Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, July : 2023

Enabling Secure and Space-Efficient Metadata Management for Encrypted Deduplication

Mrs MARESWARAMMA P [1] , P. Haritha [2]

#1 Associate Professor and Head in the department of AI & IT at DVR & Dr HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District.

#2 MCA student in the Department of computer applications at DVR & DR. HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR District.

ABSTRACT_ Scrambled deduplication consolidates encryption and deduplication in a consistent manner to give secrecy certifications to the actual information in deduplicated capacity, yet it causes significant metadata stockpiling above because of the extra stockpiling of keys. We present another encoded deduplication capacity framework called Metadedup, which smothers metadata capacity by likewise applying deduplication to metadata. Its thought expands on indirection, which adds one more degree of metadata pieces that record metadata data. We find that metadata pieces are profoundly repetitive in genuine jobs and subsequently can be successfully deduplicated. Using a distributed key management strategy, we further extend Metadedup to incorporate multiple servers to provide fault-tolerant storage and security guarantees. We widely assess Metadedup from execution and capacity productivity viewpoints. For real-world backup workloads, we demonstrate that Metadedup saves metadata storage by up to 93.94% and achieves high file write and restore throughput**.**

1.INTRODUCTION

In today's primary [31] and backup [26], [39], and [42] storage systems, CHUNK-BASED deduplication is frequently used to save a lot of space. It stores just a solitary actual duplicate of copy pieces, while referring to all copy lumps to the actual duplicate by little size references. Deduplication has been shown to effectively reduce primary storage's storage space by 50% [31] and backup storage's storage space by up to 98 percent [39]. This spurs the wide organization of deduplication in different business distributed storage administrations (e.g., Dropbox, Google Drive, Bitcasa, Mozy, and Memopal) to diminish significant capacity costs [18]. To give classification ensures, scrambled deduplication [7], [8] adds an encryption layer to deduplication, to such an extent that each piece, prior to being composed to deduplicated capacity, is deterministically encoded through symmetric-key encryption by a key got from

the lump content (e.g., the key is set to be the cryptographic hash of piece content [14]). Because of this, we are able to apply deduplication to the encrypted chunks in order to save space because duplicate chunks still have the same content even after encryption. In order to effectively manage outsourced data in cloud storage, numerous studies (e.g., [5], [7], [25], [33], and [36]) have designed various encrypted deduplication schemes. A deduplicated storage system must maintain deduplication metadata in addition to storing non-duplicate data. Deduplication metadata can be divided into two categories. The system keeps a fingerprint index that tracks the fingerprints of all chunks that have already been stored to see if they are identical. Likewise, to permit a record to be reproduced, the framework keeps a document recipe that holds the mappings from the lumps in the record to the references of the relating physical copies.\

Deduplication metadata is famously known to cause high capacity above [11], [21], [30], particularly for the exceptionally repetitive responsibilities (e.g., reinforcements) as the metadata stockpiling above turns out to be more predominant. We argue in this work that encrypted deduplication keeps key metadata, such as key recipes that keep track of the chunk-to-key mappings that make it possible to decrypt individual files, which results in even higher metadata storage overhead. Key recipes must be managed separately from file recipes, encrypted using the master keys of file owners, and stored separately for each file owner because they contain sensitive key information. In actual deployment, encrypted deduplication's storage efficiency may be compromised by such a high metadata storage overhead.

2.LITERATURE SURVEY

1.H. Wang, D. He, A. Fu, Q. Li, and Q. Wang, ‘‘Provable data possession with outsourced data transfer,’’ IEEE Trans. Services Comput., vol. 14, no. 6, pp. 1929–1939, Nov. 2021.

Abstract:
With the rapid development of cloud computing, more and more enterprises would like to upload and store their data in the public cloud. When the parts of the business of an enterprise are purchased by another enterprise, the corresponding data will be transferred to the acquiring enterprise. For the usual case, how to outsource the computation cost of data transfer to the cloud? How to ensure the remote purchased data integrity? Thus, it is important to study provable data possession with outsourced data transfer (DT-PDP). In this paper, for the first time, we propose the novel concept: DT-PDP. By taking use of DT-PDP, the following three security requirements can be satisfied: (1) the other un-purchased data security of acquired enterprise can be ensured; (2) the purchased data integrity and privacy can be ensured; (3) the data

transferability's computation can be outsourced to the public cloud servers. For the security concept of DT-PDP, we give its motivation, system model and security model. Then, we design a concrete DT-PDP scheme based on the bilinear pairings. At last, we analyze the security, efficiency and flexibility of the concrete DT-PDP scheme. It shows that our scheme is provably secure and efficient.

2. J. Chang, B. Shao, Y. Ji, M. Xu, and R. Xue, ''Secure network coding from secure proof of retrievability,'' Sci. China Inf. Sci., vol. 64, no. 12, Dec. 2021, Art. no. 229301

Abstract

In recent years, storage-as-a-service has emerged as a commercial alternative for user's local data storage due to its features include less initial infrastructure setup, relief from maintenance overhead, and universal access to the data irrespective of the location and devices [5]. However, it also faces several security threats. One of the most serious threats is the integrity of user's stored data. In particular, when storing the data file to a cloud service provider (CSP), a user (or data owner) will delete it from his/her local devices and hence lose local control of it. In this case, CSP may discard some user's rarely accessed data to save its space and earn more profit. Meanwhile, the CSP can lie about the fact. Obviously, it is extremely unfavorable for users. Proof of retrievability (PoR) protocol is just one of initial attempts to formulize the notion of "remotely and reliably checking data's integrity without downloading the whole data file".

3.N. Döttling and S. Garg, ''Identity-based encryption from the Diffie–Hellman assumption,'' J. ACM, vol. 68, no. 3, pp. 1–46, Mar. 2021

Abstract

We provide the first constructions of identity-based encryption and hierarchical identitybased encryption based on the hardness of the (Computational) Diffie-Hellman Problem (without use of groups with pairings) or Factoring. Our construction achieves the standard notion of identity-based encryption as considered by Boneh and Franklin [CRYPTO 2001]. We bypass known impossibility results using garbled circuits that make a non-black-box use of the underlying cryptographic primitives

3.PROPOSED SYSTEM

This paper proposes an information uprightness review conspire in light of the cloud and haze design, in the mean time, gives an information transmission model in the cloud and mist organization. Fog nodes in this model transmit data and compute the lowest possible communication channel to reduce communication overhead. Simultaneously, a visually impaired factor is presented in the proof age phase of the trustworthiness review to keep the enemy from working out the ciphertext in the two cross examinations and work on the security of the honesty review.

The following are the main contributions of this paper:

1) This paper proposes an information respectability review model in a cloud and haze climate, which can really lessen the correspondence above in the transmission

process and lessen the figuring strain of the cloud specialist organization.

2) A blind factor is added to the data integrity audit to prevent data leakage caused by malicious auditors challenging the data multiple times.

3) This article demonstrated the scheme's security using the provided security model. The outcomes of the experiments demonstrate that this plan performs better and is more practical.

3.1 IMPLEMENTATION

Data Owner

In this module, the data owner uploads their encrypted data in the Cloud server. For the security purpose the data owner encrypts the data file and then store in the server. The Data owner can have capable of manipulating the encrypted data file and performs the following operations Register and Login, Upload File, View Files, Update File, Verify File's Block(Data Integrity Auditing).

Cloud

The Cloud manages which is to provide data storage service for the Data Owners. Data owners encrypt their data files and store them in the Server for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the Server and then Server will decrypt them. The server will generate the aggregate key if the end user requests for file authorization to access and performs the following operations such as Login, View and Authorize User, View and Authorize Owner, View Files By Block chain, View All Transactions, Search Requests, Download Requests, View All Attackers, View File Rank Chart, View Time Delay Results, View Throughput Results.

User

In this module, the user can only access the data file with the secret key. The user can search the file for a specified keyword. The data which matches for a particular keyword will be indexed in the cloud server and then response to the end user and performing the following operations Register and Login, Search, Download, View Files, Search Request, Download Request.

TPA – responsible for Login, View File's Meta Data, View Files & Generate Secret Key, CPU Speed.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, July : 2023

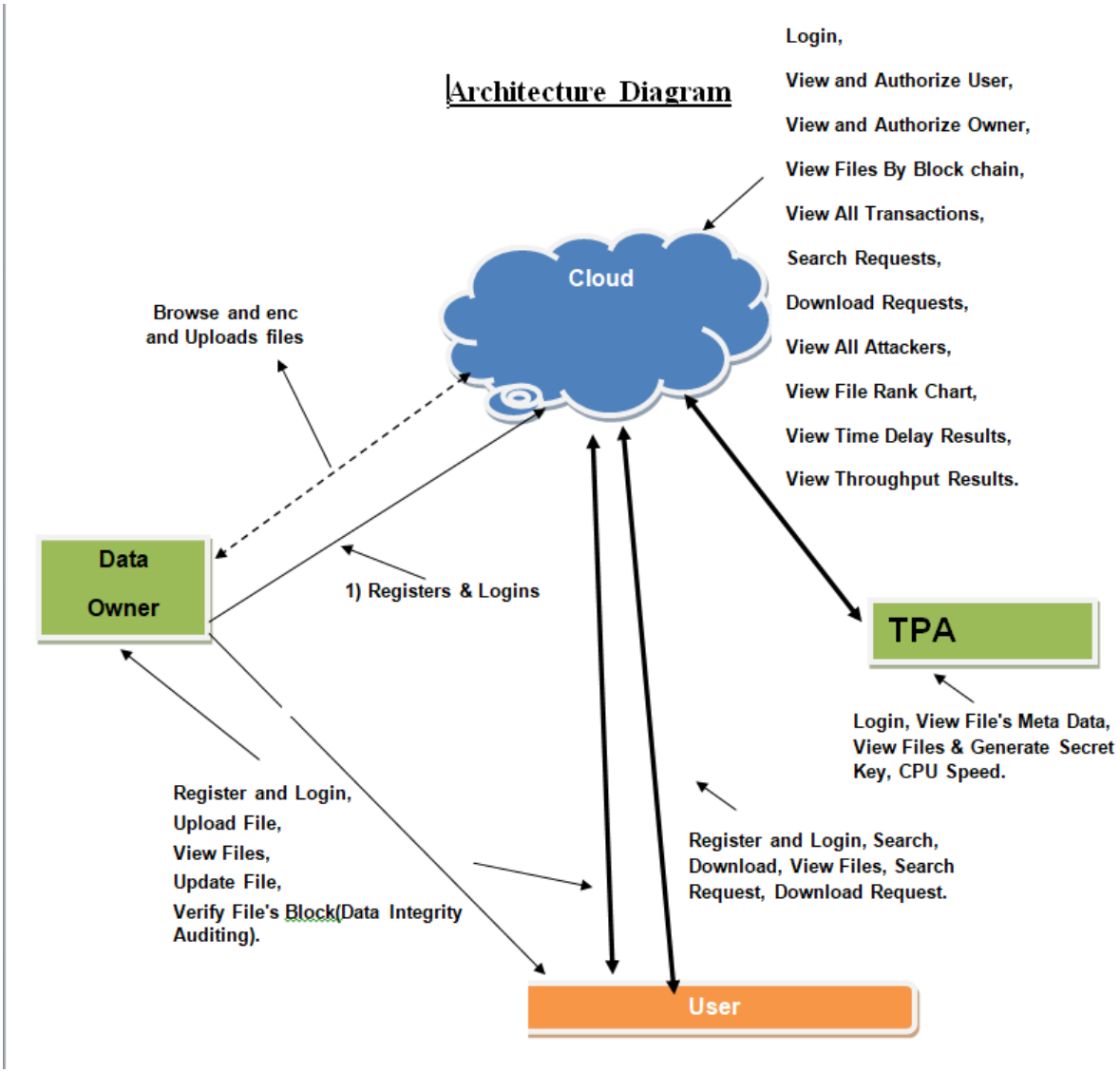

Fig 1:Architecture

4.RESULTS AND DISCUSSION

5.CONCLUSION

We offer Metadedup, which uses the power of indirection to achieve metadata deduplication. It considerably reduces the metadata storage overhead in encrypted deduplication while maintaining data and information confidentiality guarantees. Metadedup is extended for secure and fault-tolerant storage in a multi-server architecture; specifically, we present a distributed key management strategy that maintains security guarantees even when a number of servers are compromised. Metadedup is thoroughly evaluated in terms of

performance and storage efficiency. We demonstrate that Metadedup considerably reduces metadata storage space while having only a minor performance overhead when compared to network speed.

REFERENCES

[1] K. Hao, J. Xin, Z. Wang, and G. Wang, ``Outsourced data integrity

veri_cation based on blockchain in untrusted environment," *World Wide*

Web, vol. 23, no. 4, pp. 2215_2238, Jul. 2020.

[2] Y. Fan, X. Lin, G. Tan, Y. Zhang, W. Dong, and J. Lei, ``One secure data

integrity veri_cation scheme for cloud storage," *Future Gener. Comput.*

Syst., vol. 96, pp. 376_385, Jul. 2019.

[3] H. Wang and J. Zhang, ``Blockchain based data integrity veri_cation for

large-scale IoT data," *IEEE Access*, vol. 7, pp. 164996_165006, 2019.

[4] Z. Miao, C. Ye, P. Yang, R. Liu, B. Liu, and Y. Chen, ``A scheme for

electronic evidence sharing based on blockchain and proxy re-encryption,"

in *Proc. 4th Int. Conf. Blockchain Technol. Appl.*, Dec. 2021, pp. 11_16.

[5] F. Kefeng, L. Fei, Y. Haiyang, and Y. Zhen, ``A blockchain-based _exible

data auditing scheme for the cloud service," *Chin. J. Electron.*, vol. 30,

no. 6, pp. 1159_1166, Nov. 2021.

[6] K. He, J. Shi, C. Huang, and X. Hu, ``Blockchain based data integrity

veri_cation for cloud storage with T-Merkle tree," in *Proc. Int. Conf. Algo-*

rithms Archit. Parallel Process. Cham, Switzerland: Springer, Oct. 2020,

pp. 65_80.

[7] Y. Lei, Z. Jia, Y. Yang, Y. Cheng, and J. Fu, ``A cloud data access

authorization update scheme based on blockchain," in *Proc. 3rd Int. Conf.*

Smart BlockChain (SmartBlock), Oct. 2020, pp. 33_38.

[8] Y. Yuan, J. Zhang, W. Xu, and Z. Li, ``Identity-based public data

integrity veri_cation scheme in cloud storage system via blockchain," *J.*

Supercomput., vol. 78, pp. 8509_8530, Jan. 2022.

[9] S. Wang, D. Zhang, and Y. Zhang, ``Blockchain-based personal health

records sharing scheme with data integrity veri_able," *IEEE Access*, vol. 7,

pp. 102887_102901, 2019.

[10] A. Liu, Y. Wang, and X. Wang, ``Blockchain-based data-driven smart

customization," in *Data-Driven Engineering Design*. Cham, Switzerland:

Springer, 2022, pp. 89_107.

[11] K. Dhyani, J. Mishra, S. Paladhi, and I. S. Thaseen, ``A blockchain-based

document veri_cation system for employers," in *Proc. Int. Conf. Comput.*

Intell. Data Eng. Singapore: Springer, 2022, pp. 123_137.

[12] K. Xu, W. Chen, and Y. Zhang, ``Blockchain-based integrity veri_cation

of data migration in multi-cloud storage," *J. Phys., Conf. Ser.*, vol. 2132,

no. 1, Dec. 2021, Art. no. 012031.

[13] G. Xu, S. Han, Y. Bai, X. Feng, and Y. Gan, ``Data tag replacement

algorithm for data integrity veri_cation in cloud storage," *Comput. Secur.*,

vol. 103, Apr. 2021, Art. no. 102205.

[14] G. Xie, Y. Liu, G. Xin, and Q. Yang, ``Blockchain-based cloud data

integrity veri_cation scheme with high ef_ciency," *Secur. Commun. Netw.*,

vol. 2021, pp. 1_15, Apr. 2021.

[15] U. Arjun and S. Vinay, ``Outsourced auditing with data integrity

veri_cation scheme (OA-DIV) and dynamic operations for cloud data with

multi-copies," *EAI Endorsed Trans. Cloud Syst.*, vol. 7, no. 20, Jul. 2018,

Art. no. 169423.

[16] A. V. Ezhil, G. K. Indra, and K. Kulothungan, ``Auditable attribute-based

data access control using blockchain in cloud storage," *J. Supercomput.*, vol. 78, pp. 10772_10798, Jan. 2022

.