IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Enhancing Efficiency And Security In Joint Cloud Storage Through Data Deduplication

C.Vikram[1], Mr.G.Lokesh[2]

[1]PGstudent, vemu institute of technology, P.Kothakota

[2]AssistantProfessor,vemu institute of technology,P.Kothakota

ABSTRACT

Data deduplication in cloud storage is essential for reducing redundancies and bandwidth requirements. However, existing schemes reliant on a trusted key server (KS) suffer from vulnerabilities like information leakage and single-point-of-failure. To address these issues, we propose SED, a Secure and Efficient data Deduplication scheme, integrated into a Joint Cloud storage system. SED facilitates global services across multiple clouds, supports dynamic data updates, and mitigates single-point-of-failure risks. Theoretical analyses demonstrate SED's semantic security and robust anti-attack capabilities. With low computational complexity and overhead, SED effectively eliminates data redundancies, enhancing client-side usability. Comparative evaluations attest to SED's superior performance over existing schemes.

Keywords: Deduplication, cloud

INTRODUCTION:

Cloud storage is a revolutionary platform that offers expansive data storage and service access on a "pay-as-you-go" basis. Despite its convenience, the abundance of redundant data within cloud storage systems has led to a significant waste of storage resources. To address this issue, data deduplication has emerged as a highly effective technology for identifying and eliminating redundant data. By storing only a single copy of each unique piece of data, data deduplication minimizes storage requirements and enhances space utilization efficiency, benefiting both service providers and users alike. This technology has become increasingly prevalent across various cloud computing services, playing a vital role in enhancing user experience and optimizing storage space allocation.

Traditionally, data deduplication schemes have relied on a centralized framework comprising a key

server (KS), a cloud storage provider (CSP), and users. However, these classic schemes are susceptible to single-point-of-failure and "platform lock-in" issues, wherein the failure of the trusted KS can disrupt cloud storage operations and impede data outsourcing protocols. In response to these challenges, Joint Cloud computing systems have emerged as a promising alternative. Unlike traditional models, Joint Cloud systems feature a decentralized network architecture that fosters collaboration among multiple CSPs, eliminating the need for a central KS. This collaborative approach enables efficient cross-cloud services and ensures the resilience of cloud storage systems against potential failures.

In recent years, the adoption of Joint Cloud computing has gained traction within both academia and industry, driven by the pressing need to address data security concerns in cloud storage environments. With the proliferation of massive data breaches affecting billions of personal records, the encryption of outsourced data has become a paramount concern. However, traditional encryption methods pose challenges for encrypted deduplication, as ciphertexts generated by different users encrypting the same plaintext are inherently distinct.

To overcome these challenges, convergent encryption (CE) and its variants have been proposed as solutions for encrypted deduplication. These schemes leverage deterministic secret keys derived from the data itself to facilitate deduplication while ensuring data confidentiality. However, existing CE schemes are plagued by security vulnerabilities, including susceptibility to chosen-plaintext attacks, semantic security violations, and computational burdens on users.

To address these shortcomings, we propose a Secure and Efficient data Deduplication (SED) scheme based on the Joint Cloud system model. Our SED scheme aims to achieve security, functionality, and efficiency in data outsourcing and deduplication processes. By eliminating the need for a trusted KS, our scheme ensures data confidentiality, integrity, and protection against attacks. Furthermore, our SED scheme supports data renewal, sharing, and access control across multiple CSPs, enhancing functionality while minimizing computational overhead. Through our innovative approach, we strive to enhance the security and efficiency of data deduplication in cloud storage environments, paving the way for safer and more streamlined data management practices.

LITERATURE SURVEY:

Peter Christen et al

Record linkage, also known as deduplication within a single database, is vital for matching records across databases. Matched data are increasingly valuable, offering unique insights not easily obtainable otherwise. Eliminating duplicates is crucial for accurate data processing and mining, as duplicates can skew outcomes significantly. However, with the growing size of databases, matching complexities pose challenges. Various indexing techniques have emerged to address this, aiming to reduce comparison loads while maintaining high matching accuracy. This paper surveys 12 variations of 6 indexing methods, analyzing their complexity and evaluating their

performance using synthetic and real datasets. Such a comprehensive survey on record linkage and deduplication techniques is unprecedented.

GangyongJia et al

In virtualization environments, both limited main memory size and memory interference pose significant challenges. Memory deduplication, consolidating identical content pages into single copies, alleviates memory requirements. Meanwhile, memory partitioning assigns unique colors to each virtual machine, minimizing interference and enhancing performance. This paper introduces CMDP (Coordinate Memory Deduplication and Partitioning), a novel approach tackling both issues concurrently. CMDP integrates BMD (Behaviour-based Memory Deduplication) for efficient page comparison and VMMP (Virtual Machine Memory Partitioning) to reduce interference. By allocating unique page colors, VMMP optimizes resource allocation. Experimental results demonstrate CMDP's efficacy, enhancing performance by approximately 15.8% while accommodating more virtual machines simultaneously. This innovative approach offers promising prospects for optimizing virtualization environments.

Wen Xia et al

Content-Defined Chunking has emerged as a vital component in data deduplication systems, owing to its exceptional redundancy detection capabilities. However, existing CDC-based methods often incur heavy CPU overhead due to their byte-by-byte computation and judgment of rolling hashes for chunk cut-points. Introducing Fast CDC, a swift and efficient Content-Defined Chunking approach, we leverage five key techniques. These include a gear-based fast rolling hash, optimized gear hash judgment, skip sub-minimum chunk cut-points, normalize chunk-size distribution, and rolling two bytes simultaneously. Evaluation results demonstrate FastCDC's superiority, achieving 3-12X faster speeds compared to state-of-the-art CDC approaches while maintaining comparable or higher deduplication ratios. Additionally, FastCDC significantly enhances deduplication throughput in projects like Destor, showcasing its practical efficacy in real-world scenarios.

Jingwei Li *et al*

As cloud computing advances, data outsourcing to cloud services gains popularity, offering relief from cumbersome data management. However, entrusting data to untrusted cloud storage raises security concerns, particularly regarding data integrity and deduplication. In this study, we address these issues by proposing two secure systems: SecCloud and SecCloud+. SecCloud employs an auditing entity within a MapReduce cloud, reducing user computation during data uploading and auditing. Meanwhile, SecCloud+ caters to users' encryption preferences, enabling integrity auditing and secure deduplication on encrypted data. These systems ensure both data integrity and deduplication in the cloud, enhancing security while accommodating user encryption needs. This research contributes to bolstering data security in cloud environments while streamlining user processes.

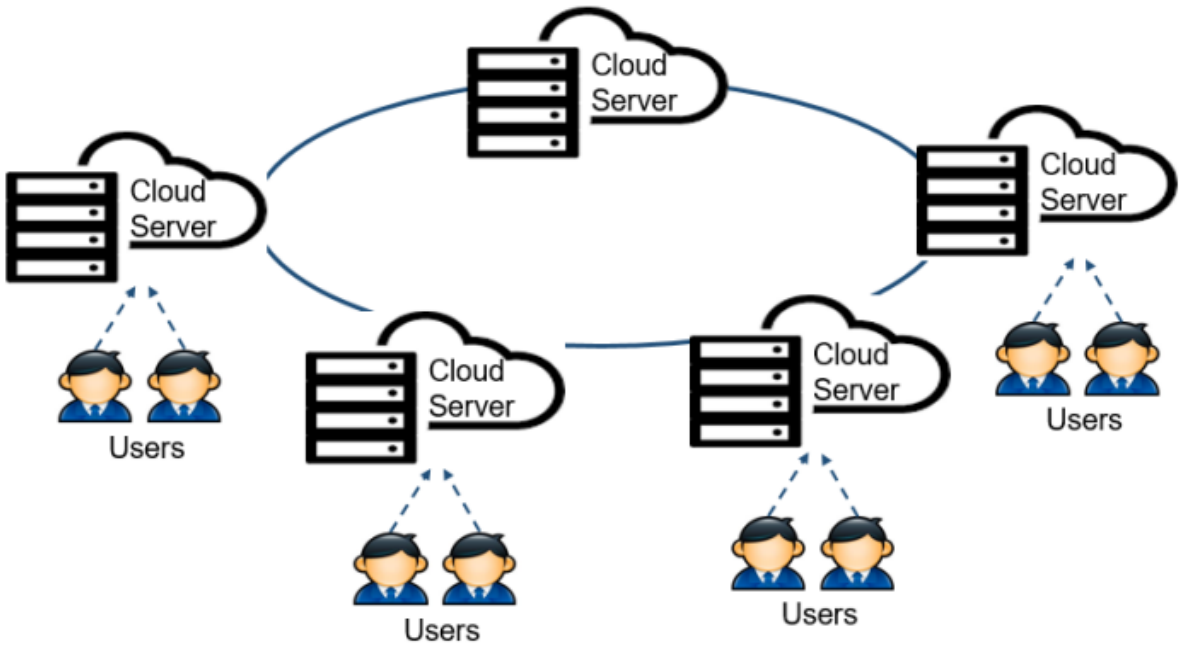
PROBLEM STATEMENT:

Data deduplication efficiently removes redundant data, optimizing storage and bandwidth in cloud computing. While widely used, encrypting data complicates deduplication as ciphertexts vary for the same plaintext, hindering duplicate detection. Traditional schemes rely on a trusted Key Server (KS) for security, risking single-point-of-failure and platform lock-in. If the KS fails, cloud storage systems halt, jeopardizing data protocols. To address these challenges, encrypted deduplication requires innovative approaches ensuring security without a trusted KS. This enhances data confidentiality and system reliability, vital in mitigating data breach risks and improving user experience in cloud storage services.

The Secure and Efficient Deduplication (SED) scheme enhances data security and efficiency by eliminating the need for a trusted Key Server (KS). SED reduces client-side overhead and improves scalability in Joint Cloud storage systems, addressing the single-point-of-failure issue. Its algorithms ensure semantic security and tag consistency, enhancing resistance against brute-force and collusion attacks. SED supports dynamic data operations, including sharing, deletion, and modification, improving usability. Notably, SED accommodates data sharing among permitted users, a unique feature. Theoretical and experimental analyses confirm SED's security and low complexity, outperforming previous schemes in security, efficiency, and functionality.

PROPOSED METHOD:

ARCHITECTURE:



METHODOLOGY:

Data Owner

The data owner uploads their data in the cloud server. For the security purpose the data owner encrypts the file and the index name and then store in the cloud. The data encryptor can have capable deleting of a specific file. And also he can view the transactions based on the files he uploaded to cloud.

Data User

User logs in by using his/her user name and password. After Login user requests search control to cloud and will Search for files based on the index keyword with the Score of the searched file and downloads the file. User can view the search ratio of the files and also the top k documents.

Cloud Server

The cloud server manages a cloud to provide data storage service. Data owners encrypt their data files and store them in the cloud for sharing with Remote User. To access the shared data files, data consumers download encrypted data files of their interest from the cloud and then decrypt them.

The cloud server authorizes the data owner and the data user and provides the search requests sent from the users. Also in this module it shows personalized search model and the interest search model. Can view all the file attackers.

RESULTS:

Files

ID	File Name	Data Owner	Date & Time
43	Android.txt	Arjun	23/10/2017 17:47:46
44	Angular_JS.txt	Arjun	23/10/2017 17:54:02
45	Bigdata.txt	Arjun	23/10/2017 17:54:36
46	PHP.txt	Arjun	23/10/2017 17:55:06
47	GST.txt	Manjunath	23/10/2017 18:30:31
48	Drmonetization.txt	Manjunath	23/10/2017 18:30:53
49	Social_Network.txt	Manjunath	23/10/2017 18:31:14
50	CAuth.jsp	Raviraj	16/11/2023 18:11:12

View files end user

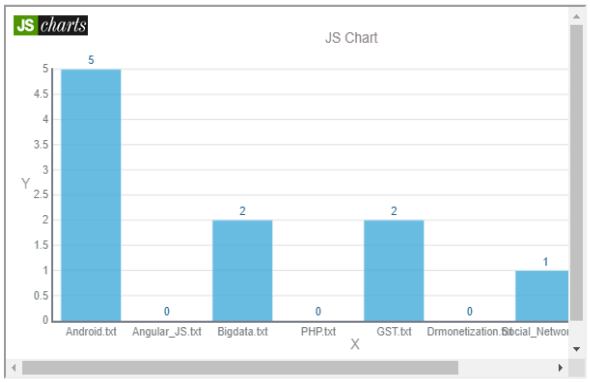
Request secret key

Transactions

ID	User Name	File Name	Task	Date & Time
79	Arjun	Android.txt	Upload	23/06/2021 17:47:46
80	Arjun	Angular_JS.txt	Upload	23/06/2021 17:54:02
81	Arjun	Bigdata.txt	Upload	23/06/2021 17:54:36
82	Arjun	PHP.txt	Upload	23/06/2021 17:55:06
83	Rajesh	Android	Search	23/06/2021 17:58:46
84	Rajesh	Android.txt	Download	23/06/2021 18:00:45
85	Rajesh	Android.txt	Download	23/06/2021 18:01:10
86	Rajesh	Android.txt	Download	23/06/2021 18:04:09
87	Manjunath	GST.txt	Upload	23/06/2021 18:30:31
88	Manjunath	Drmonetization.txt	Upload	23/06/2021 18:30:53
89	Manjunath	Social_Network.txt	Upload	23/06/2021

View file transactions

File Rank



File rank in chart

CONCLUSION

The Secure and Efficient Scheme (SED) for data deduplication, developed without the reliance on a trusted Key Server (KS), presents a significant advancement in cloud storage systems. SED effectively addresses challenges such as communication and computation overhead, scalability, and single-point-of-failure, while ensuring semantic security and tag consistency. Its robustness against common attacks and support for dynamic data operations enhance both security and usability. Through theoretical and experimental analyses, SED demonstrates superior performance in terms of security, efficiency, and functionality compared to previous schemes. This work marks a milestone by considering data sharing among permitted users, offering a comprehensive solution for modern cloud storage needs.

REFERENCES:

[1] P. Christen, "A survey of indexing techniques for scalable record linkage and deduplication," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 9, pp. 1537–1555, 2012.

[2] G. Jia, G. Han, J. J. P. C. Rodrigues, J. Lloret, and W. Li, "Coordinate memory deduplication and partition for improving performance in cloud computing," IEEE Transactions on Cloud Computing, vol. 7, no. 2, pp. 357–368, 2019.

[3] W. Xia, X. Zou, H. Jiang, Y. Zhou, C. Liu, D. Feng, Y. Hua, Y. Hu, and Y. Zhang, "The design of fast content-defined chunking for data deduplication based storage systems," IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 9, pp. 2017–2031, 2020.

[4] J. Li, J. Li, D. Xie, and Z. Cai, "Secure auditing and deduplicating data in cloud," IEEE Transactions on Computers, vol. 65, no. 8, pp. 2386–2396, 2016.

[5] L. Liu, Y. Zhang, and X. Li, "Keyd: Secure key-deduplication with identity-based broadcast encryption," IEEE Transactions on Cloud Computing, pp. 1–1, 2018.

[6] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. S. Shen, "Providing task allocation and secure deduplication for mobile crowdsensing via fog computing," IEEE Transactions on Dependable and Secure Computing, pp. 1–1, 2018.

[7] Y. Zheng, X. Yuan, X. Wang, J. Jiang, C. Wang, and X. Gui, "Toward encrypted cloud media center with secure deduplication," IEEE Transactions on Multimedia, vol. 19, no. 2, pp. 251–265, 2017.

[8] H. Wang, P. Shi, and Y. Zhang, "Jointcloud: A cross-cloud cooperation architecture for integrated internet service customization," in 2017 IEEE 37th International Conference on Distributed Computing Systems, 2017, pp. 1846–1855.

[9] K. Huang, X. Zhang, Y. Mu, F. Rezaeibagha, X. Wang, J. Li, Q. Xia, and J. Qin, "Eva: Efficient versatile auditing scheme for iot-based datamarket in jointcloud," IEEE Internet of Things Journal, vol. 7, no. 2, pp. 882–892, 2020.

[10] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in International Conference on the Theory and Applications of Cryptographic Techniques, 2013, pp. 296–312.

[11] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lock-dependent messages," in Advances in Cryptology – CRYPTO 2013. Springer Berlin Heidelberg, 2013, pp. 374–391.

[12] M. Bellare and S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," in Public-Key Cryptography – PKC 2015, J. Katz, Ed. Springer Berlin Heidelberg, 2015, pp. 516–538.

[13] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: serveraided encryption for deduplicated storage," in Usenix Conference on Security, 2013, pp. 179–194.

[14] T. Jiang, X. Chen, Q. Wu, J. Ma, W. Susilo, and W. Lou, "Secure and efficient cloud data deduplication with randomized tag," IEEE Transactions on Information Forensics and Security, vol. 12, no. 3, pp. 532–543, 2017.

[15] J. Li, X. Chen, M. Li, J. Li, P. P. C. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 6, pp. 1615–1625, 2014.

[16] X. Yang, R. Lu, J. Shao, X. Tang, and A. Ghorbani, "Achieving efficient secure deduplication with user-defined access control in cloud," IEEE Transactions on Dependable and Secure Computing, 2020.

[17] H. Yuan, X. Chen, J. Li, T. Jiang, J. Wang, and R. Deng, "Secure cloud data deduplication with efficient re-encryption," IEEE Transactions on Services Computing, pp. 1–1, 2019.

[18] Y. Shin, D. Koo, and J. Hur, "A survey of secure data deduplication schemes for cloud storage systems," ACM Computing Surveys, vol. 49, no. 4, pp. 74:1–74:38, 2017.

[19] J. Hur, D. Koo, Y. Shin, and K. Kang, "Secure data deduplication with dynamic ownership management in cloud storage," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 11, pp. 3113–3125, 2016.

[20] Y. Shin, D. Koo, J. Yun, and J. Hur, "Decentralized server-aided encryption for secure deduplication in cloud storage," IEEE Transactions on Services Computing, vol. 13, no. 6, pp. 1021–1033, 2020.