iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.48603

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: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Privacy Preserving Secure Deduplication in Cloud Storage

Nitin Kumar K[1], Praveen R[2], Leander Rithicun L[3], Prasanth Kumar M[4], Mr. Praveen Kumar M[5]
[1, 2, 3, 4]*UG scholars,* [5]*Assistant Professor, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Tamil Nadu, India.*

Abstract: ***Among the main parts of cloud computing is distributed storage, which makes it much easier for cloud clients to store and share their information in the cloud with authorized users. Secure deduplication has received a lot of attention in distributed storage since it can lessen correspondence overhead and storage space by eliminating redundancy in encrypted data.In terms of privacy and security, numerous current secure deduplication plans typically concentrate on accomplishing the accompanying properties:confidentialityof the data, consistency of the tags, command over access, and protection from animal power assaults. However, to the extent that we are aware, not a single one of them can all the while satisfy these four necessities.To resolve this issue, we present in this paper a client-definable, secure, and effective deduplication strategy.Particularly, our plan maximizes the elimination of duplicates without compromising cloud users' privacy or security by permitting just the cloud specialist co-op to approve information access for proprietors. Our authorized secure deduplication plot, according to a comprehensive security analysis, prevents brute-force attacks while maintaining data confidentiality and tag consistency.In addition, broad reproductions exhibit that our strategy outflanks the contending ones as far as the efficiency of deduplication furthermore, the overheads related with calculation, correspondence,and storage.***

List Terms: Secure deduplication, access control, Cryptography

I. INTRODUCTION

More and more data are being transferred to the cloud and imparted to approved clients because of the various advantages of distributed computing.For instance, the Cisco worldwide cloud list predicts that how much information put away in the cloud will almost arrive at 1.3 zettabytes by 2021. Subsequently,managing the always expanding measure of information presents a significant difficult for distributed storage facilities.

The information deduplication procedure has been generally created for use in distributed storage , in this scenario because it can onlystore one duplicate of repetitive information. In fact, the study demonstrates that storage for backups and archives systems have considerably more redundancy almost 90% of digital data.In fact, in backup applications and standard file systems information deduplication can diminish capacity costs by over 90% and over half,respectively.Due to these reserve funds, cloud specialist co-ops and users can save a lot of money.

However, due to concerns regarding privacy and security, clients most probably going to encode data along with their keys preceding outsourcing.Deduplication of data is hindered because indistinguishable information would be scrambled into particular ciphertexts.Data deduplication on encoded information can now be accomplished thanks to the development of convergent encryption and its variants.However, convergent encryption faces difficulties due to brute-force attacks on predictable messages.As a possible solution to this issue, server-helped encryption plans have been proposed.The copy faking assault, which forestalls authorized clients from getting precise information,unfortunately affects them.The foe produces the ciphertext and tag from unmistakable information, m, separately, in greater detail.Users verify the ciphertext's coherence with the tag after downloading, which precisely determine no matter the inaccurate information is the result of copy faking assaults during information transfer or is adulterated during information stockpiling,despite the fact that some schemes attempt to oppose this assault.

Once the inconsistent ciphertext and tag have been stored by the cloud service provider, resulting clients who transfer the tag that compares to m can get the inaccurate information form.

Because the ciphertext and comparing tag are produced freely the cloud specialist organization is unfit to confirm the consistency of the relating tag.An answer that straight forwardly figures the tag is introduced by looking at the ciphertext's hash to the received tag. This clearly supports the tag consistency check.In light of this idea, message authentication and tag consistency have been taken into consideration. Users would perform message authentication after downloading data, and the cloud specialist organization would actually look at label consistency.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

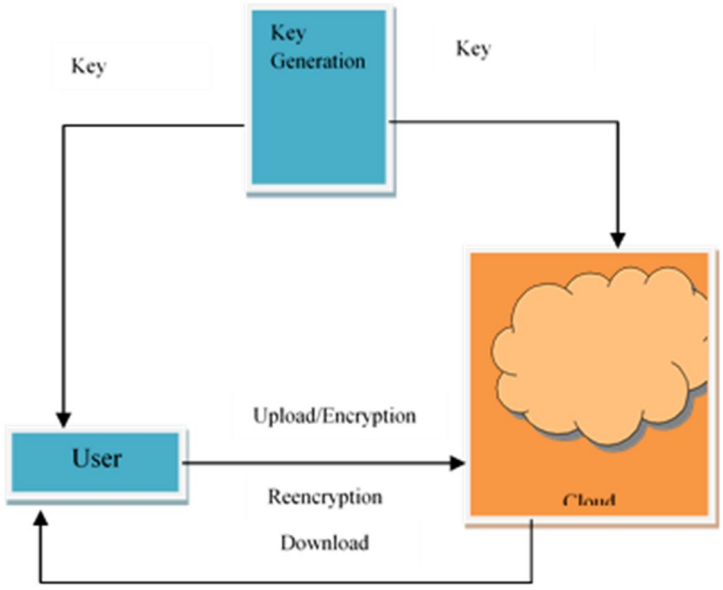


Fig. 1: System model

II. SYSTEM MODEL

Exemplifies the system's architecture, with the following descriptions of each entity's specifics:

1) *CSP*: Users can get storage services from the CSP.The CSP would like to decrease the above related with information limit by wiping out excess information and keeping just a single duplicate in the approved deduplication framework without abusing access management.
2) *Users:* Users can be authorized users or data owners in our approved deduplication framework.The client who re-appropriates scrambled information to the CSP and shares these rethought information with approved clients is typically referred to as the data owner.In general, the term "authorized user" alludes to a client who has permission from the information proprietors to get to their reevaluated information
3) *KGC*: The entire system's initial setup is the responsibility of the KGC.System parameters, the cloud service supplier's mystery key, and each client's public/confidential key pair are all generated by the KGC.After that, except if another client joins the framework, the KGC can remain offline.Keep in mind that the KGC just has to produce the newcomer's public/confidential key pair.
4) *Recncrypt*: When the cloud server-containing encrypted file is forwarded to CSP.The encrypted file is then sent to a cloud server, where a symmetric encryption calculation is utilized to re-encode the specified file.Likewise, store encrypted data on the server.

III. RELATED WORK

The research community has extensively developed a secure deduplication method that can dispose of excess information while maintaining information classification.The studies that are related can be broken down into the accompanying three categories:access control, label consistency, and information secrecy.

A. Confidentiality of Data

Douceur et al. ensure the classification of information byFirst demonstrated in [6], joined encryption, another cryptographic crude known as message-locked encryption, was formalized by Bellare et al. (12) with a semantic security evidence for unusual messages and the ability to deduplicate encrypted data.Following that, numerous implementations and variants of message-locked encryption, which is also referred to as convergent encryption, were utilized in [8, 10]. However, message-locked encryption is defenseless against brute-force attacks for messages that are predictable.Accordingly, information classification would be compromised on the grounds that the foe would have the option to figure out which information match a certain cipher text. To be able to to protect against brute-force attacks, a few servers helped protected deduplication plans have been suggested [7, 10]. In these schemes, every client connects having one or a predetermined amount of extra important servers to get the focalized key.

Although these plans penance one of the effectiveness of duplication [8, 9] or the effectiveness of calculation, correspondence, and capacity [10].The secure cross-client duplication conspire, which does not require any additional independent servers and offers high levels of security, was utilized by Liu and co. [3].In any case, all clients are mentioned to be continually online in order to assist uploaders in obtaining the merged key produced by the client who has previously transferred a similar file. This is effective in the peer to peer model, but it is less effective cloud [6].

B. Control of Access

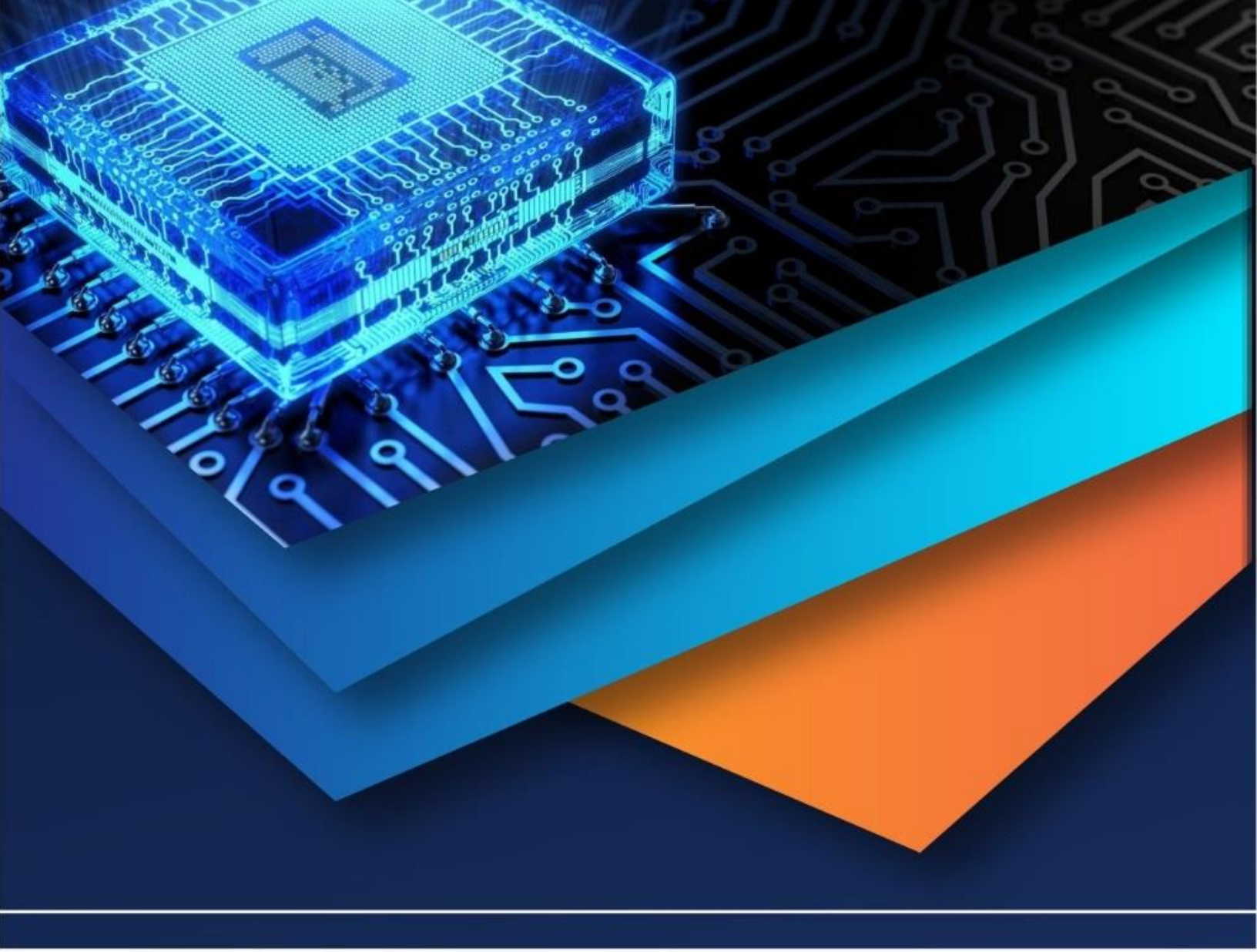
As described in [3], data owners typically share their encrypted data with authorized users and prefer to outsource it to cloud service providers. A few existing plans have consolidated secure duplication with access control. Li and other, For instance, [4] considered the various client honors in real-time copy review in addition to the actual data, in which the copy can be discarded only when the client honors of records are coordinated. Enabled authorized secure duplication by utilizing cipher text-policy attribute-based encryption. Additionally, Yan et al. used the linear secret sharing technique to achieve access control for secure duplication [6]. The proxy re-encryption mechanism with CP-ABE (PRE) were used in's scheme [9] to provide secure deduplication and access control [3]. However, to the best of our knowledge, these methods are unable to simultaneously accomplish label consistency, opposition to brute-force attacks, and efficiency.

IV. CONCLUSION

This study examines propose a user-defined, secure, and efficient duplication strategy. In particular, to accomplish approved duplication, our plan does not require the utilization of hybrid cloud architecture or the addition of an additional authorized server. Without jeopardizing data confidentiality, only the CSP can manage access rights under our plan for information proprietors. In addition, the Bloom filter is included in our plan to speed up the duplicate check. Our strategy is capable of simultaneously achievinginformation privacy, regulating access, label accuracy, and security against brute-force assaults, as demonstrated by in-depth security analyses.Also, broad execution assessments on document level and lump level deduplication demonstrate that our strategy is effective regarding deduplication productivity, computational expense, correspondence above, and capacity cost

REFERENCES

[1] “Cisco global cloud index: Forecast and methodology, 2016-2021 white paper,”https://www.cisco.com/c/en/us/solutions /collatera/service-provider/global-cloud-index-ci/white-paper-c11-738085.html.
[2] J. Gantz and D. Reinsel, “The digital universe decade-are you ready,” https: //hk.emc.com/collateral/analyst-reports/idc-digital-universeare-you-ready. pdf.
[3] H. Biggar, “Experiencing data de-duplication: Improving efficiency and reducing capacity requirements,” The Enterprise Strategy Group., 2007.
[4] D. T. Meyer and W. J. Bolosky, “A study of practical deduplication,” TOS, vol. 7, no. 4, pp. 14:1–14:20, 2012.
[5] D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Side channels in cloud services: Deduplication in cloud storage,” IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010
[6] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, “Re- claiming space from duplicate files in a serverless distributed file system,” in ICDCS, 2002, pp. 617–624.
[7] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, “A secure data deduplication scheme for cloud storage,” in Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers, 2014, pp. 99–118.
[8] P. Puzio, R. Molva, M. O¨ nen, and S. Loureiro, “Cloudedup: Secure dedu- plication with encrypted data for cloud storage,” in IEEE 5th International Conference on Cloud Computing Technology and Science, CloudCom 2013, Bristol, United Kingdom, December 2-5, 2013, Volume 1, 2013, pp. 363–370.
[9] S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Server-aided encryp- tion for deduplicated storage,” in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013, 2013, pp. 179–194.
[10] M. Miao, J. Wang, H. Li, and X. Chen, “Secure multi-server-aided data dedu- plication in cloud computing,” Pervasive and Mobile Computing, vol. 24, pp. 129–137, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY