Efficient method of Secure Authorized Data Deduplication at Small Block Level Approach in Cloud Storage

Harshita Rai[#1], Dr. Ajay Kumar Bharti[*2], Dr. Santosh Kumar[$3]

[1]School of Computer Science, Maharishi University of Information Technology, Lucknow, India.
[2]Professor, School of Computer Science, Maharishi University of Information Technology, Lucknow, India
[3]Associate Professor, School of Computer Science, Maharishi University of Information Technology, Lucknow, India

***Abstract*—In this paper, proposed a new serverside deduplication scheme for mixed data. In this it is applying data deduplication on small block level and if data block are unique data in that case first it will get encrypted then uploaded to cloud storage. Moreover, the proposed system guarantees data uprightness against any name anomaly attack. In this way, security is enhanced in the proposed system. For deduplication hash code comparison in proposed system introduced a new technique called authentication of multi-level block signature. This technique provides a mechanism which is used to reduce comparison time of hash code in data base. The adequacy examination comes to fruition show that the proposed contrive is for all intents and purposes as capable as the existing system, while the additional computational overhead is unimportant.**

***Keyword:* *Deduplication, Ownership, Cloud Storage,* multi-level block signature**

1. INTRODUCTION

Distributed storage is one of the computerized stockpiling arrangements that uses various servers (ordinarily spread over different areas) which guarantees to securely store documents, for example, site reinforcements and so forth. The information is put away on devoted servers, which gives boundless openness wherever a web association is accessible, alongside an expansion in reinforcement document security which guarantees our records not to be hacked.

At the point when a client transfers information that as of now exist in the distributed storage, the client ought to be deflected from getting to the information that were put away before he got the proprietorship by transferring it (in reverse secrecy).Existing framework are not taking care of information deduplication at little square dimension. These dynamic proprietorship changes may happen as often as possible in a functional cloud framework, and hence, it ought to be appropriately overseen with the end goal to evade the security corruption of the cloud benefit. In the previous methodology, the majority of the current plans have been proposed with the end goal to play out a POW procedure in a productive and hearty way, since the hash of the document, which is treated as a "proof" for the whole record, is powerless against being spilled to outside enemies in light 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 infers different proprietors may have transferred similar information already, he is known as a resulting uploader.

This proposed approved copy check conspire brings about insignificant overhead contrasted with ordinary tasks. In Proposed framework propose another server-side deduplication plan for blended information. It engages the cloud server to control access to re-appropriated information regardless of when the possession changes seriously by abusing randomized joined encryption and secure proprietorship pack key diffusing, a deduplication benefit over encoded information.

The proposed framework guarantees that particular embraced access to the basic information is conceivable, which is accepted to be the most essential test for able and secure scattered bigdata stockpiling benefits in the earth where information possession changes strongly. The proposed framework guarantees security in the setting of dynamic information possession by displaying a hash key structure for dynamic proprietorship gathering.

Some of its benefits are as under:-

- Higher Security. Your site reinforcements will be situated off-site and over different servers. This implies your reinforcement is better shielded from information misfortune or hacking than if it were put away on a neighbourhood server.

- Easy Access. As it's available on the web, you get to webpage reinforcements at whatever point it is required you require them, paying little heed to your area. This isn't conceivable with nearby capacity!

- No upkeep given that cloud servers are kept up by a different organization, you won't have to contract educated IT staff to keep up the server. This will spare you a large number of dollars over the long haul.

2. 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 suggested that File structures consistently contain redundant copies of information: undefined reports or sub-record areas, maybe set away on a single host, on a shared storing gathering, or moved down to discretionary limit. Deduplication accumulating structures misuse this abundance to diminish the crucial space anticipated that would contain the record systems (or fortification pictures thereof). Deduplication can work at either the sub-archive or whole record level. More fine-grained deduplication makes more open entryways for space save stores, yet in a general sense reduces the progressive configuration of a couple of records, which may have basic execution impacts when hard plates are used for limit (and on occasion requires befuddled procedures to upgrade execution.

M. Dutch[2] has expressed that concerning the understanding of data deduplication extents that data deduplication cuts down business threats, constructs wage openings, and reduces accumulating level costs, realizing a perfect whirlwind for associations sending a flexible amassing establishment. Limit adaptability propels, for instance, RAID or RAIN, shield the deduplicated data to ensure high availability of uses getting to the data. The money related parts of data deduplication make it more than persuading; it is required for any business hoping to grow their customer advantage levels. Data deduplication extents are definitely not hard to over-separate and credit favorable circumstances to, that could possibly exist.

W. K. Ng et al. [3] has proposed about another thought which we call private data deduplication tradition, a deduplication strategy for private data amassing is exhibited and formalized. Naturally, a private data deduplication tradition allows a client who holds a private data exhibits to a server who holds a framework string of the data that he/she is the proprietor of that data without revealing extra information to the server.

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Mill operator [4] has proposed about the Businesses and purchasers are winding up detectably continuously mindful of the estimation of secure, chronicled data storing. In the business field, data defending is routinely directed by law, and data mining has wound up being a guide in framing business framework. For individuals, recorded limit is being called upon to spare nostalgic and real antiquated rarities, for instance, photos, movies and individual chronicles. Further, while few would battle that business data calls for security, assurance is comparably basic for individuals; data, for instance, therapeutic records and definitive reports must be kept for drawn out extends of time however ought not be transparently accessible. Unfathomably, the growing estimation of true data is driving the necessity for cost-profitable limit; sensible limit allows the preservation of all data that may over the long haul show supportive.

3. Problem Statements

Now days everybody using cloud services, storing data, sharing data with others and people knows that cloud is a third party resource so many concern are there like data security, privacy, data ownership, space for storage, bandwidth, data deduplication etc. Major issues are data security,, data deduplication, dynamic data ownership which is not as per our expectation, we need more enhancement in these concern. As we analyze many existing work in the field of data security, dynamic data ownership. Many existing plan using deduplication but they are not addressing deduplication at small block level.

4. Contributions

We propose a data deduplication scheme over dynamically distributed data in cloud storage. The proposed system ensures that individual secure access to the common data is possible in cloud storage, which is believed to be the most basic test for successful and secure appropriated bigdata storage organizations in nature where ownership changes effectively. In existing system deduplication checked on file level and no security for data but the proposed system introduced the small block level deduplication, enhance data security and provide dynamic data ownership in cloud storage.

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 discouraged from getting to the information that were put away before he acquired the proprietorship by transferring it (in reverse secrecy). Existing framework are not dealing with information deduplication at little piece level. These dynamic proprietorship changes may happen oftentimes in a down to earth cloud framework, and in this manner, it ought to be legitimately overseen keeping in mind the end goal to dodge the security debasement of the cloud benefit. In the previous approach, the majority of the current plans have been proposed with a specific end goal to play out a POW procedure in a proficient and powerful way, since the hash of the document, which is dealt with as a "proof" for the whole record, is defenseless against being spilled to outside enemies in light 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 an ensuing uploader since this infers different proprietors may have transferred similar information already, he is known as a consequent uploader.

5.2. Proposed System

Our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations. In Proposed system we propose a new server-side deduplication plan for mixed data. It empowers the cloud server to control

access to outsourced data despite when the ownership changes intensely by manhandling randomized joined encryption and secure ownership pack key scattering, a deduplication service over encoded data.

The proposed system ensures that selective endorsed access to the common data is possible, which is believed to be the most basic test for capable and secure dispersed bigdata storage benefits in the earth where data ownership changes intensely. The proposed system ensures security in the setting of dynamic data ownership by exhibiting a hash key framework for dynamic ownership gathering.

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.2. Flow Chart for Upload Process

While uploading the file, 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.

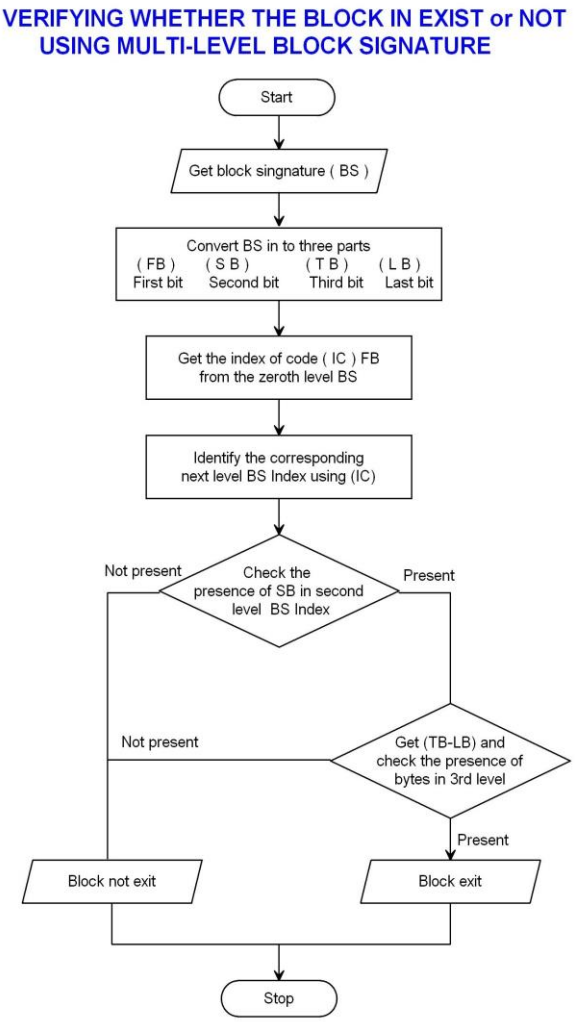

Fig.3. Flow Chart for Multi-Level Block Signature

Proposed system is providing security to the data using AES encryption as mention in uploading file flow chart Figure 2. For deduplication detection in small block level it is using concept of Multi-level block signature which improving performance of our proposed system shown in figure 3.

While hash code comparison in proposed system introduced a new technique called authentication of multi-level block signature. This technique provides a mechanism which is used to reduce comparison time of hash code in data base. Here it breaks hash code in three part, first part having two digit, second part two digit and last part having remaining digit, while comparison first it will check two digit if it match then it will check another two bit if it also match then it will compare remaining bit.

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. While uploading the file, 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 will encrypt using AES encryption then it will upload to the cloud drive. While hash code comparison in proposed system introduced a new technique called authentication of multi-level block signature. This technique provides a mechanism which is used to reduce comparison time of hash code in data base. Here it breaks hash code in three part, first part having two digit, second part two digit and last part having remaining digit, while comparison first it will check two digit if it match then it will check another two bit if it also match then it will compare remaining bit.

Table 1. Time taken to upload with concept of multi-level block signature

File Name	File Size (KB)	Time Taken to Upload (Sec.)
File-1.txt	3	15.584
File-2.txt	5	24.172
File-3.txt	10	55.285
File-4.txt	15	80.472
File-5.txt	20	115.947

While downloading the file, 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.

8. CONCLUSIONn

Distributed dynamic data ownership services is a basic and testing issue in secure deduplication over mixed data in dispersed capacity. In this proposed system, we proposed a new secure data deduplication intend to enhance procedure for the cloud data organization structure. In this way, the proposed system enhances data assurance and security in conveyed capacity for every customer who doesn't have access regarding data. Name consistency is in like manner guaranteed, while the arrangement empowers full ideal position to be taken of capable data deduplication over encoded data. To the extent the correspondence cost, the proposed system is more suitable than the existing system, while hash tag comparison our approach taking less time compare to existing system. In this way, the proposed system achieves more secure and well performing approach in dynamic storage for secure and powerful data 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.