SECURE ENCRYPTED DATA DEDUPLICATION WITH DYNAMIC OWNERSHIP UPDATING

[1]Kashoju Rama, [2.]Swetha.G

[1] M.Tech student ,Department of Computer Science and Engineering,TEEGALA KRISHNA REDDY ENGINEERING COLLEGE, Hyderabad, India

Email:manoharsunkoju120@gmail.com

[2] Assistant Professor,Department of Computer Science and Engineering,TEEGALA KRISHNA REDDY ENGINEERING COLLEGE, Hyderabad, India

Email:swethareddy630@gmail.com

Abstract:

Data grows at the impressive rate of 50% per year, and 75% of the digital world is a copy. Although keeping Copies, multiple copies of data is necessary to guarantee their availability and long term durability, in many situations the amount of data redundancy is immoderate. By keeping a single copy of repeated data, data deduplication is considered as one of the most promising solutions to reduce the storage costs, and improve users experience by saving network bandwidth and reducing backup time. However, this solution must now solve many security issues to be completely satisfying. In this paper we target the attacks from malicious clients that are based on the manipulation of data identifiers and those based on backup time and network traffic observation. We present a deduplication scheme mixing an intra and an inter-user deduplication in order to build a storage system that is secure against the aforementioned type of attacks by controlling the correspondence between files and their identifiers, and making the inter-user deduplication unnoticeable to clients using deduplication proxies and also Dynamic ownership to every uploaded file and the deleted files. Our method provides global storage space savings, per-client bandwidth network savings between clients and deduplication proxies, and global network bandwidth savings between deduplication proxies and the storage server. The evaluation of our solution compared to a classic system shows that the overhead introduced by our scheme is mostly due to data encryption which is necessary to ensure confidentiality.

I.INTRODUCTION

As data continues to proliferate, the need to efficiently manage redundancy while safeguarding against security threats becomes paramount. Traditional deduplication methods offer cost savings and bandwidth optimization but fall short in addressing evolving security concerns. Our paper proposes a novel deduplication scheme that integrates intra- and inter-user

deduplication to fortify against malicious attacks targeting data identifiers and backup processes. By meticulously managing file-identifier correspondence and concealing deduplication processes from clients, our approach establishes a secure storage framework. Through evaluation, we demonstrate substantial storage and network bandwidth savings, reaffirming our method's efficacy in upholding data integrity and privacy in today's digital landscape.The problem statement addresses the escalating challenges of data redundancy and security vulnerabilities in deduplication systems amidst exponential data growth. It highlights the necessity to curb excessive storage costs, optimize network bandwidth, and mitigate risks posed by malicious attacks targeting data identifiers and backup processes. The proposed solution aims to combine intra-user and inter-user deduplication methods to create a secure storage architecture that ensures confidentiality and data integrity. The key objectives include controlling file-identifier correspondence, concealing deduplication processes from clients, and delivering significant storage and network bandwidth savings. The study evaluates the effectiveness of the proposed scheme against conventional systems, emphasizing the overhead primarily attributed to data encryption for confidentiality assurance. Through analysis and experimentation, the research seeks to validate the feasibility and efficacy of the proposed solution in addressing the identified challenges and enhancing overall system performance and security.

II.LITERATURE SURVEY

Research on data deduplication and cloud storage security has explored both the efficiency gains and the risks involved in optimizing modern storage infrastructures. The study “Understanding Data Deduplication Ratios” highlights the critical role of deduplication in reducing storage consumption as part of efficient information lifecycle management, emphasizing how deduplication ratios influence capacity optimization. Complementing this, “A Study of Practical Deduplication” provides empirical insights from file system data collected at Microsoft, showing that whole-file deduplication achieves nearly three-quarters of the savings of block-level methods for live systems and 87% for backups, while also noting that large unstructured files dominate storage needs. From a security perspective, “Proofs of

Ownership in Remote Storage Systems" addresses vulnerabilities in client-side deduplication, where attackers can exploit hash-based proofs to illegitimately access others' files. The authors propose a secure proof-of-ownership (PoW) mechanism using Merkle trees and encoding techniques, which mitigates attacks with minimal performance overhead. Similarly, "Side Channels in Cloud Services: Deduplication in Cloud Storage" examines the privacy risks of cross-user deduplication, warning that while it is highly effective, it can unintentionally expose user data, though simple mechanisms can balance efficiency with reduced leakage risks. Finally, "Dark Clouds on the Horizon" investigates Dropbox as a case study, revealing weaknesses in its client software and protocol that allow exploitation for hiding files, storing copyright-protected content, and creating online slack space with virtually unlimited storage. Collectively, these studies underscore that while deduplication and cloud storage bring significant efficiency and scalability benefits, they also introduce new attack vectors and privacy challenges, demanding secure mechanisms to balance performance with trust.

III.EXISTING SYSTEM

The existing system encompasses current practices in data deduplication within cloud storage systems. It revolves around techniques aimed at reducing storage costs by eliminating duplicate data copies and optimizing storage resources. Trusted third party to achieve the deduplication on encrypted data is proposed. In this situation, the trusted third party has authority to access all the plaintext data, and all users compute duplicate check tags with the assistance of the trusted third party. These third party solutions are not also safe. As there is scope of data leakage through the service provider.

IV.PROPOSED SYSTEM

The Proposed solution implements a System where the user can upload the encrypted data to the cloud without any third party services.It aims to mitigate the risk of information leakage through manipulation of data identifiers, network traffic observation, and backup time analysis. The proposed system seeks to enhance confidentiality and security in data deduplication processes.particularly in inter-user and client-side scenarios, which offer substantial savings in network bandwidth and storage space. Additionally, the proposed system aims to provide a comprehensive solution capable of addressing various types of attacks that malicious users may attempt on the deduplication system simultaneously.

SecDedup is independent of any online third party, it provides mechanisms for dynamic data ownership update. Furthermore, it is suitable for both file and block oriented storage systems ,in which files and blocks are treated as the basic deduplication object units respectively. Encrypted data deduplication technology allows eliminating redundant encrypted data stored in the cloud environment, such as websites hosted in public clouds, datasets maintained by Internet service providers or comprehensive cloud storage service providers

V.SYSTEM ARCHITECTURE

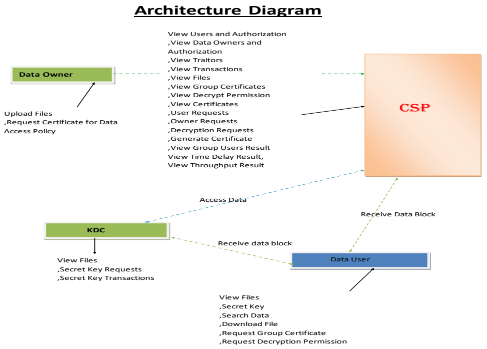
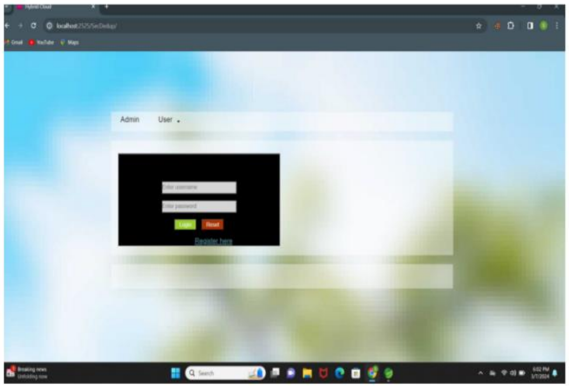


Fig 5.1 System Architecture

The system architecture is designed to ensure secure storage, controlled access, and efficient management of data in the cloud. The Data Owner is responsible for uploading files to the cloud and requesting certificates for enforcing data access policies. Once data is uploaded, it is managed by the Cloud Service Provider (CSP), which handles multiple operations such as maintaining user authorizations, processing requests from both data owners and users, generating certificates, granting or denying decryption permissions, and monitoring system performance through metrics like time delay and throughput results. To enhance security, a Key Distribution Center (KDC) is incorporated into the system. The KDC manages secret key requests and transactions, ensuring that only authorized users receive the necessary keys to access encrypted files. The Data User interacts with the CSP and KDC to perform tasks such as viewing files, searching data, downloading files, requesting group certificates, and obtaining decryption permissions. Certificates and secret keys act as the backbone of the access control mechanism, ensuring that unauthorized users cannot gain access. Together, this architecture establishes a secure, certificate-based, and key-managed framework that balances usability with strong data confidentiality, integrity, and controlled access in cloud storage environments.

VI.IMPLEMENTATION



localhost:2525/SecDedup/

Admin User

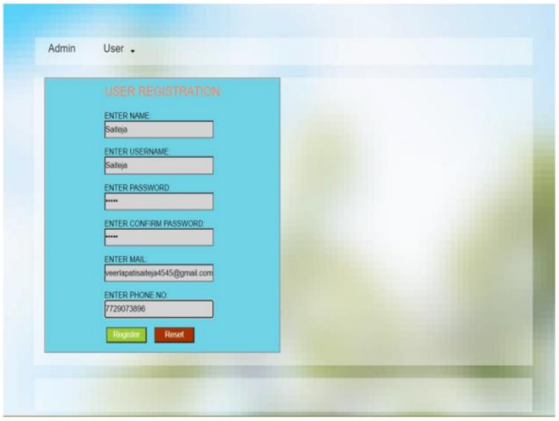
Enter username

Enter password

Login Reset

Register here

Fig 6.1 Admin Login



Admin User

USER REGISTRATION

ENTER NAME: Saiteja

ENTER USERNAME: Saiteja

ENTER PASSWORD: ••••

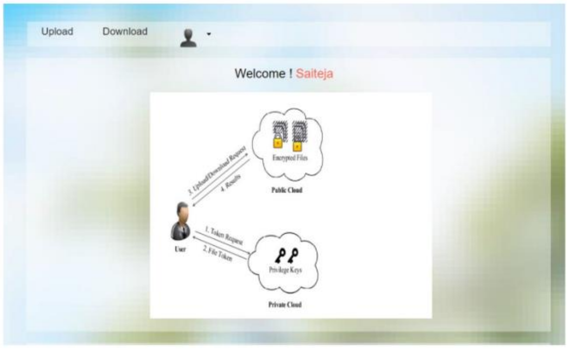
ENTER CONFIRM PASSWORD: ••••

ENTER MAIL: veerlapatisaiteja4545@gmail.com

ENTER PHONE NO: 7729073896

Register Reset

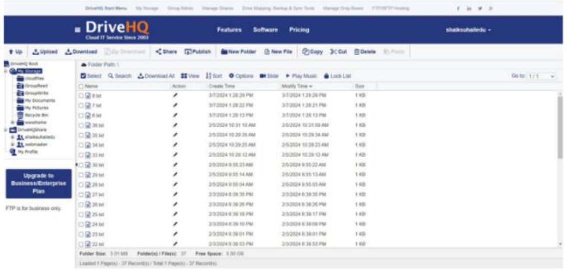
Fig 6.2 User Register



Upload Download

Welcome ! Saiteja

Fig 6.3 Main Page



Name	Action	Create Time	Modify Time	Size
8.txt		3/7/2024 1:28:29 PM	3/7/2024 1:28:29 PM	1 KB
7.txt		3/7/2024 1:28:22 PM	3/7/2024 1:28:21 PM	1 KB
6.txt		3/7/2024 1:28:13 PM	3/7/2024 1:28:13 PM	1 KB
36.txt		2/5/2024 10:31:10 AM	2/5/2024 10:31:09 AM	1 KB
35.txt		2/5/2024 10:29:35 AM	2/5/2024 10:29:34 AM	1 KB
34.txt		2/5/2024 10:29:25 AM	2/5/2024 10:29:23 AM	1 KB
33.txt		2/5/2024 10:29:12 AM	2/5/2024 10:29:12 AM	1 KB
30.txt		2/5/2024 9:55:23 AM	2/5/2024 9:55:22 AM	1 KB
29.txt		2/5/2024 9:55:14 AM	2/5/2024 9:55:13 AM	1 KB
28.txt		2/5/2024 9:55:04 AM	2/5/2024 9:55:03 AM	1 KB
27.txt		2/3/2024 6:39:35 PM	2/3/2024 6:39:35 PM	1 KB
26.txt		2/3/2024 6:39:26 PM	2/3/2024 6:39:26 PM	1 KB
25.txt		2/3/2024 6:39:18 PM	2/3/2024 6:39:17 PM	1 KB
24.txt		2/3/2024 6:39:10 PM	2/3/2024 6:39:09 PM	1 KB
23.txt		2/3/2024 6:39:01 PM	2/3/2024 6:39:01 PM	1 KB
22.txt		2/3/2024 6:38:53 PM	2/3/2024 6:38:53 PM	1 KB

Folder Size: 3.01 MB Folder(s) / File(s): 37 Free Space: 5.00 GB

Loaded 1 Page(s) - 37 Record(s) / Total 1 Page(s) - 37 Record(s)

Fig 6.4 DriveHq Page

VII.CONCLUSION

In this paper, we propose a deduplication scheme for encrypted data, named SecDedup. It does not relay on any online trusted third party, and it allows dynamic ownership updating. We use cryptographic primitives such as proxy re-encryption and bilinear mapping instead of convergent encryption. Compared with previous schemes, the security of SecDedup is significantly enhanced. In our design, the CSP can work as an intermediary, which focuses on three core functionalities, in particular, 1) updating the ownership list when a user updates or deletes his data, 2) assisting an initial uploader to validate subsequent uploaders, and 3) delivering data encryption keys in an offline manner. The optimized CSP in SecDedup can easily and effectively perform deduplication on encrypted data. Meanwhile, it is enforced that a user is allowed to access the data only if he can provide a valid access license. Simulation experiments show that our scheme is applicable and efficient.

VIII.FUTURE SCOPE

The future scope of the proposed SecDedup deduplication scheme opens up several potential areas for further research, development, and implementation. Some future directions and considerations include:

1.Real-world Implementation and Deployment: While simulation experiments demonstrate applicability, the next step could involve real-world implementation and deployment of SecDedup in practical settings to assess its performance, scalability, and usability.

2.Scalability Testing: Evaluate the scalability of SecDedup under various conditions, including increasing data volumes, user base, and concurrent operations. This can help identify potential bottlenecks and optimize the scheme for large-scale usage.

3.Usability and User Experience Studies: Conduct user studies to assess the usability and user experience of SecDedup. This could involve gathering feedback from users and administrators to identify any challenges, improvements, or additional features that could enhance the overall system.

4.Performance Optimization: Continuously optimize the performance of the SecDedup scheme, focusing on reducing computational overhead, improving response times, and enhancing overall efficiency. This may involve refining cryptographic algorithms, data structures, or implementation strategies.

5.Security Analysis and Updates: Stay vigilant in monitoring emerging security threats and advancements in cryptographic techniques. Regularly update SecDedup to address any vulnerabilities, ensuring that it remains resilient against evolving security challenges.

IX.REFERENCES

[1]X. Yang, R. Lu, J. Shao, X. Tang, and A. Ghorbani, ‘‘Achieving efficient and privacy-preserving multi-domain big data deduplication in cloud,’’ IEEE Trans. Services Comput., early access, Nov. 13, 2018, doi: 10.1109/ TSC.2018.2881147.

[2]M. Bellare, S. Keelveedhi, and T. Ristenpart, ‘‘Message-locked encryption and secure deduplication,’’ in Advances in Cryptology Eurocrypt. Berlin, Germany: Springer, 2013, pp. 296–312

. [3] M. Bellare, S. Keelveedhi, and T. Ristenpart, ‘‘DupLESS: Server-aided encryption for deduplicated storage,’’ in Proc. 22nd Usenix Conf. Secur., Usenix Assoc., 2013, pp. 179–194.

[4]P. Puzio, R. Molva, M. Onen, and S. Loureiro, ‘‘ClouDedup: Secure deduplication with encrypted data for cloud storage,’’ in Proc. IEEE 5th Int. Conf. Cloud Comput. Technol. Sci., Dec. 2013, pp. 363–

370.

[5]J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, A Secure Data Deduplication Scheme for Cloud Storage. New York, NY, USA: Ibm Corporation, 2014, pp. 99–118.

[6]P. Puzio, R. Molva, M. Önen, and S. Loureiro, ‘‘PerfectDedup: Secure data deduplication,’’ in Proc. 10th Int. Workshop Data Privacy Manage., Secur. Assurance (DPM), 4th Int. Workshop QASA, in Lecture Notes in Computer Science, vol. 9481. Vienna, Austria: Springer, Sep. 2015, pp. 150–166.

[7]C. Hui, R. H. Deng, Y. Li, and G. Wu, ‘‘Attribute-based storage supporting secure deduplication of encrypted data in cloud,’’ IEEE Trans. Big Data, vol. 5, no. 3, pp. 330–342, Sep. 2016.

[8]X. Ge, J. Yu, H. Zhang, C. Hu, Z. Li, Z. Qin, and R. Hao, ‘‘Towards achieving keyword search over dynamic encrypted cloud data with symmetrickey based verification,’’ IEEE Trans. Depend. Sec. Comput., early access, Jan. 30, 2019, doi: 10.1109/TDSC.2019.2896258

[9]B. Libert and D. Vergnaud, ‘‘Unidirectional chosen-ciphertext secure proxy
re-encryption,’’ IEEE Trans. Inf. Theory, vol. 57, no. 3, pp. 1786–1802, Mar. 2011.