WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Big Data Dynamic Ownership Management with Secure Data Deduplication in Cloud Storage

Sumit Yadav[1], Divyansh Yadav[2], Er. Shadab Ali[3]

[1]UG student of Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Uttar Pradesh, Lucknow divyanshyadav737373@gmail.com

[2]UG student of Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Uttar Pradesh, Lucknow sumitya611@gmail.com

[3]Associate Professor of Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Uttar Pradesh, Lucknow shadab@srmcem.ac.in

Abstract:

As the exponential growth of data in the digital world has given rise to cloud storage as the solution of choice for cost-efficient and scalable data management, big data in the cloud presents massive challenges related to secure storage, tracking data ownership, and handling redundant data. This paper proposes a new framework for dynamic ownership management with secure deduplication of data in cloud storage systems. Through the integration of cryptographic ownership transfer protocols and privacy-preserving deduplication mechanisms, the model ensures data integrity, confidentiality, and effective storage utilization. Thorough analysis and simulation prove the effectiveness and security of the system in real-world big data scenarios

Big Data, Cloud Storage, Secure Deduplication, Ownership Management, Cryptography, Dynamic Access Control.[4].

Introduction

The data explosion has reshaped how organizations handle storing and managing data. Cloud storage provides cost-effectiveness and elasticity but, at the same time, poses questions on ownership of data, privacy, and duplication of data accumulation. Deduplication, the method used to remove duplicate data blocks, conserves storage space but becomes dangerous when used with encrypted data and ownership change.

This work suggests an end-to-end secure, scalable framework for managing dynamically ownership rights of data in the cloud store such that security and privacy aren't compromised through deduplication. [2].

Distributed computing provides flexible, low-effort, and space-free online services ranging from simple reinforcement services to distributed storage platforms. The rapid growth of information amounts stored in the distributed storage has led to an increased demand for mechanisms for saving disk space and network transfer rate. In order to reduce asset usage, most distributed storage administrations such as Dropbox, Wuala, Mozy, and Google Drive make use of a deduplication technique whereby the cloud server retains only one duplicate of redundant data and provides links to the duplicate rather than keeping other actual duplicates of such data regardless of the number of customers make a request to store data.[1].

Literature Review

Practical Deduplication Studies

D.T. Meyer and W.J. Bolosky [1] noted that file systems tend to have redundant copies of the data, e.g.,

same files or sub-file areas saved across systems or backup storage. Redundancy is eliminated by deduplication systems to cut the amount of storage needed. It is done at the file level or sub-file, with lower granularity providing more savings but maybe slowing down due to fragmenting layouts, particularly on conventional hard disks.

Understanding Deduplication Ratios

M. Dutch [2] noted that efficient deduplication decreases business risks, diminishes storage expense, and offers revenue possibilities. Technologies such as RAID or RAIN are implemented to maintain data availability and integrity in deduplicated environments and accommodate scalable and trustworthy storage infrastructures.

Architecture of System

Initial Uploader: A user sends a file to the cloud for the very first time.

The file is encrypted by the user before sending, to provide privacy and security.

The encrypted file is sent to the Cloud Services Provider, where it gets stored.

Cloud Services Provider**: This includes:**

Cloud services: They are the units of storage and processing that have the responsibility to handle data.

Cloud server: Serves as the primary controller that deduplicates, stores, and may execute verification processes.

Subsequent Uploader: When the received encrypted file is from the initial uploader, the system stores it.

A different user attempts to upload the same file (logically similar data).

This user also encrypts the data before sending.

The system verifies whether the uploaded content (after going through some sort of secure verification or deduplication process) already exists.

Deduplication Process: If the file already exists (based on secure deduplication logic), the cloud server does not store the file again.

Instead, it associates the new upload request with the old data.

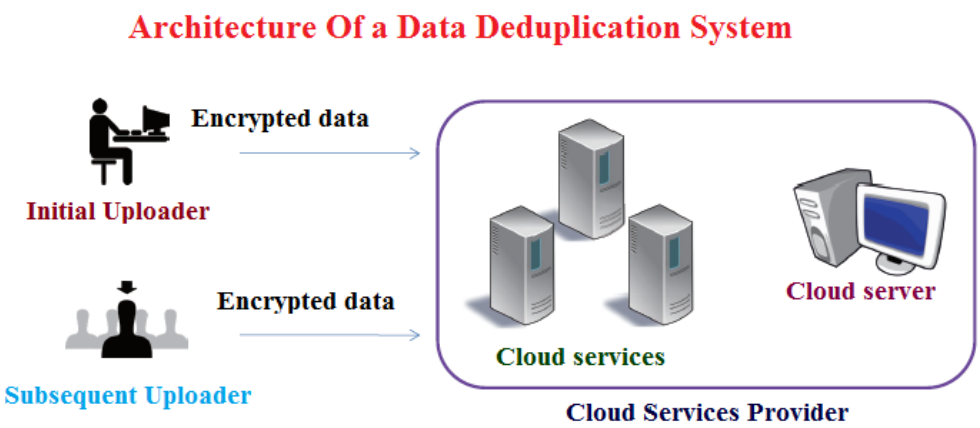
This prevents redundancy, conserving storage space and bandwidth, even if many users upload the same file.

Key Concept

Data Privacy**:** Files are encrypted prior to upload.

Storage Efficiency**:** Duplicate copies of the same information are not stored once.

Ownership Management: The system controls who uploaded what and associates many users with the same stored data securely.



Conclusion:

The presented system provides a safe and effective way of data deduplication in cloud storage for addressing the issue of dynamic data ownership. It ensures that encrypted data are deduplicated without compromising security or ownership control. Naming convention consistency is maintained while improving data assurance and limiting access restrictions for unauthorized parties. The system exhibits better performance and lower communication costs compared to current methodologies. By utilizing hash tag comparison, it drastically reduces processing time. Main benefits are increased storage efficiency, cost reduction, and increased data security. It also provides support for scalability, hence suitability in Big Data environments.

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 discuss "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 introduce the notion of "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," Proceedings of the 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 have recently proposed a new data encryption scheme. deduplication system, Proc. International Conference on Communications, Circuits and Ssytems (ICCCAS), pp. 265–269, 2010.

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

[10]. Data theft by former employees, http://www.theaustralian.com.au/australian\0002it/datatheftattributed-to-former-employees/story-e6frgakx-1226572351953,2002.

[11]. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer introduce the idea of "Reclaiming." "Space from duplicate files in a serverless distributed file system," Proceedings of the International Conference on Distributed Computing Systems (ICDCS), pages 617 to 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]. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui present their paper titled "A secure cloud." backup system with assured deletion and version control," Proc. International Workshop on Cloud Computing Security, 2011.