Migration Letters

Volume: 21, No: S5 (2024), pp. 1620-1627

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Privacy Persevering Data Deduplication And Proxy Re-Encryption Scheme For Secure Cloud Storage

P. Anantha Prabha[1], M. Asif Abdul Khader[2], K.V. Hamsanand[3], R. Hari Balaji[4]

Abstract –
Proxy re-encryption (PRE) emerges as a viable result for secure cloud data sharing applications, ensuring a streamlined and secure conversion of ciphertext. One of the most significant characteristics of cloud computing is storage, which lets users assign data storage to the cloud and allows safe sharing with authorized users. Existing secure data exchange schemes prioritize characteristics such as data secrecy, tag consistency, and access control for enhanced security and privacy. However, none of the currently known schemes can simultaneously fulfil all four requirements. To overcome this limitation, the proposed system utilize identity-based conditional PRE method that effectively supports both authorization processes and ciphertext evolution. The System also introduces an efficient secure deduplication technique and Proxy Re-Encryption (PRE) with access control defined by the user. A breakthrough solution that strikes a delicate balance between minimizing data duplication and safeguarding the privacy of cloud users is introduced. This is achieved through a novel approach that empowers Cloud Service Provider (CSP) to grant authorization for accessing data on behalf of data owners., all while maintaining strict confidentiality through secure deduplication techniques based on symmetric key cryptography. The rigorous security analysis confirms the robustness of this approach, and extensive simulations demonstrate its superior performance in terms of computational efficiency, communication overhead, storage optimization, and overall deduplication effectiveness, outperforming competing schemes.

Keywords: *Cryptography, Cloud Computing, proxy re-encryption, symmetric key, Deduplication.*

I. INTRODUCTION

Cloud computing provides users with virtually limitless "virtualized" resources delivered as services over the entire Internet, abstracting away concerns related to platforms and implementations. Cloud service providers currently provide economical access to storage that is highly available and computing capabilities on a massive parallel scale. As the adoption of cloud computi[1]ng continues to grow, a significant volume of data is stored in the cloud and distributed among users, with access to the stored data being regulated by specific credentials. Effectively managing the growing volume of data presents a significant challenge for Cloud Service Provider(CSP). Despite the numerous benefits provided by data deduplication, concerns arise regarding security and privacy. The vulnerability of users' sensitive data to both insider and outsider attacks is a significant issue. Traditional encryption, which ensures data secrecy, faces incompatibility with data deduplication. The objective was to tackle the concern, wherein a proxy possessing the delegation token could delegate the entire ciphertext of the delegator.

By employing privacy-preserving approaches, the use of PRE in this data deduplication improves the security and privacy of data stored. In the case of data deduplication, it may involve identifying and eliminating redundant copies of data. Proxy

[1,2,3,4] Department of Computer Science and Engineering, Sri Krishna College of Technology.

re-encryption could be utilized to enable secure data sharing or access control highlighting how privacy is maintained throughout the process. This involves encryption techniques that protect sensitive information from unauthorized access or exposure. The remaining of the paper is as follows. Section II presents a literature survey for re-encryption methods and data duplication in the system. Section III gives the proposed system with extensive methodologies. Section IV depicts the experimental setup and result analysis of the proposed system. Section V has a conclusion regarding the overall proposed system.

II . LITERATURE REVIEW

Recently, there's has been an increased awareness on information security in cloud computing due to the growing demand for services that process and store data. Hashem et al.[1,17] provided an overview of cryptography algorithms to identify the most effective ones for securing data on cloud servers. With an emphasis on symmetric key cryptography, the review included both asymmetric and symmetric key cryptography. They discussed the difficulties with cloud computing security and emphasized that cloud service providers must guarantee security as part of their quality of the service offering. Standardized encryption/decryption techniques are desperately needed to safeguard stored data in light of the inherent security risks found in cloud environments, such as IAM and hijacking. Because cloud service providers have unlimited access to data and keys, confidence in them has been called into doubt. In response, Pradeep et al.[9] suggested a framework for safe file sharing together with a mechanism intended to prevent key leakage. The focus was on utilizing asymmetric keys supplied by a reliable third party within the cloud to achieve security.

In addition, OutFS, a user-side encrypted file system designed for transparent encryption of outsourced data on the cloud, was introduced by Khashan et al.[11]. With well thought-out key management, OutFS used a hybrid encryption strategy that included symmetric and asymmetric techniques. The incorporation of an identity-based encryption system (IBE) into OutFS was emphasized as a means of providing strong data sharing security. Performance investigation demonstrated the effectiveness of OutFS, demonstrating significant throughput rates for both writing and reading files that are outsourced. Security investigation verified that OutFS is resistant to a variety of threats, such as eavesdropping and brute-force attacks. To prevent key loss or theft and to ensure the protection of uploaded data, Huthaifa et al.[15] discussed an algorithm addresses the large storage space that cloud computing offers, which is accessible via the internet, emphasizing the significance of protecting personal and corporate data by using genetic and XOR algorithm techniques.

A Homomorphic conditional proxy re-encryption system was proposed by Chunguang et al.[2]. The system enables cloud computing and encrypted data sharing by combining homomorphic encryption with conditional proxy re-encryption. In order to control and minimize decryption complexity, the system is incorporated into a homomorphic proxy re-encryption mechanism during the scheme's creation. For data protection, users encrypted their data before uploading, Linmei Jiang et al.[3] defined a secure storage system with encrypted data sharing using conditional re-encryption technology. The addition or removal of users from sharing groups was made possible by this approach, which allowed for both broadcast and dynamic data sharing without changing the encryption public keys. The introduction of conditional PRE (C-PRE), particularly sender-specified PRE (SS-PRE), enabled secure decryption abilities.

Peng Zeng et al.[5,16] provided a formal definition and security model for SS-PRE, presenting the construction of secure SS-PRE schemes (IND-CPA and IND-CCA). The SS-PRE scheme surpassed traditional C-PRE schemes in terms of efficiency, exhibiting benefits in terms of calculation cost and ciphertext size. The conversion of ciphertexts under various access regulations was made easier by the attribute-based proxy re-encryption, which extended the use of traditional proxy re-encryption to the attribute-based setting. The key-policy attribute-based proxy re-encryption proposed by Chunpeng Ge et al. [6] has been shown to be secure against chosen-ciphertext attacks in the model and can handle any monotonic access structures on users' keys. Additionally, Revocable Proxy Re-Encryption with Ciphertext Evolution is a system proposed by Shimao Yao et al. [12] that

is intended to make cloud data exchange safe and effective. Key elements of the method were maintained, including with fine-grained authorization, protection against Chosen-Ciphertext attacks (CCA), multi-use, and constant ciphertext length. Especially, it allowed authorization revocation through developing ciphertext and permitted ciphertext updating to match new keys.

To enhance confidentiality, Juntao et al.[13] suggested a digital content management system built on Blockchain technology. With the use of Ethereum smart contracts, they achieved flexible and equitable decryption key exchanges by combining Multiauthority Ciphertext-Policy Attribute-Based Encryption (MA-CPABE) with proxy re-encryption. In Blockchain-based IoT data sharing, Kwame et al.[14] proposed an approach for secure data sharing utilizing edge devices for intensive computations and information-centric networking for efficient content delivery.

The exponential growth of digital data in cloud storage systems was a serious concern, necessitating the use of efficient solutions like deduplication to reduce the burden of duplicate data. Ravneet Kaur et al.[4] performed a thorough examination of existing deduplication techniques and taxonomies, including approaches based on cloud data storage and those intended for text and multimedia data. They also addressed about the difficulties of deduplication, especially when it comes to managing and storing encrypted data. Building on these analyses, Zheng Yan et al.[10] proposed a distinct data storage control strategy that includes duplication management as well as access control across several Cloud Service Providers. Myungkeun Yoon et al.[7] also presented a constant-time chunking algorithm for eliminating redundancy in network packets, which greatly improved throughput compared to previous schemes. However, dynamic ownership and key management are frequent problems for already-existing schemes. Hyunsoo Kwon et al. [8] suggested a safe deduplication system that incorporates trustworthy key management based on pairing-based encryption and supports dynamic ownership changes without requiring extra costs in order to address these issues.

III. PROPOSED SYSTEM

The proposed system employs a comprehensive strategy designed to address issues with data ownership, making use of PRE to effectively handle the storage of encrypted data with deduplication. The primary focus is to address problems associated with duplicate data when data owners upload their data, ensuring a data storage environment that is both efficient and safe. Moreover, the method ensures reliable data deduplication efficiency, which makes it very suitable for the dynamic and variable data sizes found in cloud computing environments.

A. System Design:

The proposed scheme contains three entities as shown in Figure.1, the user, the Cloud Server(CS), and the Proxy Server. The system requires a trusted proxy to run which will be owned by the cloud computing service. The user trusts the proxy sever to re-encrypt the file without the system knowing how to decrypt it with the given key.

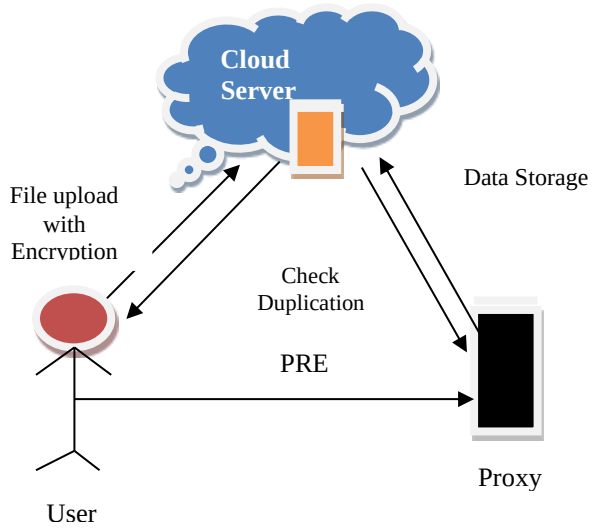

Fig.1. The system model

In Figure 1, the user possesses the data intended for storage in the cloud. Within the storage system, a duplication check is done prior to storing the data. The data is encrypted, then securely uploaded into the storage system via proxy re-encryption if no duplicates are found.

B. Data Encryption and check Duplication

The data owner chooses specific files from various sources for encryption. To execute this process, the data owner needs to possess their encryption key for authorization purposes. The proposed system employs the Advanced Encryption Standard (AES) as its primary encryption algorithm. The symmetric encryption technique known as AES is known for its effectiveness and security. AES is frequently used by data owners for encryption before sending sensitive data to the cloud. The algorithm's strength lies in its ability to secure data with a symmetric key, ensuring confidentiality during data transmission and storage. The equation represents AES encryption: Cipher Text(Z) = AES(X, K) where, X is Plain Text , and K is the encryption Key

Before initiating the file encryption process, it is imperative for the data owner to ensure that the file is not redundantly stored in the storage system. To accomplish this, the proposed system integrates a Content-Based Chunking algorithm, a sophisticated technique that transcends traditional boundaries by breaking files into fixed-size or variable-size chunks based on their content rather than their physical structure. This method discerns similarities within the actual data, enabling the swift and effective detection of duplicate information not only within a single file but also across multiple files.

C. Proxy Re-Encryption

PRE serves as a transformative layer in the proposed system's security architecture. In the context of data sharing within the cloud, PRE allows a designated proxy entity to rewrite ciphertext originally encrypted by a user. This rewriting process occurs without revealing the underlying plaintext, offering a secure means of data sharing.

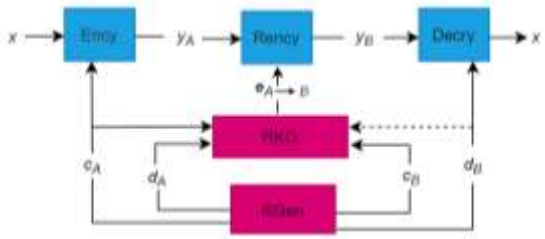

Fig.2. Proxy Re-encryption Scheme

A PRE scheme is represented as a set of algorithms (KGen, ReKGen, Encr, ReEncr, Decr):

• KGen(n) → (c_A, d_A). The public and secret key pair (c_A, d_A) of user A is generated by the key generation algorithm KGen when it receives the security parameter 'n'.

• ReKGen(c_A, d_A, c_B, d_B) → $e_{A→B}$. The re-encryption key generation algorithm ReKGen generates a re-encryption key $e_{A→B}$ when it receives the pair of public and secret keys (c_A, d_A) for user A and (c_B, d_B) for user B as inputs.

• Encr(c_A, x) → z_A. The encryption algorithm Encr creates a ciphertext z ∈ Z with the public key c_A and a message x ∈ X as input.

• ReEncr($e_{A→B}$, z_A) → z_B. Given a re-encryption key $e_{A→B}$ and a ciphertext z_A ∈ Z as input, the re-encryption algorithm ReEncr produces a second ciphertext z_B ∈ Z or indicates that z_A is invalid.

X and Z represents plain text and cipher text, respectively.

D. Decryption:

When authorized users access the shared data, they employ AES decryption. The symmetric key, securely transmitted through the PRE process, is used to decrypt the ciphertext. This seamless decryption process ensures that authorized users can retrieve the original plaintext data securely, completing the secure data sharing cycle within the proposed system. The equation represents AES decryption: Decrypt Text(D) = $AES^{-1}(Z,K_{user})$ where C is CipherText , and K_{user} is user's private decryption key.

In addition to safeguarding data confidentiality, the implementation of AES decryption and secure symmetric key transmission guarantees the integrity of the shared data by preventing unauthorized changes or tampering during the retrieval process.

IV. EXPERIMENTS AND RESULTS

The Proposed System employed Java to simulate a cloud/client/proxy system for evaluating the effectiveness of data storage and fine access in cloud storage solution. The technique is versatile, especially for encrypted and secure data storage. The proposed secure data storage architecture is designed for scenarios where multiple individuals, such as financial specialists, data analysts, and senior executives with authority, are interested in the same or similar data. It caters to the needs of individuals within the same corporate organization who wish to store business data securely in the cloud. Moreover, the solution facilitates the secure transmission of corporate data through channels while preserving privacy.

A. Data Handling

In the system, the duration for generating the PRE is evaluated, with the exclusion of the I/O time needed for loading or storing data to disk. Data Handling, a crucial element impacting the quality of service (QoS) for clients, is handled by the data owner. The experimental results are illustrated in Figure 3. To enhance clarity, the y-axis is normalized to the time cost with a base of 20, where the time is represented in milliseconds required for conducting the data pre-processing.

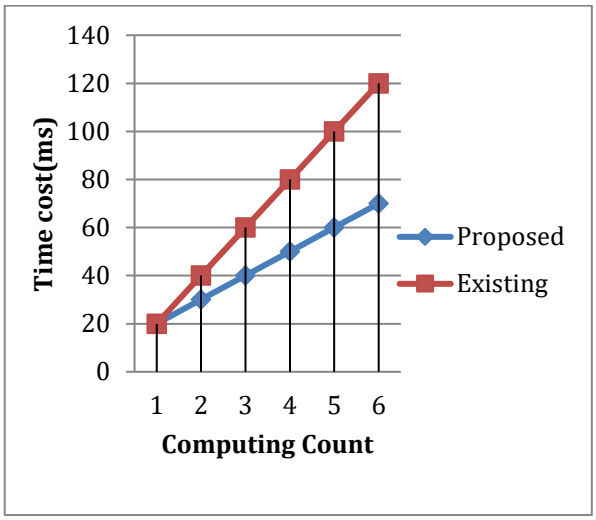

Fig.3. Computing Count with Deduplication and PRE

Figure 3 presents a comparison of the time cost between direct computing and proposed approach. It is evident that the time cost associated with computing involving Deduplication and PRE is significantly lower than that of existing methods. These observations collectively demonstrate the effectiveness of proposed outsourcing scheme in reducing the computation costs for users' devices.

B. Result Analysis:

Predominantly, the performance distinctions between the approach and other schemes are evaluated.

Java version	1.8.11 (Java 8)
OS	Windows 11
Processor	Intel® Core™ i5-1135U CPU @ 2.60 GHz
Ram	16.00 GB

Table I. System Specification

A comparison of the techniques used for encryption, secret key creation, and decryption is shown in the following Table II. The time it takes for these algorithms to process data is used to evaluate the system's performance. The performance evaluation is produced by calculating the average time form all systems involved in the testing procedure.

Algorithm	**Existing**	**Proposed**
Encryption	59.509ms	58.647ms
Secret Key Generation	32.247ms	32.247ms

Decryption	89.759ms	79.247ms

Table II. Comparison of Algorithm's Calculation Time.

The testing process involves various systems, and the average time across all systems is calculated to provide a comprehensive performance assessment. The chosen computer specifications Table I, provide a context for understanding how algorithm performance aligns with different system configurations. This insight aids in evaluating the proposed scheme's adaptability and effectiveness across diverse computing environments. The algorithms of both the existing scheme and the proposed scheme undergo performance assessments, with the results presented in Table II. In all comparisons, it is evident that the suggested scheme consistently surpasses the performance of the existing.

C. Security Analysis

In tackling the issue of the privileged differential, the proposed approach tries to evaluate security through two major aspects: duplicate check authorization and data confidentiality. Proxy Re-Encryption (PRE) undergoes rigorous security analysis to ensure robust protection of sensitive data in a cloud computing environment. The security assessment involves evaluating the resilience of the PRE scheme against various cryptographic attacks and potential vulnerabilities. The secure transmission and delegation of decryption capabilities are verified to prevent unauthorized access or information leakage. Furthermore, the PRE scheme is assessed for its ability to maintain the integrity and authenticity of re-encrypted data during the proxy re-encryption process. The comprehensive security analysis ensures that Proxy Re-Encryption stands as a trustworthy and effective cryptographic tool, meeting the stringent security requirements for secure data sharing and management in cloud environments.

V. CONCLUSION

Handling encrypted data through deduplication is essential for the successful development of a cloud storage service, especially when dealing with extensive data storage. the suggested system combines advanced encryption standards (AES) with PRE to address the issue of data duplication and improve the security of data that is kept on cloud servers. The system guarantees an elevated level of cryptographic security by using AES as the encryption/decryption technique, protecting sensitive data during both the storage and transmission phases. Proxy Re-Encryption adds an additional layer of security and permits regulated and safe data sharing across the cloud ecosystem. Additionally, the use of advanced algorithms and methodologies to execute a data duplication check guarantees that redundant data is effectively handled and recognized, maximizing storage capacity and lowering the possibility of storing similar data. Overall, the proposed system stands as an effective and comprehensive system for secure data management and sharing in cloud environments, addressing key challenges related to privacy, duplication, and cryptographic integrity. The goal of future research will be to improve and refine PRE schemes such that they are ideal for distributed environment.

REFERENCES

[1] Hashem and H. Ramadan. "Using Cryptography Algorithms to Secure Cloud Computing Data and Services", Amer. J Eng. Res. (AJER), vol. 6, no. 10, pp. 334-337, 2017.

[2] Chunguang Ma, Juyan Li and Weiping Ouyang. "Lattice-Based Identity-Based Homomorphic Conditional Proxy Re-Encryption for Secure Big Data Computing in Cloud Environment", IEEE , 2017.

[3] Linmei Jiang and Donghui Guo(senior Member, IEEE). "Dynamic Encrypted Data Sharing Scheme Based on Conditional Proxy Broadcast Re-Encryption for Cloud Storage", IEEE Access, 2017.

[4] Ravneet Kaur, Inderveer Chana and Jhilik Bhattacharya. “Data deduplication techniques for efficient cloud storage management”, IEEE, 2017.
[5] Peng Zeng And Kim-Kwang Raymond Choo,(Senior Member, IEEE) “A New Kind of Conditional Proxy Re-Encryption for Secure Cloud Storage”, IEEE, 2018.
[6] Chunpeng Ge,Willy Susilo,Liming Fang, Jiandong Wang and Yunqing Shi “A CCA-secure key-policy attribute-based proxy re-encryption in the adaptive corruption model for dropbox data sharing system”, IEEE, 2018.
[7] Myungeun Yoon, eT al. “A Constant-Time ChunkingAlgorithm for packet level Deduplicatiom”, 2018.
[8]Hyunsoo Kwon, Changee Hahn, Kyungtaae Kang And Junbeom Hur “Secure Deduplication with reliable and revocable key management in Fog Computing”,2018.
[9] K. V. Pradeep et al., "An Efficient Framework for sharing a file in a secure manner using Asymmetric key Distribution Management in Cloud Environment", journal of computer networks and communications HINDAWI Limited, PP. 1-8, 2019.
[10] Zheng Yan, Lifang Zhang Di, Wenxiu Ding, Qinghua Zheng “Heterogeneous Data Storage Management with Deduplication in Cloud Computing”, IEEE Access, 2019.
[11] O. A. Khashan, "Secure outsourcing and sharing of cloud data using a user-side encrypted file system", IEEE Access, vol. 8, pp. 210855-210867, 2020.
[12] Shimao Yao , Ralph voltaire J. Dayot, Hyung-Jin Kim, And In-Ho Ra(Member, IEEE). “A Novel Revocable and Identity-Based Conditional Proxy Re-Encryption Scheme With Ciphertext Evolution for Secure Cloud Data Sharing”, IEEE, 2021
[13] Juntao Gao, H. Yu, X. Zhu, and X. Li, “Blockchain-based digital rights management scheme via multiauthority ciphertext-policy attribute-based encryption and proxy re-encryption,” IEEE Syst. J., vol. 15, no. 4, pp. 5233–5244, Dec. 2021.
[14] Kwame Opuni-Boachie Obour Agyekum , Qi Xia , Emmanuel Boateng Sifah , Christian Nii Aflah Cobblah , Hu Xia and Jianbin Gao. “A Proxy Re-Encryption Approach to Secure Data Sharing in the Internet of Things Based on Blockchain”, IEEE, VOL. 16, NO. 1, 2022.
[15] Huthaifa A. Al Issa and Al-Balqa. “Encryption and Decryption Cloud Computing Data Based on XOR and Genetic Algorithm”, 2022.
[16] Radhini, M. P., P. Anantha Prabha, and P. Parthasarathi. "Distributed Attribute based Encryption for Secure Sharing of Personal Medical Records in Cloud."2014.
[17] M.P. Radhini, P. Ananthaprabha, P. Parthasarathi, “Secure Sharing of Medical Records using Cryptographic Methods in Clouds”, Volume 3, Issue 4, April 2014.