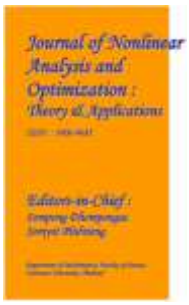
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2 : 2024
ISSN : **1906-9685**



ENSURING VACCINE SAFETY: ENCHANCED BLOCK CHAIN POWERS SPACE OPTIMIZED STORAGE

[1]K.Jaya Krishna, [2]Garlapati Uday kiran,

[1]Associate Professor, Department of Master of Computer Applications, QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

[2]PG Scholar, Department of Master of Computer Applications, QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

ABSTRACT

Data deduplication is a technique to eliminate duplicate data in order to save storage space and enlarge upload bandwidth, which has been applied by cloud storage systems. However, a cloud storage provider (CSP) may tamper user data or cheat users to pay unused storage for duplicate data that are only stored once. Although previous solutions adopt message-locked encryption along with Proof of Retrievability (PoR) to check the integrity of deduplicated encrypted data, they ignore proving the correctness of duplication check during data upload and require the same file to be derived into same verification tags, which suffers from brute-force attacks and restricts users from flexibly creating their own individual verification tags. In this paper, we propose a verifiable deduplication scheme called VeriDedup to addressthe above problems. It can guarantee the correctness of duplication check and support flexible tag generation for integrity check over encrypted data deduplication in an integrative way. Concretely, we propose a novel Tag-flexible Deduplication-supported Integrity Check Protocol (TDICP) based on Private Information Retrieval (PIR) by introducing a novel verification tag called note set, which allows multiple users holding the same file to generate their individual verification tags and still supports tag deduplication at the CSP. Furthermore, we make the first attempt to guarantee the correctness of data duplication check by introducing a novel User Determined Duplication Check Protocol (UDDCP) based

on Private Set Intersection (PSI), which can resist a CSP from providing a fake duplication check result to users. Security analysis shows the correctness and soundness of our scheme. Simulation studies based on real data show the efficacy and efficiency of our proposed scheme and its significant advantages over prior arts.

Index : deduplication, encrypted data, tdicp, pir, psi, csp

I. INTRODUCTION

Data deduplication, a technique widely employed in cloud storage systems, serves to optimize storage space and upload bandwidth by identifying and eliminating duplicate data. While offering significant benefits, the adoption of data deduplication has also introduced concerns regarding data integrity and user trust. Specifically, there is a risk of cloud storage providers (CSPs) exploiting deduplication processes for financial gain, potentially compromising user data integrity and privacy. Prior approaches to address these concerns have primarily focused on encryption and Proof of Retrievability (PoR) mechanisms, yet they have often overlooked the critical aspect of verifying the correctness of deduplication checks during data upload. In response to these challenges, this paper presents VeriDedup, an innovative verifiable deduplication scheme tailored to ensure the integrity of data deduplication processes within cloud storage infrastructures. VeriDedup introduces two key protocols: the Tag- flexible Deduplication-supported Integrity Check Protocol (TDICP) and the User Determined Duplication Check Protocol (UDDCP).TDICP leverages Private Information Retrieval (PIR) techniques to enable flexible tag generation for integrity checks over encrypted data. A novel verification tag, termed the note set, is introduced, empowering multiple users to generate individual verification tags while still facilitating tag deduplication at the CSP level. This approach not only enhances flexibility for users but also strengthens data integrity checks within the cloud storage environment.Furthermore, the paper introduces UDDCP, which is based on Private Set Intersection (PSI) principles, to ensure the accuracy of data duplication checks. By employing UDDCP, VeriDedup mitigatesthe risk of CSPs providing false duplication check results to users, thereby bolstering trust and confidence in cloud storage services. Security analysis conducted on VeriDedup confirms its correctness and robustness, while simulation studies based on real-world data

demonstrate its efficacy and efficiency compared to existing solutions. Overall, VeriDedup represents a significant advancement in addressing critical concerns surrounding data integrity, trust, and efficiency in cloud storage deduplication processes, offering a comprehensive and reliable solution for users and organizations leveraging cloud storage services.

II. LITERATURE SURVEY

1) Secure and efficient proof of storage with deduplication

AUTHORS: Q. Zheng, S. Xu.

The paper "Secure and efficient proof of storage with deduplication" by Q. Zheng and S. Xu, presented in the 2nd ACM Conference on Data and Application Security and Privacyin 2012, explores methods for securely and efficiently proving storage integrity while alsoconsidering data deduplication techniques. Deduplication is a process that reduces storage space by eliminating duplicate copies of data. This paper likely discusses how to combineproof of storage mechanisms with deduplication to ensure both security and storage efficiency in data management systems.

2) A secure two-phase data deduplication scheme

AUTHORS: P. Meye, P. Raipin, F. Tronel, E. Anceaume.

The paper titled "A secure two-phase data deduplication scheme" presents a method for efficiently and securely removing duplicate copies of data in storage systems. Data deduplication is an important technique used in various applications to save storage space and reduce redundancy. The authors propose a two-phase approach to data deduplication, focusing on both performance and security aspects. The first phase involves identifying duplicate data blocks using techniques such as content-based hashing or fingerprinting. In the second phase, a secure mechanism is employed to handle the deduplication process while ensuringdata integrity and confidentiality. By implementing a secure two-phase data deduplication scheme, the proposed method aims to address challenges related to data redundancy, storage optimization, and data security in modern computing environments. This research contributes to the development of efficient and reliable data management techniques for large-scale storage systems.

3) Encrypted data management with deduplication in cloud computing

AUTHORS: Z. Yan, M. Wang, Y. Li, A.V. Vasilakos.

The paper titled "Encrypted data management with deduplication in cloud computing," published in IEEE Cloud Computing in April 2016, addresses a critical issue in cloud computing: ensuring data security and efficiency through encryption and deduplication techniques. Encryption is used to protect sensitive data from unauthorized access, while deduplicationaims to eliminate redundant data by storing only unique data instances. The combination of these techniques is crucial for managing large volumes of data in cloud environments securely and cost-effectively. The authors likely discussed various encryption and deduplication strategies specifically tailored for cloud computing environments. These strategies may include efficient key management for encrypted data, secure deduplication algorithms to avoid data leaks, and performance optimizations to minimize overhead.

4) Game theoretical analysis on encrypted cloud data deduplication

AUTHORS: X. Liang, Z. Yan, X. Chen, L.T. Yang, W. Lou, Y. T. Hou

The paper titled "Game theoretical analysis on encrypted cloud data deduplication," published in October 2019, delves into the realm of cloud data management and security. Encrypted cloud data deduplication refers to the process of identifying and eliminating redundant data while ensuring data confidentiality through encryption. The authors employ game theory, a mathematical framework for analysing strategic interactions between different entities, to explore the dynamics of encrypted cloud data deduplication. By applying game theory, they aim to understand the incentives and strategies of various stakeholders, such as cloud service providers and data owners, in thecontext of deduplication while maintaining data privacy. This research is significant as it addresses the challenges of optimizing data storage efficiency and security in cloud environments. By using game theoretical analysis, the paper contributes to advancing strategies and protocols for encrypted cloud data deduplication, which is crucial for enhancing data privacy and resource utilization in cloudcomputing systems.

5) Investigating the adoption of hybrid encrypted cloud data deduplication with game theory

AUTHORS: X. Liang, Z. Yan, R. H. Deng, Q. Zheng.

The paper titled "Investigating the adoption of hybrid encrypted cloud data deduplication with game theory" explores the use of game

theory in enhancing the efficiency and securityof cloud data deduplication processes. Cloud data deduplication aims to reduce storage overhead by identifying and eliminating redundant data chunks across multiple users or files. The term "hybrid encrypted" suggests a combination of encryption techniques used in cloud environments, where sensitive data is encrypted to ensure privacy and security. Game theory, a mathematical framework for modeling strategic interactions among rational decision-makers, is employed in this context to analyze decision-making strategies among different entities involved in cloud data deduplication. The authors likely investigate various aspects such as the trade-offs between storage savings and computational overhead, the impact of different encryption schemes on deduplication efficiency, and strategies to incentivize data ownersto participate in deduplication processes while maintaining data confidentiality.

III. PROBLEM STATEMENT

The user then checks the integrity of the stored file by verifying the response. However, existing PoR solutions mainly aim to improve the performance at the user side and assumethat the CSP has infinite computation and storage resources. While, in practice, the CSP performs data deduplication in order to achieve the most economic usage of its storage. Unfortunately, existing solutions mentioned above are incompatible with deduplication. This is because the verification tags of these schemes are created with user individual used private keys unknown to each other, thus different verification tags are generated, given the same file held by different users.

3.1 Existing System Disadvantages:

we propose a note insertion mechanism based on PIR to let the data holder insert a specific set called note set that contains several randomized bit sequences, which conform to a hidden function, as verification tags into the encrypted blocks of a uploaded file. The data owners/holders who are proved to have the ownership of the corresponding blocks can verify the integrity of the uploaded blocks through a challenge on whether the notes are conform to the hidden function. Attention need be given that the verification tags generated by multiple data holders with various notes can also be deduplicated in VeriDedup, so that the CSP will no longer be required to maintain multiple pieces of verification tags from the same block of different data

holders for integrity check, which reduces storage consumption of performing deduplication.

IV. PROPOSED SYSTEM

Specifically, the main contributes of this project are summarized as below: We propose a novel protocol named TDICP based on PIR to check the integrity of uploaded files in the CSP with deduplication employed. TDICP allows users to generate their own individual verification tags for integrity check while the verification tags can alsobe deduplicated at the CSP although different. We propose another novel protocol named UDDCP to guarantee the correctness of duplication check based on PSI, so that the CSP is impossible to cheat the user to pay for unused storage space due to deduplication. We construct a novel deduplication scheme called VeriDedup that contains the above two novel protocols and other essential properties, such as PoW and data access key assignmentby re-shaping our previous scheme in order to overcome its shortcomings regarding integrity and duplication proof. We prove the security of TDICP and UDDCP by constructing several games and conduct both theoretical analysis and experimental simulation to evaluate their performance. Our results show their efficacy and efficiency.

4.1 Proposed System Advantages:

To reduce the deduplication data for cloud. To improve the user accessibility. Cloud data security.

4.2 Proposed System Limitations:

Computational Cost: The system relies on cryptographic protocols like PIR (Privacy-Preserving Information Retrieval) and PSI (Private Set Intersection) for integrity and duplication checks. These protocols can be computationally expensive, impacting both the user and the cloud service provider (CSP). Scalability: As the data volume in the cloud storage grows, the verification process using VeriDedup might become less scalable. Managing a large number of verification tags and PSI operations could become cumbersome. Threat Model: The system focuses on preventing the CSP from cheating users on storage costs and data integrity. However, it might not be entirely secure against other threats like unauthorized data access or insider attacks within the CSP. Limited File types: VeriDedup might not be equally efficient for all file types. It might be better suited for specific data formats where deduplication is more effective (e.g., documents, backups). Proof of Work (PoW): The use of PoW can introduce additional overhead for users. Depending on the PoW scheme, it might

require solving complex puzzles, impacting user experience.

V. SYSTEM ARCHITECTURE

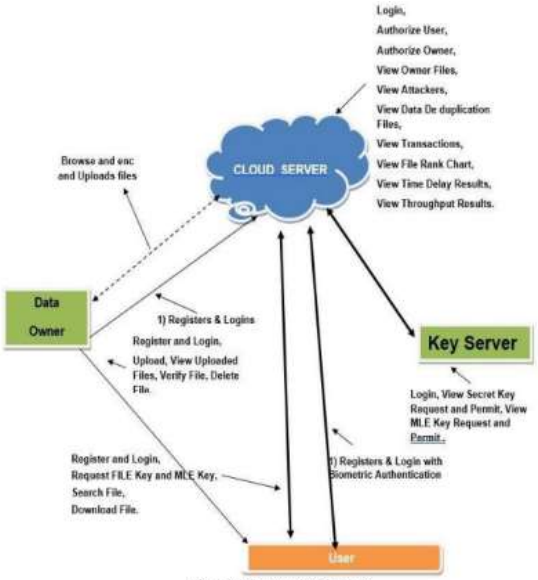


Fig 4.2.1 System Architecture

VI. METHODOLOGY

6.1 Integrity check:

It can guarantee the correctness of duplication check and support flexible tag generation for integrity check over encrypted data deduplication in an integrative way. Concretely, we propose a novel Tag-flexible Deduplication-supported Integrity Check Protocol based on Private Information Retrieval by introducing a novel verification tag called note set, which allows multiple users holding the same file to generate their individual verification tags and still supports tag deduplication at the CSP. But multiple users holding the same file stored at the cloud may create different tags as their willingness for data integrity check, which improves integrity check security by overcoming brute-force attacks, but impacts deduplication.

6.2 Duplication check:

Another security issue ignored by the previous literature is the correctness guarantee of data duplication check provided by the CSP. Severalschemes motivate the CSP to perform deduplication, but ignore that the CSP could cheat the users by providing a fake duplication check result. The reason is simple since the CSP can gain an extra profit by asking the usersto pay normal storage fee without granting a deserved discount while performing deduplication to save storage space.

6.3 Private information retrieval:

The TDICP explores a new verificati on tag called note set in which each note isa randomized bit sequence that is conform to a function f. The note set is inserted into the files based on Private Information Retrieval.Meanwhile, the UDDCP explores a new challenge and response mechanism based on Private Set Intersection to let the user instead of the CSP tell whether the file is duplicate first, so that the CSP cannot cheat the user on the result of duplication check during file upload.

6.4 Data deduplication

VeriDedup follows the construction of our previous deduplication scheme and improves it by using PSI and PIR to ensure both data integrity and duplication check correctness over encrypted data deduplication. Specifically, compared with previous work, we introduce a PSI based challenge and response mechanism to the duplication check procedure in order to let the data holder first tell whether the uploaded blocks are duplicate or not instead of the CSP. In addition, we employ AA to verify the computations of the CSP during the duplication check, so that the CSP cannot cheat the users to upload data blocks that have been stored already.

VII. ALGORTHIMS

Algorithm 1 Submit Merkle Tree Root

Require: the packing time *date*; the Merkle tree root *root*;
Ensure: execution status `ret_code` $\in \{0, 1\}$;
1: get the smart contract table object T;
2: create entry object e;
3: assign values to e, $e.Date \leftarrow date$, $e.Proof \leftarrow root$;
4: set the remaining attributes as empty;
5: `ret_code` $\leftarrow T$.insert($date, e$);
6: **return** `ret_code`;

Algorithm 2 Submit Signatures and Parameters

Require: the table key *date*; aggregate signature (R, s); large prime number p, q; the generator g of the cyclic group $\mathbb{G}$; the group public key X;
Ensure: execution status `ret_code` $\in \{0, 1\}$;
1: get table object T;
2: create entry object e;
3: assign signature values to e;
4: set the *condition* as $entry.Key == date$;
5: select entry from T where $entry.Key$ equals $date$;
6: `ret_code` $\leftarrow T.update(date, e, condition)$;
7: **return** `ret_code`;

VIII. EXPECTED RESULTS

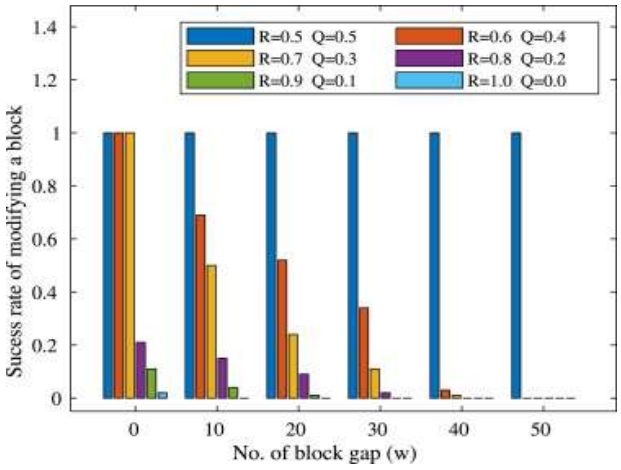


Fig 1 : The success rate of modifying with any transaction over blockchain network to attempt modification attack over vaccine supply chain.

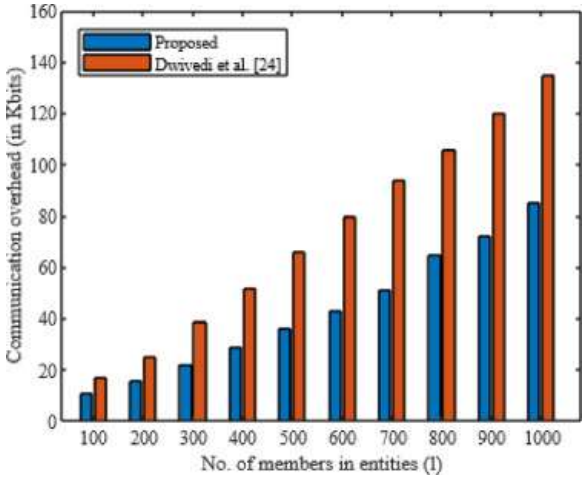


Fig. 2 : Comparison of communication overhead.

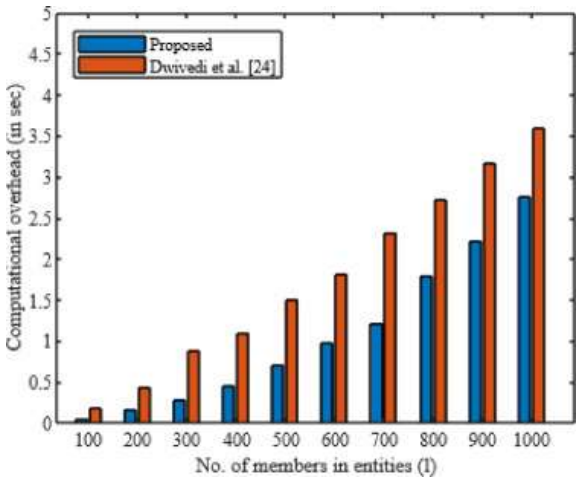


Fig 3 : Overall computational overhead of VaccineChain.

IX. CONCLUSION

VeriDedup to check the integrity of an outsourced encrypted file and guarantee the correctness of duplication check in an integrated way. The integrity check protocol TDICP of VeriDedup allows multiple data holders to verify the integrity of their outsourced file with their own individual verification tags without interacting with the data owner. On the other hand, we employed a novel challenge and response mechanism in the duplication check protocol UDDCP of VeriDedup to let the data holder instead of the CSP first tell whether a file is duplicate in order to guarantee the correctness of duplication check. Security and performance analysis show that VeriDedup is secure and efficient under the described security model. The result of our computer simulation further shows its efficiency compared with highly related prior arts.

X. FUTURE ENHANCEMENT

Dynamic Data Updates: Currently, VeriDedup might not support efficient updates to existing data. Developing a mechanism for users to securely update their data while maintaining integrity and deduplication benefits would be valuable.

Enhanced Privacy: While VeriDedup protects data integrity, exploring techniques like homomorphic encryption could enable users to perform limited computations on their outsourced data without compromising confidentiality.

Decentralized Storage Integration: Integrating VeriDedup with decentralized storage solutions could offer additional benefits like distributed trust and censorship resistance. However, ensuring compatibility and maintaining efficiency within decentralized architectures would be crucial.

Performance Optimization for Specific Workloads: VeriDedup demonstrates overall efficiency. However, tailoring the cryptographic protocols to specific data types (e.g., scientific datasets or multimedia) could further optimize performance for those workloads.

Lightweight Client Implementations: Developing lightweight client-side implementations of VeriDedup protocols could benefit resource-constrained devices and improve overall system scalability.

XI. REFERENCE

[1] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, "Verifiable computation over large database with incremental updates," IEEE Trans. Computers, vol. 65, no. 10, pp. 3184–3195, 2016.

[2] M. Gerla, J.Wang, and G. Pau, "Pics-on-wheels: Photo surveillance in the vehicular cloud," International Conference on Computing, Networking and Communications, pp. 1123–1127, 2013.

[3] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 9, pp. 2386–2396, 2014.

[4] H. Yuan, X. Chen, T. Jiang, X. Zhang, Z. Yan, and Y. Xiang, "Dedupdum: Secure and scalable data deduplication with dynamic user management," Inf. Sci., vol. 456, pp. 159–173, 2018.

[5] H. Huang, X. Chen, Q. Wu, X. Huang, and J. Shen, "Bitcoinbased fair payments for outsourcing computations of fog devices," Future Generation Comp. Syst., vol. 78, pp. 850–858, 2018.

[6] IDC. (2014) The digital universe of opportunities: Rich data and the increasing value of the internet of things. [Online]. Available: https://www.emc.com/leadership/digitaluniverse/2014iview/index.htm

[7] W. J. Bolosky, S. Corbin, D. Goebel, and J. R. Douceur, "Single instance storage in windows 2000," in Conference on Usenix Windows Systems Symposium, 2000.

[8] Dropbox. (2007). [Online]. Available: http://www.dropbox.com

[9] GoogleDrive. (2012). [Online]. Available: http://drive.google.com

[10] Memopal. (2018). [Online]. Available: http://www.memopal.com

[11] Netapp. (2008) Netapp deduplication helps duke institute for genome sciences and policy reduce storage requirements for genomic information by 83 percent. [Online]. Available: http://www.netapp.com

[12] M. Dutch, "Understanding data deduplication ratios," in SNIA Data Management Forum, 2008, pp. 1–13.

AUTHOR PROFILE:

[1] Mr. K. Jaya Krishna, currently working as an Associate Professor in the Department of Master of Computer Applications, QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He did his MCA from Anna University, Chennai, M.Tech

(CSE) from JNTUK, Kakinada. He published more than 10 research papers in reputed peer reviewed Scopus indexedjournals. He also attended and presented research papers in different national and international journals and the proceedings were indexed IEEE. His area of interest is Machine Learning, Artificial intelligence, Cloud Computing and Programming Languages.

[2] Mr. Garlapati Uday Kiran, currently pursuing Master of Computer Applications at QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh. He Completed B.Sc. in Computer Science from BA & KR Degree College, Ongole, Andhra Pradesh. His areas of interests are Cloud Computing & Machine learning.