Data Deduplication Using Hybrid Encryption Algorithm

Anusha.P
UG Scholar, Department of Computer Science and Engineering,Velammal engineering college, chennai
Malarvizhi.R
UG Scholar, Department of Computer Science and Engineering,Velammal engineering college, chennai
Mrs.Sumathi
Asst.ProfessorI, Department Of Computer Science and EngineeringVelammal Engineering College.

ABSTRACT

Secure data de-duplication can significantly reduce the communication and storage overheads in cloud storage services,and has potential applications in our big data-driven society. Existing data deduplication schemes are generally designed to either resist brute-force attacks or ensure the efficiency and data availability, but not both conditions. In cloud environments, users store their data or files in cloud storage but it is not infinitely large. In order to reduce the requirement of storage and bandwidth, data deduplication has been applied. Users can share one copy of the duplicate files ordata blocks to eliminate redundant data. If the storage and verification servers both reply to the user with 'no deduplication', the user transfers his data to the servers . Otherwise, once the alter box is consistently found, the user gives up uploading data for deduplication. Specially , when the alter box is only found in one server, it implies that the results are inconsistent and at least one of servers is invalid. The security and efficiency analysis is also presented in this paper.

Keywords- **Deduplication, Cloud computing, Encryption, data availability, accountability, Big data.**

I. INTRODUCTION

Cloud storage usage is likely to increase in our big data driven society. For example, IDC predicts that the amount of digital data will reach 44 ZB in 2020 [1]. Other studies have also suggested that about 75% of digital data are identical (or duplicate) [2], and data redundancy in backup and archival storage system is significantly more than 90% [3]. Many papers have discussed the security issues on cloud computing [4, 12]. There are three factors on security of cloud computing: data confidentiality, integrity, and availability (CIA) [4]. Data deduplication is a specialized data compression technique for eliminating the redundant data.

The data are really stored on the storage media if they were uploaded first; otherwise, the redundant data are replaced with a pointer to the unique data copy. To achieve data deduplication, there are two basic approaches [5]: target-based approach and source-based approach. However, storage of high redundant data makes inefficient use of cloud storage resources and upload bandwidth [6]. the verifiable deduplication of image, which is an usual data type of large-scale and high dimensionality stored on internet and cloud [7], has received limited attention so far. This complicates data deduplication efforts, as identical data encrypted by different users (or even the same user using different keys) will result in different ciphertexts [7], [8].Thus, how to efficiently perform data deduplication on encrypted data is a topic of ongoing research interest. In recent times, a number of data deduplication schemes have been proposed in the literature. These schemes are designed to realize encrypted data deduplication (see [9], [10], [11]). example, the _R-MLE2 (Dynamic) scheme proposed in [12] seeks to improve the efficiency of duplicate cipher text identification. However, the scheme suffers from brute-force attacks, the most popular attack in secure data deduplication schemes. In this paper, we resolve the secure deduplication problem by introducing a verification cloud server besides the storage server, which play the role of the verifier in the model. For efficiency consideration, The outsourced data should not berequired by the verifier for the verification purpose [13]. [14] proposed a scheme to deduplicate encrypted big data stored in the cloud based on ownership challenge and proxy re-encryption. Although this scheme is efficient, it is vulnerable to brute-force attacks. A number of incidents have shown that such information may be more invasive to one's privacy than the core data itself(e.g. NSA PRISM [15]). However, such information disclosure is inevitable in existing deduplication schemes. Therefore, we aim to minimize information leakage as much as practical, in the sense that only the entity (the cloud storage provider) that operates the deduplication knows it. Furthermore, if the duplicate information is leaked, then the cloud storage provider will be held accountable[16]. In order to efficiently complete deduplication, the hash value of encrypted image is calculated and transferred to both two cloud servers as the fingerprint used in the deduplication process [18], [19], which is proved to be secure.

II. RELATED WORK

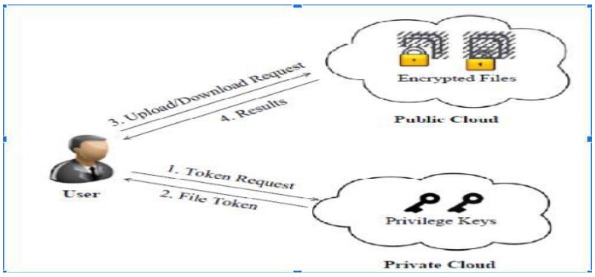
Data deduplication :As the volume of data stored in the cloud increases quickly the term deduplication technique has been more and more concerned recently. The strategies of deduplication can be categorized to two strategies: file-level and block-level deduplication .The file-level deduplication eliminates duplicate data copy at the file granularity if two files have the same hash value and are identified as identical [22][10][11].This method needs low computational overhead but has low duplicate elimination effectiveness. The other block-level deduplication is also a popular technique which first divides each input file into several blocks of fixed-size or variable size and then use hash value of each block to eliminates the block already stored in cloud.

III. PROPOSED SYSTEM

In this paper, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model.

The user is only allowed to perform the duplicate check for files marked with the corresponding privileges. We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality.

IV. ARCHITECTURE



V. CONCLUSION

In this paper, we proposed an efficient and privacy-preservingbig data deduplication in cloud storage for a three-tier crossdomain architecture. We then analyzed the security of our proposedscheme and demonstrated that it achieves improved privacypreserving, accountability and data availability, while resisting brute-force attacks. We also demonstrated that the proposed scheme outperforms existing state-of-the-art schemes, in terms of computation, communication and storage overheads. In addition, the time complexity of duplicate search in our scheme is an efficient logarithmic time.Future research includes extending the proposed scheme to fully protect the duplicate information from disclosure, even by a malicious CSP, without affecting the capability to perform data deduplication. Future research agenda will also include extending the scheme to be resilient against a wider range of security threats by external attackers, as well as improving the time complexity of duplicate search.All files have priviledge keys of their own which enchances security in the system. Token for accessing a file is sent as a otp to users email FURTHER ENCHANCEMENT: Working on the area ,To generate otp to the users mobile numbers

REFERENCES

[1] IDC, “Executive summary: Data growth, business opportunities, and the it imperatives,,http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm, 2014.

[2] J. Gantz and D. Reinsel, “The digital universe decadeareyou ready,” https://hk.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf.

[3] H. Biggar, “Experiencing data de-duplication: Improving efficiencyand reducing capacity requirements,” The Enterprise Strategy Group.,2007. [Online]. Available: http://journals.sagepub.com/doi/abs/10.1177/

[4] 000944550704300309

[5] J. Harauz, L. M. Kaufman, and B. Potter, “Data Security in the World ofCloud Computing,” IEEE Security and Privacy, Vol.7, No. 4, 2009, pp.61-64

[6] D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Side Channels in Cloud Services: Deduplication in Cloud Storage,” IEEE Security and Privacy, Vol. 8, No. 6, 2010, pp. 40-47.

[7] AtulKatiyar and JonWeissman. Videdup: an application-aware framework for video de-duplication. In Proceedings of the 3rdUSENIX conference on Hot topics in storage and file systems, pages 7–7. USENIX Association, 2011.

[8] Michael S Lew, NicuSebe, ChabaneDjeraba, and Ramesh Jain. Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2(1):1–19, 2006.

[9] J. Paulo and J. Pereira, “A survey and classification of storage deduplication systems,” ACM Comput. Surv., vol. 47, no. 1, pp. 11:1–11:30, 2014. [Online]. Available: http://doi.acm.org/10.1145/2611778

[10] S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Server-aidedencryption for deduplicated storage,” in Proceedings of the 22th USENIXSecurity Symposium, Washington, DC, USA, August 14-16, 2013, 2013,pp.179–194. [Online]. Available: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/bellare

[11] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-lockedencryption and secure deduplication,” in Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on theTheory and Applications of Cryptographic Techniques, Athens, Greece,May 26-30, 2013. Proceedings, 2013, pp. 296–312. [Online]. Available:http://dx.doi.org/10.1007/978-3-642-38348-9 18

[12] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev,“Message-locked encryption for lock-dependent messages,” in Advancesin Cryptology - CRYPTO 2013 - 33rd Annual Cryptology

Conference,Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, PartI, 2013, pp. 374–391. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40041-4 21

[13] T. Jiang, X. Chen, Q. Wu, J. Ma, W. Susilo, and W. Lou, "Secure andefficient cloud data deduplication with randomized tag," IEEE Trans.Information Forensics and Security, vol. PP, no. 99, pp. 1–1, 2016.

[14] Qian Wang, Cong Wang, Jin Li, Kui Ren, and Wenjing Lou.Enabling public verifiability and data dynamics for storage security in cloud computing. In Computer Security–ESORICS2009, pages 355–370. Springer, 2009.

[15] Z. Yan, W. Ding, X. Yu, H. Zhu, and R. H. Deng, "Deduplication onencrypted big data in cloud," IEEE Trans. Big Data, vol. 2, no. 2, pp.138–150, 2016. [Online]. Available: http://dx.doi.org/10.1109/TBDATA.2016.2587659

[16] "Prism (surveillance program)," https://www.theguardian.com/us-news/prism.

[17] [16] R. Bhaskar, S. Guha, S. Laxman, and P. Naldurg, "Verito: Apractical system for transparency and accountability in virtual economies," in 20th Annual Network and Distributed System SecuritySymposium, NDSS 2013, San Diego, California, USA, February24-27, 2013, 2013. [Online]. Available: http//:internetsociety.org/doc/verito-practical-system-transparency-and-accountability-virtual-economies

[18] Halevi, Shai and Harnik, Danny and Pinkas, Benny andShulman-Peleg, Alexandra, Proofs of ownership in remote storage systems, Proceedings of the 18th ACM conference onComputer and communications security, page 491–500, 2011.

[19] Wang, Cong and Wang, Qian and Ren, Kui and Lou, Wenjing, Privacy-preserving public auditing for data storage security in cloud computing, INFOCOM, 2010 Proceedings IEEE, pp. 1–9, 2010.

[20] Somani, Uma and Lakhani, Kanika and Mundra, Manish,Implementing digital signature with RSA encryption algorithm to enhance the Data Security of cloud in Cloud Computing,Parallel Distributed and Grid Computing (PDGC), pp. 211–216, 2010.

[21] Wang, Qian and Ren, Kui and Yu, Shucheng and Lou, Wenjing,Dependable and secure sensor data storage with dynamicintegrity assurance, ACM Transactions on Sensor Networks(TOSN), vol. 8(1), 2011.