This is advocated by current security patterns and various industry directions, for example, PCI DSS. In any case, ordinary encryption makes deduplication outlandish for the accompanying reason. Deduplication procedures exploit information similitude to recognize similar information and diminish the storage room.

Conversely, encryption calculations randomize the encoded records with a specific end goal to make cipher text indistinct from hypothetically arbitrary information. Encryptions of similar information by various clients with various encryption keys brings about various cipher texts, which makes it troublesome for the cloud server to decide if the plain information are the same and deduplicate them.

Say a client Alice encodes a document M under her mystery key skA and stores its comparing cipher text CA. Weave would store CB, which is the encryption of M under his mystery key skB. At that point, two issues emerge: (1) in what capacity can the cloud server recognize that the hidden document M is the same, and (2) regardless of the possibility that it can identify this, how might it enable both sides to recoup the put away information, in view of their different mystery keys? Direct customer side encryption that is secure against a picked plaintext assault with arbitrarily picked encryption keys avoids deduplication.

One credulous arrangement is to enable every customer to encode the information with general society key of the distributed storage server. At that point, the server can deduplicate the distinguished information by decoding it with its private key combine. Be that as it may, this arrangement permits the distributed storage server to get the outsourced plain information, which may abuse the protection of the information if the cloud server can't be completely put stock in [13],[14].

Merged encryption [15] settle this issue adequately. A merged encryption calculation encodes an info record with the hash estimation of the information document as an encryption key. The cipher text is given to the server and the client holds the encryption key. Since united encryption is deterministic1, indistinguishable documents are constantly scrambled into indistinguishable cipher text , paying little heed to who encodes them.

In this manner, the distributed storage server can perform deduplication over the cipher text, and all proprietors of the record can download the cipher text (after the confirmation of-possession (PoW) handle alternatively) and decode it later since they have a similar encryption key for the document.

United encryption has for some time been examined in business frameworks and has distinctive encryption variations for secure deduplication [8],[16],[17],[18], which was formalized as message blocked encryption later in [20]. Be that as it may, focalized encryption experiences security blemishes with respect to label consistency and proprietorship repudiation.

2 Problem Statements

Deduplication is best when different clients outsource similar information to the distributed storage, yet it raises issues identifying with security and possession. Verification of-proprietorship plans permit any proprietor of similar information to demonstrate to the distributed storage server that he claims the information vigorously. In any case, numerous clients are probably going to scramble their information before outsourcing them to the distributed storage to protect security, yet this hampers deduplication on account of the randomization property of encryption. As of late, a few deduplication plans have been proposed to take care of this issue by enabling every proprietor to have a similar encryption scratch for similar information.

Be that as it may, the greater part of the plans experience the ill effects of security blemishes, since they don't consider the dynamic changes in the responsibility for information that happen as often as possible in a handy distributed storage benefit. In this paper, we propose a novel server-side deduplication conspire for encoded information.

It enables the cloud server to control access to outsourced information notwithstanding when the possession changes progressively by abusing randomized focalized encryption and secure proprietorship amass key dispersion.

This counteracts information spillage not exclusively to denied clients despite the fact that they beforehand claimed that information, additionally to a genuine however inquisitive distributed storage server. Likewise, the proposed plot ensures information honesty against any label irregularity assault.

3 Contributions

We propose a deduplication conspire over scrambled information. The proposed plot guarantees that lone approved access to the mutual information is conceivable, which is thought to be the most essential test for effective and secure distributed storage administrations in nature where proprietorship changes powerfully.

It is accomplished by misusing a gathering key administration component in every possession gathering. When contrasted with the past deduplication conspires over scrambled information, the proposed plot has the accompanying points of interest as far as security and productivity. To start with, dynamic proprietorship administration ensures the retrogressive and forward mystery of deduplicated information upon any possession change.

Instead of the past plans, the information encryption key is refreshed and specifically appropriated to substantial proprietors upon any possession change of the information through a stateless gathering key dispersion system utilizing a parallel tree. The possession and key administration for every client can be directed by the semi-trusted cloud server sent in the framework. In this way, the proposed conspire delegates the most difficult assignments of possession administration to the cloud server without releasing any classified data to it, as opposed to the clients.

Second, the proposed plot guarantees security in the setting of PoW by presenting a re-encryption system that uses an extra gathering key for dynamic possession gathering. In this manner, in spite of the fact that the encryption key (that is the hash estimation of the document) is uncovered in the setting of PoW, the protection of the outsourced information is as yet saved against outside foes, while deduplication over encoded information is still empowered and information honesty against harm assaults is ensured.

4 LITERATURE SURVEY

On the basis of extensive literature survey related to the data deduplication with dynamic ownership management in cloud storage has been taken into consideration in this section.

D. T. Meyer, and W. J. Bolosky [1] has proposed that File frameworks regularly contain repetitive duplicates of data: indistinguishable documents or sub-record locales, perhaps put away on a solitary host, on a mutual stockpiling group, or moved down to optional capacity. Deduplicating stock piling frameworks exploit this excess to decrease the fundamental space expected to contain the record frameworks (or reinforcement pictures thereof). Deduplication can work at either the sub-document or entire record level. All the more fine-grained deduplication makes more open doors for space reserve funds, yet fundamentally lessens the successive format of a few records, which may have critical execution impacts when hard plates are utilized for capacity (and at times requires confused strategies to enhance execution.

M. Dutch[2] has stated that with respect to the comprehension of information deduplication proportions that information deduplication brings down business dangers, builds income openings, and diminishes stockpiling level expenses, bringing about an ideal tempest for organizations sending a versatile stockpiling foundation. Capacity flexibility advances, for example, RAID or RAIN, shield the deduplicated information to guarantee high accessibility of utilizations getting to the information. The financial aspects of information deduplication make it more than convincing; it is required for any business looking to expand their client benefit levels. Information deduplication proportions are anything but difficult to over-break down and credit advantages to, that could conceivably exist.

W. K. Ng et al. [3] has proposed about another idea which we call private information deduplication convention, a deduplication method for private information stockpiling is presented and formalized. Instinctively, a private information deduplication convention permits a customer who holds a private information demonstrates to a server who holds an outline string of the information that he/she is the proprietor of that information without uncovering additional data to the server.

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller [4] has proposed about the Businesses and buyers are ending up noticeably progressively aware of the estimation of secure, chronicled information stockpiling. In the business field, information safeguarding is regularly commanded by law, and information mining has ended up being an aid in forming business system.

For people, recorded capacity is being called upon to save nostalgic and authentic ancient rarities, for example, photographs, films and individual archives. Further, while few would contend that business information calls for security, protection is similarly imperative for people; information, for example, medicinal records and authoritative reports must be kept for drawn out stretches of time however should not be openly available. Incomprehensibly, the expanding estimation of authentic information is driving the requirement for cost-productive capacity; reasonable capacity permits the conservation of all information that may in the long run demonstrate helpful.

5. SYSTEM ANALYSIS

5.1. Existing System

At the point when a client transfers information that as of now exist in the distributed storage, the client ought to be dissuaded from getting to the information that were put away before he acquired the proprietorship by transferring it (in reverse secrecy)2. These dynamic proprietorship changes may happen every now and again in a functional cloud framework, and accordingly, it ought to be legitimately overseen keeping in mind the end goal to maintain a strategic distance from the security debasement of the cloud benefit.

In the previous approach, the greater part of the current plans have been proposed keeping in mind the end goal to play out a PoW procedure in an effective and powerful way, since the hash of the record, which is dealt with as a "proof" for the whole document, is defenceless against being spilled to outside enemies in view of its moderately little size.

An information proprietor transfers information that don't as of now exist in the distributed storage, he is called an underlying uploader; if the information as of now exist, called a consequent uploader since this suggests different proprietors may have transferred similar information beforehand, he is known as a resulting uploader.

5.2. Proposed System

A few deduplication plans have been proposed to tackle this issue by enabling every proprietor to have a similar encryption key for similar information. In any case, the majority of the plans experience the ill effects of security blemishes, since they don't consider the dynamic changes in the responsibility for information that happen as often as possible in a pragmatic distributed storage benefit. In this paper, we propose a novel server-side deduplication conspire for scrambled information.

It enables the cloud server to control access to outsourced information notwithstanding when the proprietorship changes powerfully by abusing randomized joined encryption and secure possession bunch key dispersion. A deduplication plot over encoded information. The proposed conspire guarantees that exclusive approved access to the mutual information is conceivable, which is thought to be the most essential test for proficient and secure distributed storage benefits in the earth where proprietorship changes powerfully.

It is accomplished by misusing a gathering key administration instrument in every proprietorship gathering. The proposed plot guarantees security in the setting of PoW by presenting a re-encryption system that uses an extra gathering key for dynamic proprietorship gathering. The vast majority of the plans have been proposed to give information encryption, while as yet profiting by a deduplication strategy, by empowering information proprietors to share the encryption enters within the sight of within and outside enemies. Since scrambled information are given to a client.

5.3. Advantages of the Proposed System

- ➢ Generate data tags before uploading as well as audit the integrity of data having been stored in cloud.
- ➢ Enables secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication.
- ➢ Integrity auditing and secure deduplication directly on encrypted data.

6. DATA DEDUPLICATION ARCHITECTURE

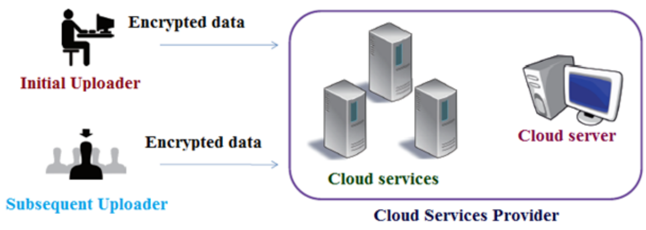

Fig. 1. Architecture of a data deduplication system

we describe the data deduplication architecture and define the security model. According to the granularity of deduplication, deduplication schemes are categorized into (coarse-grained) file-level or (fine-grained) block-level schemes. Since block-level deduplication can easily be deduced from file-level deduplication, we consider only file-level deduplication for simplicity's sake. Thus, a data copy refers to a whole file in this paper.

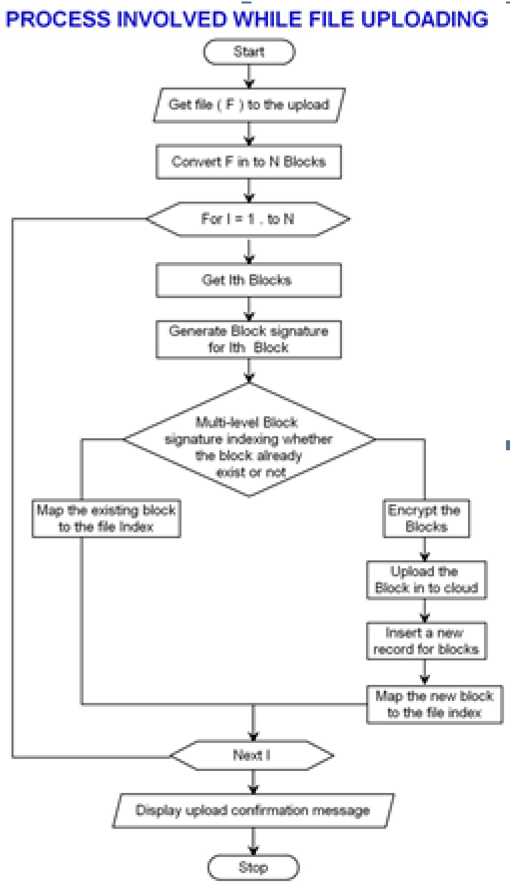
PROCESS INVOLVED WHILE FILE UPLOADING
Start
Get file (F) to the upload
Convert F in to N Blocks
For I = 1 , to N
Get Ith Blocks
Generate Block signature for Ith Block
Multi-level Block signature indexing whether the block already exist or not
Map the existing block to the file Index
Encrypt the Blocks
Upload the Block in to cloud
Insert a new record for blocks
Map the new block to the file index
Next I
Display upload confirmation message
Stop

Fig.2. Flow Chart for Upload Process

7. RESULT

The accompanying depictions layout the outcomes or yields that we are going to get once regulated execution of the considerable number of modules of the framework.

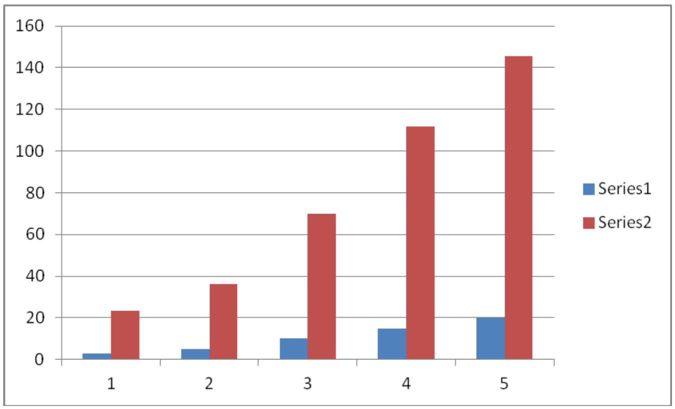

Fig. 3. **Upload Process Result**

While uploading the file, shows in figure 3, first step is break the file in small blocks based on given block size after that hash code get generated for all blocks, while generating hash code it will check whether it is new block of data or duplicate block of data based on hash code if hash code matched with existing hash code means it is duplicate block of data and if it is not matching means it is new data, all new block of data we will encrypt using AES encryption then we will upload to the cloud drive. As graph showing the result if file size is less it will take less time to upload and if file size is big it will take more time to execute.

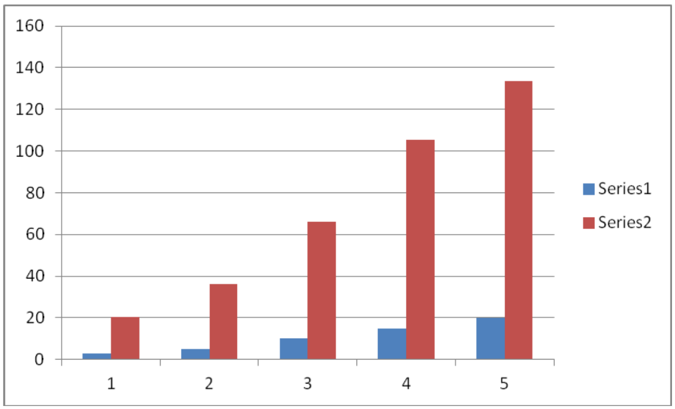

Fig.4. **Download Process Result**

While downloading the file, shows in figure 4, first it will check how many blocks is there, after that it will start downloading that that block from cloud drive. While downloading blocks from cloud drive it will decrypt block content and after downloading the all blocks it will merge all block, to make a single file. So if file size is less it will take less time to download and file size is big it will take more time to download.

8. CONCLUSION

Dynamic possession administration is a critical and testing issue in secure deduplication over scrambled information in distributed storage. In this investigation, we proposed a novel secure information deduplication plan to improve a fine-grained proprietorship administration by abusing the normal for the cloud information administration framework. In this manner, the proposed plot improves information protection and privacy in distributed storage against any clients who don't have substantial responsibility for information, and in addition against a genuine however inquisitive cloud server. Label consistency is likewise ensured, while the plan enables full favorable position to be taken of proficient information deduplication over encoded information. As far as the correspondence cost, the proposed plot is more effective than the past plans, while as far as the calculation cost, taking extra 0:1 and 0:2 ms contrasted with the RCE conspire, which is irrelevant practically speaking. In this manner, the proposed plot accomplishes more secure and fine-grained proprietorship administration in distributed storage for secure and effective information deduplication.

Future Enhancement

In future enhancement we can add to upload many more file big data etc, if file size is very big(big data) that also we can use in this application, we can improve performance by reducing the time while uploading the file.

REFERENCES

[1] D. T. Meyer, and W. J. Bolosky, “A study of practical deduplication,” Proc. USENIX Conference on File and Storage Technologies, 2011.

[2] M. Dutch, “Understanding data deduplication ratios,” SNIA Data Management Forum, 2008.

[3] W. K. Ng, W. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage,” Proc. ACM SAC’12, 2012.

[4] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” Proc. StorageSS’08, 2008.

[5] N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, “Reconciling end-to-end confidentiality and data reduction in cloud storage,” Proc. ACM Workshop on Cloud Computing Security, pp. 21–32, 2014.

[6] P. S. S. Council, “PCI SSC data security standards overview,” 2013.

[7] D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Side channels in cloud services, the case of deduplication in cloud storage,” IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.

[8] C. Wang, Z. Qin, J. Peng, and J. Wang, “A novel encryption scheme for data deduplication system,” Proc. International Conference on Communications, Circuits and Ssytems (ICCCAS), pp. 265–269, 2010.

[9] Malicious insider attacks to rise, http://news.bbc.co.uk/2/hi/7875904.stm

[10] Data theft linked to ex-employees, http://www.theaustralian.com.au/australian-it/datatheftlinked- to-ex-employees/story-e6frgakx-1226572351953,2002.

[11] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, “Reclaiming space from duplicate files in a serverless distributed file system,” Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.

[12] P. Anderson, L. Zhang, “Fast and secure laptop backups with encrypted de-duplication,” Proc. USENIX LISA, 2010.

[13] Z. Wilcox-O’Hearn, B. Warner, “Tahoe: the least-authority filesystem,” Proc. ACM StorageSS, 2008.

[14] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, J. C. S. Lui, “A secure cloud backup system with assured deletion and version control,” Proc. International Workshop on Security in Cloud Computing, 2011.

[15] J. Xu, E. Chang, and J. Zhou, “Leakage-resilient client-side deduplication of encrypted data in cloud storage,” ePrint, IACR, http://eprint.iacr.org/2011/538.

[16] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure deduplication,” Proc. Eurocrypt 2013, LNCS 7881, pp. 296–312, 2013.Cryptology ePrint Archive, Report 2012/631, 2012.

[17] S. Halevi, D, Harnik, B. Pinkas, and A. Shulman-Peleg, “Proofs of ownership in remote storage systems,” Proc. ACM Conference on Computer and Communications Security, pp. 491–500, 2011.

[18] M. Mulazzani, S. Schrittwieser, M. Leithner, and M. Huber, “Dark clouds on the horizon: using cloud storage as attack vector and online slack space,” Proc. USENIX Conference on Security, 2011.

[19] A. Juels, and B. S. Kaliski, “PORs: Proofs of retrievability for large files,” Proc. ACM Conference on Computer and Communications Security, pp. 584–597, 2007.

[20] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable data possession at untrusted stores,” Proc. ACM Conference on Computer and Communications Security, pp. 598–609, 2007.

[21] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, “Secure deduplication with efficient and reliable convergent key management,” IEEE Transactions on Parallel and Distributed Sytems, Vol. 25, No. 6, 2014.

[22] G.R. Blakley, and C. Meadows, “Security of Ramp schemes,” Proc. CRYPTO 1985, pp. 242–268, 1985.

[23] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, “A hybrid cloud approach for secure authorized deduplication,” IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.

[24] M. Bellare, S. Keelveedhi, T. Ristenpart, “DupLESS: Serveraided encryption for deduplicated storage,” Proc. USENIX Security Symposium, 2013.

[25] M. Bellare, S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," Proc. PKC 2015, pp. 516–538, 2015.

[26] Y. Shin and K. Kim, "Equality predicate encryption for secure data deduplication," Proc. Conference on Information Security and Cryptology (CISC-W), pp. 64–70, 2012.

[27] X. Jin, L. Wei, M. Yu, N. Yu and J. Sun, "Anonymous deduplication of encrypted data with proof of ownership in cloud storage," Proc. IEEE Conf. Communications in China (ICCC), pp.224-229, 2013.

[28] D. Naor, M. Naor, and J. Lotspiech, "Revocation and tracing schemes for stateless receivers," Proc. CRYPTO 2001, Lecture Notes in Computer Science, vol. 2139, pp. 41–62, 2001.