IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204271 |

Survey on Hashing based Deduplication System in Cloud Computing

Atharva Girish Kirve, Akshay Kaduba Khandebharad, Shrikant Gokul Patil, Prof. Poonam Jadhavar

Department of computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

ABSTRACT*:* Data deduplication, a technique for removing redundant data by storing only one duplicate of a file, saves a lot of storage and bandwidth for cloud storage providers. Individuals and businesses can outsource data storage to faraway servers using cloud storage services. Deduplication has become a widely deployed technology in cloud data centers to improve IT resources efficiency. An attacker can utilize the duplicate check to see if a file (for example, a pay slip with a given name and salary amount) has already been stored (by someone else), exposing the user's personal information. Most prior techniques that rely on the assistance of a trusted key server (KS) are vulnerable and limited due to information leakage, low attack resistance, high computational cost, and so on. If the trustworthy KS fails, the entire system collapses, resulting in single-point-of-failure. In existing system data duplication suffer from security issues. So we are solving these issues with help of hash code generation and double encryption techniques. It integrates cloud data deduplication with access control. Encrypted data can be securely accessed because only authorized data holders can obtain the symmetric keys used for data decryption. Extensive performance analysis and test showed that our scheme is secure and efficient under the described security model and very suitable for big data deduplication. We evaluate its performance based on extensive analysis and computer simulations. The results show the superior efficiency and effectiveness of the scheme for potential practical deployment, especially for big data deduplication in cloud storage.

KEYWORDS-Access control, cloud computing, data deduplication.

I. INTRODUCTION

Cloud computing offers a new way of providing services by reorganizing various resources on the Internet. The important and popular cloud service is data storage. In order to preserve the privacy of the data subjects, the data is usually stored encryption. However, encrypted data introduces new challenges for data deduplication in the cloud, which becomes crucial to large amounts of data conservation and processing in the cloud. Deduplication of traditional schemes cannot work on encrypted data. Existing encryption solutions data deduplication suffers from weak security. They cannot withstand the control and revocation of data access. Therefore, some of them. They can be easily implemented in practice. In this document, we propose a deduplication ofencrypted data stored in the cloud based on challenge properties and redefining proxy. Integrated cloud data deduplication with access control. We value their performance based on extensive analytics and computer simulations. The results show the highest efficiency and efficiency scheme potential practical distribution, especially for large data deduplication storage in the cloud

Motivation:

- To remove the duplicate copies of data and improve the reliability.
- Avoid storing deduplication data like file name also file contents. In our system generate hash code for file contents for checking data are duplicated or not. Also purpose of our system is providing double encryption techniques.

Goals/Objectives:

- To improved integrity.
- To increase the storage utilization.
- To remove the duplicate copies of data and improve the reliability.
- To improve the security.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204271 |

II. RELATED WORK

Sr.No.	Paper name and Author Name	Algorithm	Advantages/Disadvantages	Refer Point
1	M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.	AES Algorithm	The advantage of server-aided MLE is the prospect of multi-tiered security.	Deduplication, security, cloud storage.
2	J. Li, Y. K. Li, X. F. Chen, P. P. C. Lee, and W. J. Lou, "A hybrid cloud approach for secure authorized eduplication," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1206–1216, May 2015.		Adv: 1) External adversaries whichaim to extract secret information as much as possible from both public cloud and private cloud. 2) internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes.	Deduplication, authorized duplicate check, confidentiality, hybrid cloud
3	P. Meye, P. Raipin, F. Tronel, and E. Anceaume, "A secure two phase data deduplication scheme," in Proc. HPCC/CSS/ICESS, 2014, pp. 802–809,	SHA256 algorithm	Adv: Now, by taking advantage of the intra-user deduplication performed between clients and their DP, from a client point of view, the visibility of what is stored in the storage system is limited to his/her own account. Dis: To address the confidentiality issues against attacks that can be performed by the CSP.	Cloud storage; Intra-user deduplication; Inter-user deduplication; Confidentiality; Deduplication proxy;
4	Z. Yan, X. Y. Li, M. J. Wang, and A. V. Vasilakos, "Flexible data access control based on trust and reputation in cloud computing," IEEE Trans. Cloud Comput., vol. PP, no. 99, Aug. 2015.	Diffie Hellman Algorithm	Adv: It is customary to define security in a context like ours by using a model involving a game between two parties where one of the parties is an adversary who tries to get an advantage in the game. Dis: Because a user could access cloud services many times in a same context, skipping the re-encryption key generation can greatly improve the efficiency and capacity of our scheme.	Trust; reputation; access control; cloud computing.
5	P. Puzio, R. Molva, M. Onen, and S. Loureiro, "ClouDedup: Secure deduplication with encrypted data for cloud storage," in Proc. IEEE Int.	Bluefish Algorithm	Adv: The advantages of deduplication unfortunately come with a high cost in terms of new security and privacy challenges. We propose ClouDedup, a secure and efficient storage service	Deduplication, Encryption, Security

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204271 |

	Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370.		which assures block-level deduplication and data confidentiality at the same time. Dis: Furthermore, we will work on finding possible optimizations in terms of bandwidth, storage space and computation.	
6	C. Y. Liu, X. J. Liu, and L. Wan, "Policy-based deduplication in secure cloud storage," in Proc. Trustworthy Comput. Serv., 2013, pp. 250–262.	AES Algorithm	Adv: De-duplication takes advantage of data similarity in order to achieve storage reduction.	Cloud Storage, Encryption, Data De-duplication, Proxy Re-encryption, Convergent Encryption.
7	Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355.	MDS Algorithm, SHA-l Algorithm	Adv: The advantage of comparing blocks or files bit to bit is that it is accurate, but it is also time consuming. The advantage of comparing blocks or files by hash value is that it is very fast, but there is a chance of accidental collision. Dis: The disadvantages of this storage concept were the difficulty in sharing data via the Internet and the enormous wastage of storage space to keep replications, so 'once write multi read' fell into disuse.	Cloud storage, deduplication, efficiency, load balance
8	Z. C. Wen, J. M. Luo, H. J. Chen, J. X. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc. Int. Conf. Intell. Netw. Collaborative Syst., 2014, pp. 85–90.	RSA Algorithm	Adv: Firstly, before each user uploads an encrypted image, he calculates its hash value as the fingerprint. Secondly, the fingerprint is sent to both cloud servers for checking duplicates. If the storage and verification servers both reply to the user with 'no deduplication', the user transfers his data to the servers.	Deduplication, Image, Cloud computing, Encryption

III. EXISTING SYSTEM APPROACH

Existing solutions for deduplication suffer from brute-force attacks. They cannot friendly support data access control and revocation at the same time. Most existing solutions cannot ensure reliability, security and privacy with sound performance. First data holders may not be always online or available for each a management, which could come storage delay. Second deduplication could become too complicated in the term of communication and computation to involve data holder into deduplication process. Third, it may intrude the privacy of data holder in a process of discovering duplicated data. Forth a data holder may have no idea how to issue data access right or deduplication key to users in some situation when it does not know other data holders due to data suffer distribution. Therefore, CSP cannot cooperate with data holders on data storage deduplication in many situations.

on ownership challenge and PRE. Our scheme can flexibly support data update and sharing with deduplication even when the data holder are offline. Encrypted data can be securely accessed because only authorized data holders can

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204271 |

obtain the symmetric keys for the data decryption. Extensive performance analysis and test showed that our scheme is secured and efficient under the described security model and very suitable for deduplication. The result of our computer simulations further showed the practically of our scheme.

IV. PROPOSED SYSTEM APPROACH

In recent time, there are many problems of storage places in cloud. If data holderstore file in cloud which is already available in cloud, so this is waste of memory.
So, in this paper we remove the deduplication. Also, security issues are occurduring the cloud server data storage. We are solving these problems with the help ofdouble encryption technique and hash code techniques for generate hash value

Data Holders

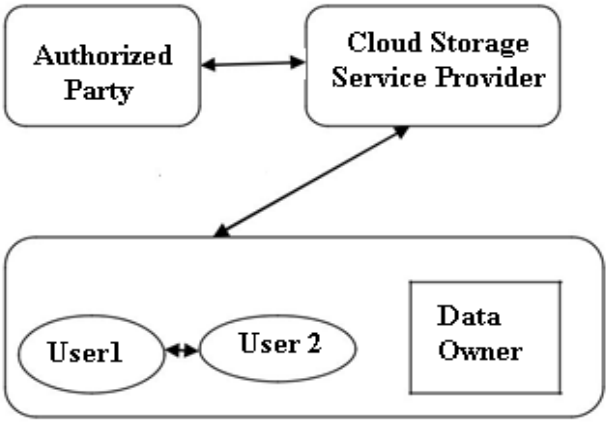

Fig 1: System architecture

- **CSP:** The CSP is allowing to data owner for storageservices of data. It cannot be fully trusted. Therefore it is curious about the contents of stored data. It should perform honestly on data storage in order to gain profit.

- **Data Holder:** The data holder can uploads and savestheir data and files in the CSP. In this system is possible to number of data holders could save their files in encrypted raw data in the CSP. The file is regarded as data owner by the data holder that produces or creates the
File. The data holder is in normal form than the higher priority of owner.

- **AP:** an authorized party (AP) that is fully trusted by thedata holders. The data holders to verify data ownership and handle data deduplication. It does not collude with CSP. In this case, CSP should not know the plain user data in its storage. AP cannot know the data stored in CSP.

V. CONCLUSION

Discuss about to managing encrypted data with deduplication is for achieving a successful cloud storage service, for big data storage. In this paper, proposed a scheme to manage the encrypted big data in cloud with deduplication based on PRE and ownership challenge. This scheme can flexibly support data update and sharing with deduplication even when the data holders are offline. Only authorized data holders can use the symmetric keys used for data decryption by encrypted data can be securely accessed. The performance analysis and graphs showed that there scheme is secure under the described security model for big data deduplication. The results of the computer simulations further showed in the variation of graph. Future work is to include optimizing there design and implementation for big data storage in cloud. To ensure that CSP behaves as expected in deduplication management Studying verifiable computation. The part of cloud computing has brought many researchers from different fields; yet, much effort remains to reach use and the

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204271 |

broad acceptance of cloud computing technology. In the future scope propose a big data deduplication. For the future environment system can focus on personalize search on user feedback sessions as well as recommendation based on user point of interest with database security is the interesting part of system

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.
[2] J. Li, Y. K. Li, X. F. Chen, P. P. C. Lee, and W. J. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1206–1216, May 2015, doi:10.1109/TPDS.2014.2318320.
[3] P. Meye, P. Raipin, F. Tronel, and E. Anceaume, "A secure two phase data deduplication scheme," in Proc. HPCC/CSS/ICESS, 2014, pp. 802–809, doi:10.1109/HPCC.2014.134.
[4] Z. Yan, X. Y. Li, M. J. Wang, and A. V. Vasilakos, "Flexible data access control based on trust and reputation in cloud computing," IEEE Trans. Cloud Comput., vol. PP, no. 99, Aug. 2015, doi:10.1109/ TCC.2015.2469662, Art. no. 1.
[5] P. Puzio, R. Molva, M. Onen, and S. Loureiro, "ClouDedup: Secure deduplication with encrypted data for cloud storage," in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370, doi:10.1109/CloudCom.2013.54.
[6] C. Y. Liu, X. J. Liu, and L. Wan, "Policy-based deduplication in secure cloud storage," in Proc. Trustworthy Comput. Serv., 2013, pp. 250–262, doi:10.1007/978-3-642-35795-4_32.
[7] Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355, doi:10.1109/CSCWD.2011.5960097.
[8] Z. C. Wen, J. M. Luo, H. J. Chen, J. X. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc. Int. Conf. Intell. Netw.Collaborative Syst., 2014, pp. 85–90, doi:10.1109/INCoS.2014.111.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details