iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33534

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

A Result Review Analysis of on Distributed De-Duplication System using File Level and Block Level

Miss. Sneha Lande[1], Dr. M. M. Bartere[2]
[1, 2]*G.H. Raisoni University, Amravati, India*

Abstract: ***As the world moves to digital storage for archival purposes, there is an increasing demand for systems that can provide secure data storage in a cost-effective manner. Now days with the huge increasing of population and the using of technology, it leads to many problems. The growth in technology is increasing the amount of storage or communication and technique devices. By the unpredictable development of digital data, DE-duplication techniques are broadly engaged to backup data and decrease network and storage transparency by notice and eradicate redundancy among data.***
As an alternative of maintaining multiple data copies with the same content, DE duplication reducing redundant data by maintaining only single copy and referring other redundant data to that copy. By identifying common chunks of data both within and between files and storing them only once, DE duplication can yield cost savings by increasing the utility of a given amount of storage. DE duplication has inward much concentration from both academic world and industry since it can really recover storage utilization and keep storage space, particularly for the applications with high DE duplication ratio such as archival storage systems.

I. INTRODUCTION

The authenticity of many legal, financial, and other documents is determined by the presence or absence of an authorized handwritten signature. The recipient of the signed document can verify the claimed identity of the sender using the signature. Also, if the sender later repudiates the contents of the document, then recipient can use the signature to prove the validity of the document. With the computerized message systems replacing the physical transport of paper and ink documents, an effective solution for authentication of the electronic data is necessary. Various methods have been devised to solve this problem, but the use of `digital signature' is definitely the best solution amongst them.

A digital signature is nothing but an attachment to any piece of electronic information, which represents the content of the document and the identity of the originator of that document uniquely. The digital signature is intended for use in electronic mail, electronic funds transfer, electronic data interchange, software distribution, data storage, and other applications which require data integrity assurance and data origin authentication.

Digital signatures may also be generated for stored data and programs so that the integrity of the data and programs may be verified at any later time. Although there are various approaches to implement the digital signature, this report discusses the `Digital Signature Standard'. It specifies the Digital Signature Algorithm (DSA) which is appropriate for applications requiring a digital rather than written signature.

The DSA is considered as the standard procedure to generate and verify digital signatures. Authentication of any system means providing a security to that system.

One of the authentication techniques is a textual password based, but this can be easily broken. To overcome this drawback we need to secure the text so this can be done by using digital signature. By using Image authentication technique, the text will be encrypted in an image a secured with digital signature. And then transmit over the network. Due to this the text is unable to break and hence data will be transmitted securely.

The private key us used in the signature generation process and must remain secret to prevent other non-identifiable entities from using it to generate fraudulent signatures. There are also algorithms that are in place to prevent the private key falling into another's hands or another person who is aware of the private key from using it to sign a different message. As such these digital signatures cannot be forged.

On the other hand, the public key is used in the verification process. While it need not remain a secret, the integrity of the public key must be maintained. In this sense, anyone with the public key can verify the signed message using the public key. An approved hash function is used to convert the signed message to a fixed-length representation of the message. The verifier requires assurances that the public key to be used belongs to the signer and that the originator of the document also owns the private key.

II. BACKGROUND AND RELATED WORK

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

It introduced the idea of security and scheme for symmetric encryption in concentrate security framework. They give different idea of security and analyze the good involution of reduction among them. They provide method of encryption using a block cipher, cipher block chaining and counter mode.

Its have two goals .First is to study the idea of security for symmetrical encryption and second is to provide concrete security analysis of fixed symmetric encryption device.

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proc. Of StorageSS, 2008.

They developed a solution that provides both data security and space efficiency in single-server storage and distributed storage systems to solve the problem such that deduplication exploits identical content, while encryption tries to make all content appear random ,the same content encrypted with two different keys results in very different cipher text. Deduplication and encryption are opposed to one another. Deduplication takes benefit of data similarity to achieve a reduction in storage space & the goal of cryptography is to make cipher text indistinguishable from theoretically random data

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

They presents an algorithm which takes benefits of the data which is common between users to reduce the storage requirements, and increase the speed of backups.

This algorithm supports clientend per-user encryption which is important for confidential personal data, also supports a unique feature that allows immediate detection of common sub trees, avoiding the necessity to query the backup system for every file. This system has shown that a community of laptop users shares a considerable amount of data in between. This gives the potential to significantly decrease backup times and storage requirements. However, they have shown that manual selection of the relevant data -eg, backing up only home directories is a poor strategy; this become fails to take backup of important files, at the same time as unnecessarily duplicating other files.

III. PROPOSED METHODOLOGY

To protect private data the secret sharing technique is used which is corresponding to distributed storage systems. In this paper the secret sharing technique is used for protection of private data. In detail a file is divides and encode into sections by using secret sharing technique. These sections will be distributed over many independent storage servers. A cryptanalysis hash value of the content will also be calculated and send to storage server as the mark of the fragment stored at each server. only the data user who first upload the data is required to calculate and distribute such secret shares and following users own same data copy do not need to calculate and stores these shares. Retrieve data copies owner must access a minimum number of storage server by a validation and obtain the secret shares to alter the data. In different way, the authorized uses will access the secret shares data copy.

A. *Proposed Objectives*

1) To authenticate a data and to make available integrity and validity assurances on the data.
2) To implement File-level Distributed Deduplication System.
3) To implement Block-level Distributed Deduplication System.
4) To maintain the consistency and integrity of data within file.

B. *De-Duplication Technique*

1) Input: Any real time database of file transfer.
2) Output:
a) Finding File Level Authentication
b) Finding Block Level Authentication
c) Group of Algorithm Analysis

C. Proposed Techniques

File-Level Distributed De-duplication System

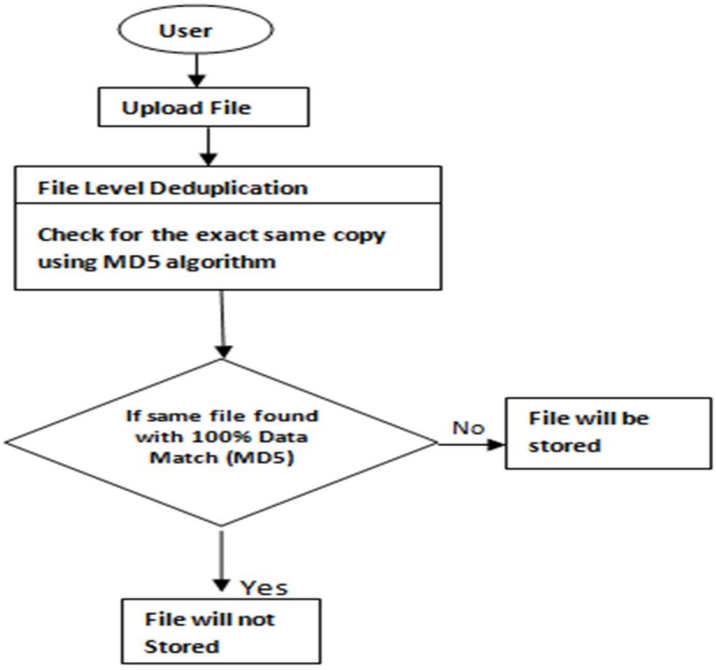


Fig1: Flowchart Working Process of File Level deduplcation

It support capable duplicate check, tags for each file will be calculated and send to storage cloud service provider. To prevent alignment invasion organized by the cloud based service provider, tag collected at different storage servers. System Setup: In our structure, the storage cloud service provider is considered to be n with identities denoted by id1, id2,…,idn respectively. To upload file F, the client communicate with cloud based service provider to perform the elimination of duplicate data .For downloading file F, the client downloads the secret shares of the file from k out of storage servers.

Block-Level Deduplication System

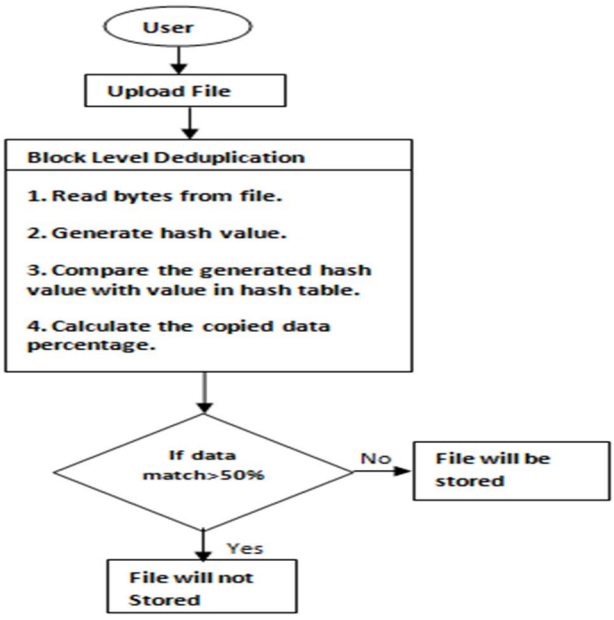


Fig1: Flowchart Working Process of Block Level deduplcation

In this part, we appear how to derive the fine grained block level distributed deduplication.In this system, the client also demands to perform the file level deduplication before uploading file. The user partition this files into blocks, if no duplication is found and performs block-level deduplication system. The system set up is similar to file-level deduplication and also block size parameter will be defined.

IV. RESULT ANALYSIS

This part focuses on result and its analysis based on the tools used for this method. The tools help in analyzing the user and the admin analysis graphs which help both user and the admin to know both techniques file level and block level are good. To evaluate the behavior of the file deduplcation the parameters like no of file transfer is used.

An example to show the analysis is given below.

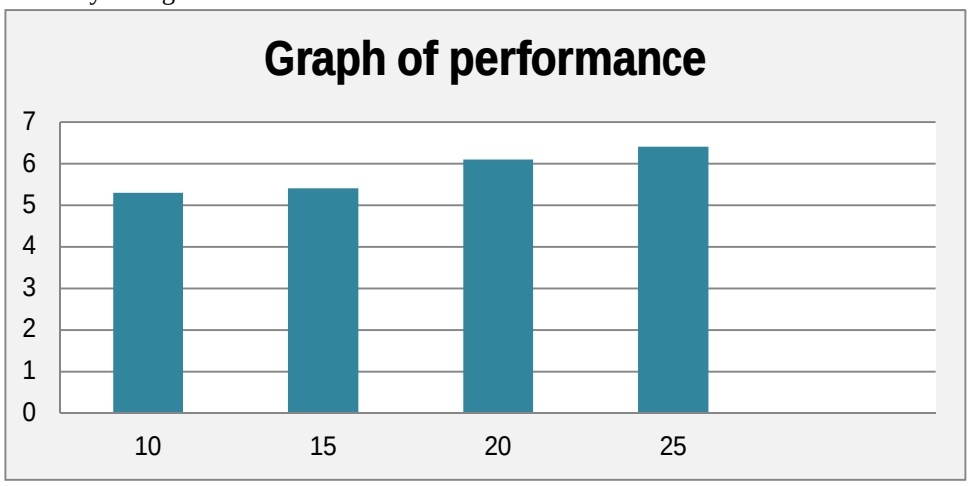


Fig 4: Result of Analysis

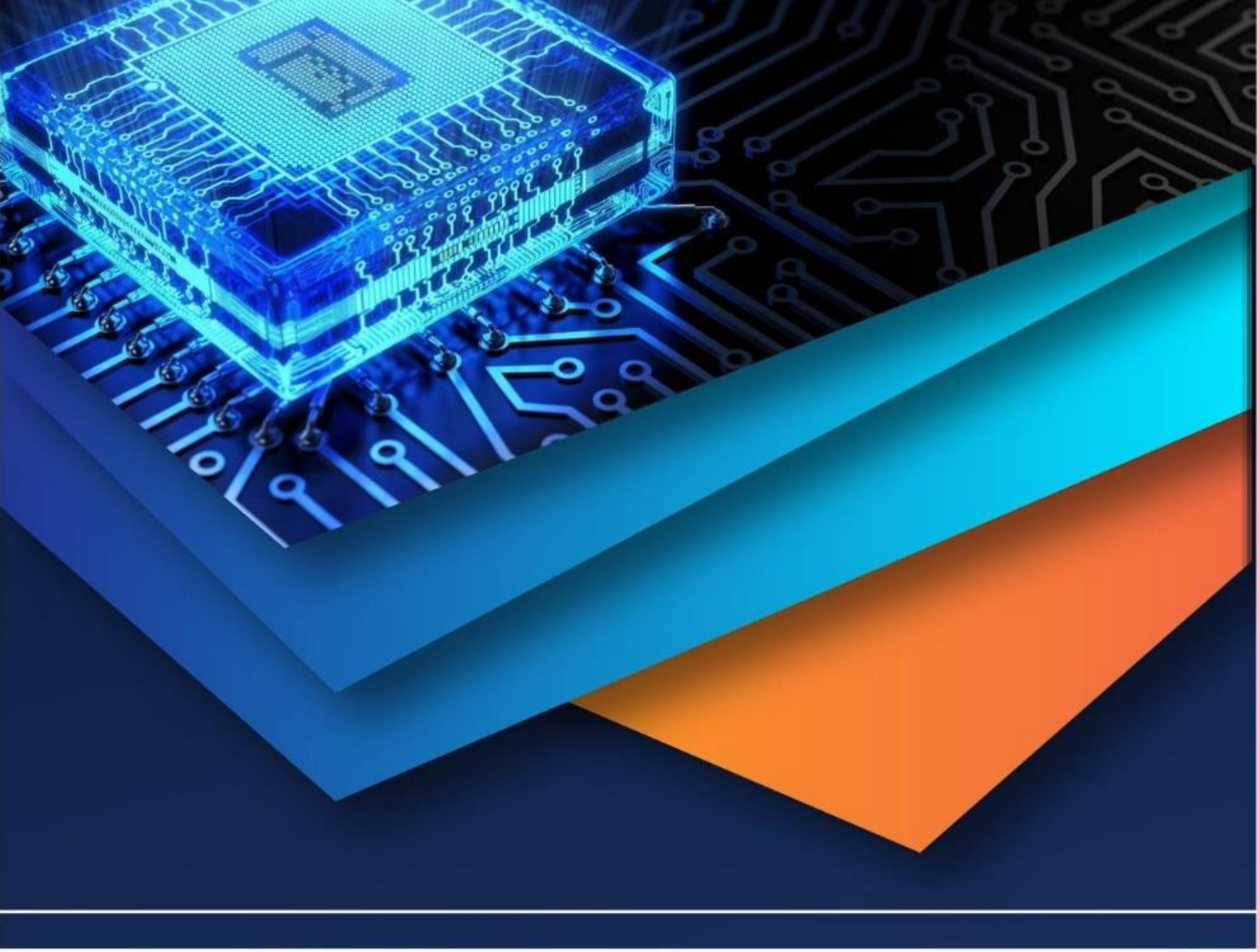
In the above graph the x axis indicates the number of user files in database. The y axis indicated the total computational time in ms to perform encryption operation using De-Duplication technique.

V. CONCLUSION

We proposed the distributed de-duplication systems to improve the reliability of data while achieving the confidentiality of the users outsourced data without an encryption mechanism. We proposed the file level de-duplication, block level de-duplication, tag generation and message authentication code technique. The security of tag consistency and integrity will be achieved.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Dupless: Serveraided encryption for deduplicated storage,” in USENIX Security Symposium, 2013
[2] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” in Proc. Of StorageSS, 2008.
[3] P. Anderson and L. Zhang, “Fast and secure laptop backups with encrypted de- duplication,” in Proc. of USENIX LISA, 2010
[4] J. S. Plank, S. Simmerman, and C. D. Schuman, “Jerasure: A library in C/C++ facilitating erasure coding for storage applications - Version 1.2,” University of Tennessee, Tech. Rep. CS-08-627, August 2008.
[5] J. S. Plank and L. Xu, “Optimizing Cauchy Reed-solomon Codes for fault-tolerant network storage applications,” in NCA-06: 5th IEEE International Symposium on Network Computing Applications, Cambridge, MA, July 2006.
[6] C. Liu, Y. Gu, L. Sun, B. Yan, and D. Wang, “R-admad: High reliability provision for large-scale de-duplication archival storage systems,” in Proceedings of the 23rd international conference on Supercomputing, pp. 370–379.
[7] 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.
[8] P. Anderson and L. Zhang, “Fast and secure laptop backups with encrypted de-duplication,” in Proc. of USENIX LISA, 2010.
[9] Z. Wilcox-O’Hearn and B. Warner, “Tahoe: the least-authority filesystem,” in Proc. of ACM StorageSS, 2008.
[10] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, “A secure cloud backup system with assured deletion and version control,” in 3rd International Workshop on Security in Cloud Computing, 2011.
[11] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” in Proc. of StorageSS, 2008.
[12] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, “A secure data deduplication scheme for cloud storage,” in Technical Report, 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY