User-Defined Access Control for Efficient Secure Deduplication in the Cloud

D. Bujjibabu 1, Dudipalli Sindhuja 2

#1Professor in the department of MCA at QIS College of Engineering and Technology (Autonomous), Vengamukkapalem, Prakasam (DT)
#2MCA Student in the department of MCA at QIS College of Engineering and Technology (Autonomous), Vengamukkapalem, Prakasam (DT)

ABSTRACT_ Distributed storage as one of the main administrations of distributed computing which fundamentally works with cloud clients to re-appropriate their information to the cloud for capacity and offer them with approved clients. Secure deduplication has received a lot of attention in cloud storage because it can reduce communication overhead and storage space by eliminating redundancy in encrypted data. In terms of privacy and security, numerous current secure deduplication schemes typically concentrate on achieving the following properties: confidentiality of the data, consistency of the tags, management of access, and resistance to brute-force attacks However, as far as we are aware, none of them can simultaneously fulfill these four requirements. To defeat this deficiency, in this article, we propose a proficient secure deduplication plot that upholds client characterized admittance control. Particularly, our plan maximizes the elimination of duplicates without compromising cloud users' privacy or security by allowing only the cloud service provider to authorize data access on behalf of owners. Our authorized secure deduplication scheme, according to a comprehensive security analysis, prevents brute-force attacks while maintaining data confidentiality and tag consistency. In addition, extensive simulations demonstrate that our strategy outperforms the competing ones in terms of the efficiency of deduplication and the overheads associated with computation, communication, and storage.

1.INTRODUCTION

With the rapid growth of data volumes, there are increasing demands for safe places to store private data. An effective solution to this issue is to outsource big data to the cloud [1, 2]. Duplicate data still eats up a lot of storage space and network bandwidth and complicates data management, despite all the benefits of cloud computing [3]. The cloud storage provider is able to save storage space by storing only a single copy of the data that is owned by multiple owners thanks to the deduplication process [5], which is a process that identifies the same data by data similarity. However, there are still issues with dynamic ownership management and access control with the current deduplication schemes. First and foremost, many users encrypt data prior to uploading it to cloud storage to safeguard their privacy. Deduplication will be hampered if the same data is encrypted using different keys because the encryption key is generated at random. Some deduplication schemes suggest that the owners of the same file share the same encryption key to solve this issue [6–14]. However, the majority of them do not take into account the frequent dynamic ownership shifts in cloud storage services [15]. The cloud clients ought to be renounced from the legitimate possession list once they demand the distributed storage supplier for information cancellation/change. Second, numerous schemes [4], [15], and [19] were proposed to address dynamic ownership management by utilizing Authority Party (AP) or The Public Cloud Provider (P ub CSP) as trusted or semi-trusted third parties to perform proxy reencryption work. From one viewpoint, it very well may be challenging to execute a confided in outsider in commonsense applications [20]. On the other hand, when a third party conspires with unauthorized users, some schemes cannot withstand collusion attacks. In this paper, we propose a clever plan, which focuses on effectively taking care of the issue of deduplication with regular cloud client disavowal and new cloud client participating in distributed computing. Specifically, not the same as existing information deduplication strategies, which utilize either trusted/semi-believed outsider to do intermediary re-encryption work, our proposed plot plans a half breed cloud engineering, which incorporates a public cloud and further presents a confidential cloud. In our scheme's implementations, the introduced private cloud acts as both a data owner and a proxy to 1) manage the dynamic ownership when the real data owner is offline or revokes ownership, and 2) control access to outsourced data through re-encryption techniques. Moreover, we propose to improve our plan concerning productivity by 1) guaranteeing that the information proprietor performs encryption just when he/she is the underlying uploader; 2) introducing an entrance control procedure that confirms the legitimacy of the information clients before they download information; 3) requiring the cloud user to be on the ownership list before the public cloud server can send ciphertext to them As a result, the expense of extensive communication will decrease.

2.LITERATURE SURVEY

1.Z. Yan, L. Zhang, W. Ding, and Q. Zheng, "Heterogeneous data storage management with deduplication in cloud computing," IEEE Trans. Big Data, vol. 5, no. 3, pp. 393–407, Sep. 2019

Abstract:

Cloud storage as one of the most important services of cloud computing helps cloud users break the bottleneck of restricted resources and expand their storage without upgrading their devices. In order to guarantee the security and privacy of cloud users, data are always outsourced in an encrypted form. However, encrypted data could incur much waste of cloud storage and complicate data sharing among authorized users. We are still facing challenges on encrypted data storage and management with deduplication. Traditional deduplication schemes always focus on specific application scenarios, in which the deduplication is completely controlled by either data owners or cloud servers. They cannot flexibly satisfy various demands of data owners according to the level of data sensitivity. In this paper, we propose a heterogeneous data storage management scheme, which flexibly offers both deduplication management and access control at the same time across multiple Cloud Service Providers (CSPs). We evaluate its performance with security analysis, comparison and implementation. The results show its security, effectiveness and efficiency towards potential practical usage.

2. X. Liu, K. R. Choo, R. H. Deng, R. Lu, and J. Weng, "Efficient and privacy-preserving outsourced calculation of rational numbers," IEEE Trans. Dependable Secure Comput., vol. 15, no. 1, pp. 27–39, Feb. 2018

In this paper, we propose a framework for efficient and privacy-preserving outsourced calculation of rational numbers, which we refer to as POCR. Using POCR, a user can securely outsource the storing and processing of rational numbers to a cloud server without compromising the security of the (original) data and the computed results. We present the system architecture of POCR and the associated toolkits required in the privacy preserving calculation of integers and rational numbers to ensure that commonly used outsourced operations can be handled on-the-fly. We then prove that the proposed POCR achieves the goal of secure integer and rational number calculation without resulting in privacy leakage to unauthorized parties, and demonstrate the utility and the efficiency of POCR using simulations.

3.DaSCE: Data Security for Cloud Environment with Semi-Trusted Third-Party IEEE Transactions on Cloud Computing (Volume: 5, Issue: 4, 01 Oct.-Dec. 2017) **AUTHORS: Ali, M., Malik, S. and Khan, S.,**

Off-site information storage is an utility of cloud that relieves the clients from focusing on records storage system. However, outsourcing records to a third-party administrative manipulate entails serious safety concerns. Data leakage may additionally appear due to assaults by means of different customers and machines in the cloud. Wholesale of statistics through cloud carrier issuer is but any other trouble that is confronted in the cloud environment. Consequently, high-level of protection measures is required. In this paper, we advocate Data Security for Cloud Environment with Semi-Trusted Third Party (DaSCE), a records safety machine that presents (a) key administration (b) get entry to control, and (c) file certain deletion. The DaSCE makes use of Shamir's (k, n) threshold scheme to manipulate the keys, the place okay out of n shares are required to generate the key. We use a couple of key managers, every web hosting one share of key. Multiple key managers keep away from single factor of failure for the cryptographic keys. We (a) put into effect a working prototype of DaSCE and consider its overall performance based totally on the time fed on in the course of more than a few operations, (b) formally mannequin and analyze the working of DaSCE the usage of High Level Petri nets (HLPN), and (c) affirm the working of DaSCE the usage of Satisfiability Modulo Theories Library (SMT-Lib) and Z3 solver. The outcomes disclose that DaSCE can be successfully used for safety of outsourced statistics with the aid of using key management, get right of entry to control, and file certain deletion.

3.PROPOSED SYSTEM

To overcome this challenge, in this paper, we design an efficient secure cross-user deduplication scheme with user-defined access control.

In our authorized deduplication system, the proposed scheme should ensure that outsourced encrypted data can be available not only to the data owner but also to all authorized users selected by this data owner. On the contrary, unauthorized users cannot access the content of encrypted data outsourced by users of interest

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 cloud End users. To access the shared data files users will request the permission of content key and the MSK master secret key. And the cloud will provide the permission .and also views all the transactions and attackers related to the files.

KGC

Authority generates the content key and the secret key requested by the end user. Authority can view all files with the content key and master secret key generated with the corresponding data owner details of the particular file.

END USER

User has to register and login for accessing the files in the cloud. User is authorized by the cloud to verify the registration. User has to request for the MSK master secret key and content key to download the file. User can only download and search the file if the data owner of the particular file has provided the permissions.

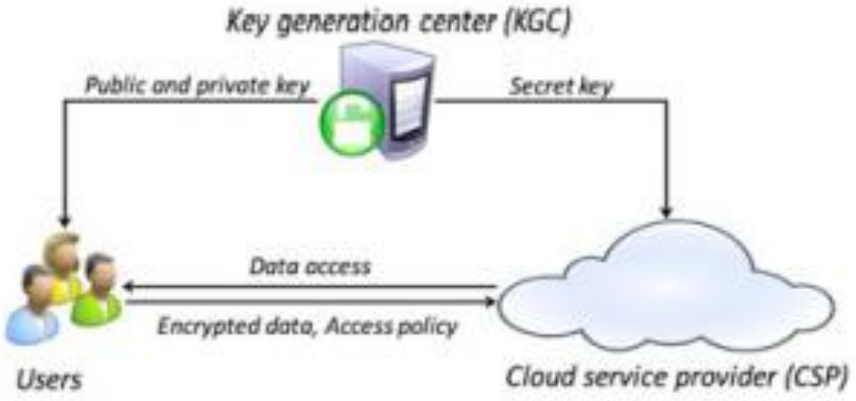

Key generation center (KGC)
Public and private key
Secret key
Data access
Encrypted data, Access policy
Users
Cloud service provider (CSP)

Fig 1: Architecture

4.RESULTS AND DISCUSSION

Table 1:Comparision Table

Scheme	Without IS in data upload phase	Confidentiality	Resistance to brute-force attacks	Tag consistency	Access control
Li *et al.*'s scheme [17]	○	●	◑	○	●
Cui *et al.*'s scheme [18]	○	●	◑	◑	◑
Yan *et al.*'s scheme [19]	○	●	○	○	●
Liu *et al.*'s scheme [35]	●	●	●	●	○
Our scheme	●	●	●	●	●

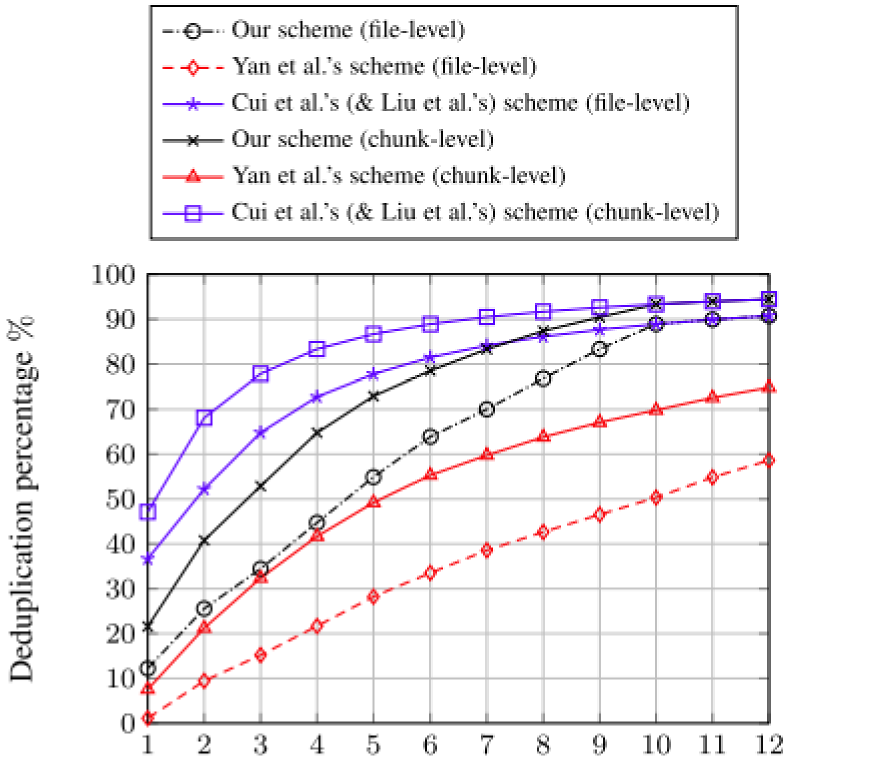
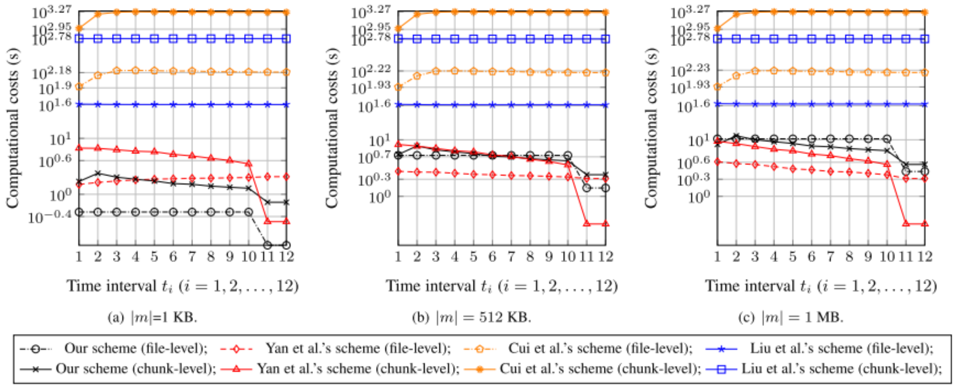


Fig. 2. Effectiveness of deduplication

We compare the effectiveness of deduplication for four schemes in Fig. 2, which demonstrates that the effectiveness of chunk-level deduplication is superior to the file-level deduplication. Furthermore, the deduplication effectiveness of Cui et al.'s scheme [18] and Liu et al.'s scheme [35] is the best, followed by our scheme, the worst is Yan et al.'s scheme [19]. Although the deduplication effectiveness of our scheme is less than that of schemes in [18], [35], the gap is shrinking over time and our scheme can further satisfy access control. Specifically, Liu et al.'s scheme does not consider access control, and thus the CSP checks for duplicates in all stored data. In other words, the CSP just stores one copy of each uploaded data, ensuring the best deduplication effectiveness.



(a) $|m|$=1 KB.

(b) $|m| = 512$ KB.

(c) $|m| = 1$ MB.

Fig. 3. Comparison of computational costs in the phase of data upload

We depict the comparison of computational costs in Fig. 3. From the figure, it can be obviously shown that computational costs of chunk-level deduplication are larger than that of file-level deduplication. Further, when the message length (jmj) is relatively small, computational efficiency is mainly affected by deduplication-related operations (e.g., tag generation, duplicate search and verification). As a result, Fig. 3a demonstrates that computational costs associated with deduplication operations in our scheme are the lowest. As jmj increases, the computational efficiency of our scheme is slightly lower than that of Yan et al.'s scheme (see Figs. 3b and 3c). The main reason is that our scheme generates the tag with the encrypted data, and thus computational costs are influenced by the symmetric encryption algorithm (i.e., jmj). From Fig. 3, we can also observe that the computational efficiency of our scheme is far superior to Liu et al.'s scheme and Cui et al.'s scheme, especially for the chunk-Leve deduplication. The main factors include the tag generation, the operation of duplicate verification and the time complexity of duplicate search. Specifically, Liu et al.'s scheme requires users to interactively run the same-input-PAKE protocol many times to generate the tag, which incurs high computational costs. Cui et al.'s scheme checks duplicates by performing time-consuming pairing operations and the time complexity of duplicate search increases linearly with the number of stored files (i.e., Oð2nÞ pairing operations for n stored data). On the contrary, our scheme does not require interactive protocols and the corresponding duplicate search is based on an efficient Bloom filter technique, where the time complexity of duplicate search is OðfÞ (f n).

.5.CONCLUSION

In this paper, we present a user-defined access control efficient secure deduplication scheme. In particular, our method does not require the use of a hybrid cloud architecture or the addition of a second authorised server to achieve the authorised deduplication. Only the CSP in our system is capable of managing access rights on behalf of data owners without jeopardising data privacy. Also included in our plan is the Bloom filter, which effectively completes the duplicate check. According to thorough security analyses, our scheme can simultaneously achieve data confidentiality, access control, tag consistency, and resistance to brute-force attacks. Additionally, thorough performance analyses of file-level and chunk-level deduplication demonstrate the effectiveness of our plan in terms of deduplication efficiency, computational cost, communication overhead, and storage cost.

REFERENCES

[1] Joseph A Akinyele, Christina Garman, Ian Miers, Matthew W Pagano, Michael Rushanan, Matthew Green, and Aviel D Rubin. Charm: a framework for rapidly prototyping cryptosystems. Journal of Cryptographic Engineering, 3(2):111–128, 2013.

[2] Ittai Anati, Shay Gueron, Simon Johnson, and Vincent Scarlata. Innovative technology for cpu based attestation and sealing. In Workshop on hardware and architectural support for security and privacy (HASP), volume 13, page 7. ACM New York, NY, USA, 2013.

[3] Alexandros Bakas and Antonis Michalas. Modern family: A revocable hybrid encryption scheme based on attribute-based encryption, symmetric searchable encryption and SGX. In SecureComm 2019, pages 472–486, 2019.

[4] Amos Beimel. Secure schemes for secret sharing and key distribution. PhD thesis, PhD thesis, Israel Institute of Technology, Technion, Haifa, Israel, 1996.

[5]JohnBethencourt,AmitSahai,andBrentWaters. Ciphertext-policy attribute-based encryption. In S&P 2007, pages 321–334. IEEE, 2007.

[6] Victor Costan and Srinivas Devadas. Intel sgx explained. IACR Cryptology ePrint Archive, 2016(086):1–118, 2016.

[7] Ben Fisch, Dhinakaran Vinayagamurthy, Dan Boneh, and Sergey Gorbunov. IRON: functional encryption using intel SGX. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, pages 765–782, 2017.

[8] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In Advances in Cryptology-CRYPTO 1999, pages 537–554. Springer, 1999.

[9] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In ACM CCS 2006, pages 89–98. ACM, 2006.

[10] Jinguang Han, Willy Susilo, Yi Mu, Jianying Zhou, and Man Ho Allen Au. Improving privacy and security in decentralized ciphertext-policy attribute-based encryption. IEEE transactions on information forensics and security, 10(3):665–678, 2015.

Author's Profiles

Dr.D.Bujji Babu, currently working as a Professor and Head in the Department of Master of Computer Applications , QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He did his M. Tech (CSE) from JNTUK, Kakinada and Ph.D(CSE) from Acharya Nagarjuna University. He published more than 50 research papers in reputed peer reviewed Scopus indexed journals. He also attended and presented research papers in different national and international journals and the proceedings were indexed IEEE, Springer Link series. He visited the countries Kuching, Malaysia for attending and presenting his research articles. 3 Patent journals are published and in pipeline for grant. He wrote more than a dozen of monographs and published by the Technical Publishers. He Published Two Course content modules for the students of Acharya Nagarjuna University. He is the recognized research supervisor under JNTUK, Kakinada and guided several UG and PG Projects, currently supervising 3 research scholar under JNTUK. His area of interest is Software Engineering, Data Mining, Data Science, Big Data and Programming Languages. He is the Principal Investigator for the DST Sponsored Project and Co-PI for another DST Project.

Ms. D. Sindhuja as MCA Student in the department of Qis College of engineering and Technology (autonomous), Vengamukkapalem, Prakasam (DT). She has Completed B.Sc.

in Computer Science from Grk Degree &p.g College. Her areas of interests are Cloud Computing& machine learning.