IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Novel Secure Cloud Storage Paradigm Based On Intelligent Deduplication, Attribute-Flexible Encryption, And Immutable Blockchain Auditing

DR. A SOMASUNDARAM[1], JIBIN JOY[2],

[1]Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India, Somasundaram.a@gmail.com

[2]Research Scholar (Ph.D.), Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India, jibinjoysamuel@gmail.com

ABSTRACT

In this paper, FuzzyDedup-Chain, a new unified framework, is presented to combine adaptive block-level deduplication, Fuzzy Identity-Based Encryption with Proxy Re-Encryption (Fuzzy IBE-PRE), and blockchain-based auditing to achieve better efficiency, security, and transparency of cloud data management. Conventional cloud systems find it hard to optimize storage, provide flexible access control and auditability akin to the impregnable qualities of the Tasers. In order to overcome these shortcomings, the given architecture proposes an Adaptive-Flat-Block-Deduplication (AFBD) mechanism that dynamically chooses the best block size to minimize redundancy without causing a high cryptographic overhead. It has a Selective-Re-Encryption (SR-PRE) mechanism to delegate access rights by re-encrypting only metadata reference not an entire data block, which greatly lowers the re-encryption latency. More so, a PoNR protocol-backed by a blockchain guarantees clear and verifiable reporting of deduplication efficiency without revealing plaintext content. The audit layer uses aggregate digital signatures; this allows scalability of storage events and verification that is tamper proof. The combination of these elements forms secure, efficient, and auditable cloud architecture that can handle the current threats, minimizes the cost of storage, and fine-grained access control. The suggested model can be regarded as a full-fledged solution to the next-generation cloud-based environments that require operational efficiency and solid security assurances.

Keywords: Cloud Security; Deduplication; Fuzzy Identity-Based Encryption; Proxy Re-Encryption; Blockchain Auditing; Data Sharing; Access Control; Aggregate Signatures; Proof-of-Non-Redundancy.

1. INTRODUCTION

The development of cloud computing has been high and this has affected the manner in which organizations store, manipulate and share information online. With the continuously growing amounts of data, cloud infrastructures are being strained to offer good use of storage, good data sharing, and auditing. Conventional cloud architectures would typically concentrate on these needs one by one, storage efficiency, in deduplication, security, in encryption, and in transparency, in log-based auditing. However, the lack of connection between all these layers results in redundant functionality, high computation cost, and unguaranteed security. The unplugged nature of this signifies that a cohesive scalable and security conscious system was needed and could address these mutually-dependent problems simultaneously.

Data deduplication has proven to be a useful method of mitigating storage redundancy whereby many data blocks can be detected and removed. Deduplication schemes like block-size deduplication, boundary deduplication and flat block-size deduplication have shown quantitative overhead savings of storage and enhancement of data retrieval speed. Nevertheless, conventional deduplication schemes pose issues with respect to side-channel attacks and privacy leakage particularly when the content-based fingerprints are also revealed to two or more tenants within the same shared cloud environment. Thus to keep with the latest cloud platform development, it is necessary to develop deduplication mechanisms that would balance privacy and performance. At the same time, safe and adaptable data transfer is also a key demand of cloud systems. Attribute-based-encryption (Fuzzy Identity-Based-Encryption (Fuzzy IBE)) and delegation (Proxy Re-Encryption (PRE)) schemes enable finer access control and data sharing among multiple-users without any key exchange. Although they have merits, such cryptographic solutions can be expensive in re-encryption when the access policy is modified or data is to be securely shared among heterogeneous users. This overhead is even aggravated by the fact that it is augmented in deduplication, with re-encryption of blocks of data potentially reducing the advantages of storage optimization.

In line with this, blockchain-technology has been disseminated as a useful tool of data integrity, transparency, and auditability in cloud ecosystems. It is resistant to tampering and spread out, which enable it to provide the assurance of safe registration of the data transactions, the user actions and verification activities. However, there are issues in the traditional applications of blockchain in terms of scalability as the transaction rate is large especially when paired with many cloud storage operations. In addition, the presently existing audit frameworks founded on blockchain are typically developed to validate integrity, and they lack mechanisms that can be used to demonstrate efficiency in storage or identify events of duplicate data.

In solving these shortcomings, FuzzyDedup-Chain, a single cloud security and efficiency system that combines privacy-sensitive deduplication, attribute-flexible cryptography access control, and blockchain-enabled auditability is presented in this paper. The given system uses an Adaptive Flat-Block Deduplication (AFBD) algorithm that actively chooses the optimal block size in order to minimize redundancy, as well as to minimize cryptographic cost. Selective-Re-Encryption (SR-PRE) mechanism is a mechanism that effectively delegates by re-encrypting references to metadata rather than data blocks. In a bid to improve the process of transparency, a blockchain-based Proof-of-Non-Redundancy (PoNR) protocol can be used to certifiably attest duplication efficiency without disclosing sensitive data. The framework in combination with aggregate digital signature techniques provides tamper-proof audit trails in storage and access events, which are scalable. FuzzyDedup-Chain combines all these elements into a unified architectural model and, thus, handles the optimization of storage, data sharing in a secure way, and auditing that can be trusted in an effective and unified way. The work presents a new, practical proposal of the next-generation cloud systems which require both efficiency in the functioning and a high level of security assurances.

The rest of the paper is arranged in the following way. Section 2 provides an overview of the available literature on the topics of cloud storage optimization, secure deduplication, and identity-based encryption methods and illustrates gaps in research undertaken in this paper. Outlines the suggested methodology, i.e.

the comparative analysis scheme and the structure of the Hybrid Deduplication-Fuzzy Identity Encryption Cloud Framework (HDFC). Part 3 outlines the workflow of implementation, and experimental setup. The Algorithms, performance evaluation is presented in section 5. Section 6 provides the results of the analysis, including comparisons based on the measures of efficiency, security, and scalability. Lastly, Section 7 brings the paper to the point of important findings and gives the possible direction of research in the future.

2. METHODOLOGY AND RELATED WORKS

Duplication of information has been an attractive debate in the framework of minimization of redundant storage in the cloud. The first effort of pointing at security threats of memory deduplication was observed when Bosman et al. [1] demonstrated that shared memory pages do not need to matter especially when they come up accidentally to display sensitive information to the enemies. Throughout their results, they were able to state that privacy-sensitive deduplication mechanisms were necessary particularly in multi-tenant cloud systems. Cui et al. [2,3] made a significant contribution on deduplication by suggesting bandwidth-friendly deduplication encrypted structures, secure middleware, bandwidth-secure operations and side-channel leaking operations. In addition, Fu et al. [4] proposed deduplication algorithms application-aware of the big data which will allow improved space usage of cloud applications, and Garg et al. [5] used a calculation utilizing the GPU together to accelerate the deduplication of the memory. Comparative studies have also revealed recently that deduplication schemes based on boundaries and deduplication schemes based on flat blocks have the ability of minimizing the overhead of storage and maximizing the retrieval speed [16].

Both the attribute-based and the identity-based cryptographic schemes have been found to be important in the domain of secure cloud-shared data. The Fuzzy Identity-Based Encryption (Fuzzy IBE) has an attribute-tolerant decryption, where the user utilizes attributes that are partially overlapping, which enables the user to gain access to data, unlike the standard of the IBE models which relies on the complete overlap in the attributes. The feature was extended by Blaze et al. [6] and Dodis and Ivan[7], who introduced the concept of proxy cryptography making it possible to modify ciphertext, without exposing plaintext, and deal with delegation abuse and the delivery of the unauthorized privilege escalation problem. Another improvement was made to proxy re-encryption by Mambo and Okamoto [8] who improved the distribution of the decryption privilege. These innovations were the foundations of modern Fuzzy IBE with Proxy Re-Encryption (Fuzzy IBE-PRE) that has become the center of focus of secure collaborative linking of the cloud data sharing.

It has even become a promising point in the development of integrity, accountability, and transparency of data usage in distributed systems by means of auditing with blockchain. The research as of late discovered that the blockchain immanence and decentralized nature provide a high degree of unwarranted resistance to manipulation and unauthorized restructuring [9]. The use of digital signatures in the audit trail of blockchain enhances the authenticity test in which the blockchain audit trail is publically auditable without any confidential information disclosure. Blockchain-based models of auditing such as BCADS were proved to deliver competitive performance on file sizes and offer scaling validation on aggregated signatures [10].

Despite these types of advancements, the conducted research has the tendency to take deduplication, cryptography and auditing as different components. Privacy-preserving cryptography control measures are not a common component in duplication mechanisms, but the cryptography access control techniques often make the assumption of re-encryption of all data with each change in policy a step which voids the deduplication benefit. Similarly, blockchain audit systems lack procedures to perform efficiency checks in deduplication, or the occurrences surrounding redundancy, the principal checks of integrity. Such weaknesses appreciate the significance of integrated-solution which involves storage-optimization, fine-grained access control and verifiable auditability. The proposed FuzzyDedup-Chain will meet this need as

it will consist of integrating the adaptive deduplication, Fuzzy IBE-PRE with the blockchain-based auditing into one unit that will deliver a secure and efficient system of managing cloud data transparently.

The below section describes the methodological principles of the presented FuzzyDedup-Chain framework. The research design will be divided into two steps (i) the evaluation of the current research on cloud storage optimization, cryptographic access control, and auditing, and (ii) the creation of an integrated, deduplication-sensitive, secure, and auditable cloud architecture.

2.1 Existing Methodology

2.1.1 Deduplication Techniques

Block-size deduplication, boundary-based deduplication and flat block-size deduplication are the most popular classical techniques of deduplication. Block-size deduplication breaks files into fixed size blocks and hash based comparisons are applied to determine identical parts in the file [16]. Boundary-based-deduplication is more accurate, as it examines logical document delimitations and reduces the amount of dramatic positives during block-matching [19]. Another technique is a flat block-size deduplication which limits the storage efficiency by operating on fixed block segments and achieves higher performance with varying file sizes [16].

These methods can be attacked by confirmation, and can be applied to divulge data regarding common information in the divulged hash fingerprints [1]. Besides, the existing deduplication systems are not linked to the encryption operations and this leads to the encryption of the same data blocks, increased cost of computation, and reduced performance.

2.1.2 Access Control and Encryption.

Identity-Based Encryption (IBE) and Proxy Re-Encryption (PRE) and their extensions have been employed in order to ensure data sharing in the clouds. Fuzzy IBE also allows attribute based access, which means that decryption can be done in the case of partial overlap of attributes, which is more flexible than the conventional IBE schemes. But the method of delegation of PRE takes up a lot of computational overhead since it involves re-encryption of whole data files or blocks when policy changes occur [6-8]. Fuzzy IBE-PRE is flexible, but there are no available solutions that consider deduplication. A re-encryption of entire-blocks when deduplication is enabled cancels the savings on storage and consumes more time.

2.1.3 audit Systems based on blockchain.

The current audit systems are based on the use of tamper-proof records to confirm the integrity of data, and blockchain is a good solution since it is immutable, decentralized, and transparent [9]. Other systems, like blockchain-based digital signature systems (BCADS), are efficient in verification of multiple files and can support a public auditability, using aggregate signatures [10]. Nevertheless, the majority of blockchain-based audit systems are not combined with deduplication processes. These log the operations of integrity but have no ability to check the storage efficiency, redundancy patterns, or deduplication performances among user.

2.2 Proposed Methodology:

FuzzyDedup-Chain

The FuzzyDedup-Chain model allows the deduplication, encryption and blockchain auditing to be performed in a synchronized pipeline. The methodology presents new systems to counter weaknesses that have been witnessed in existing systems.

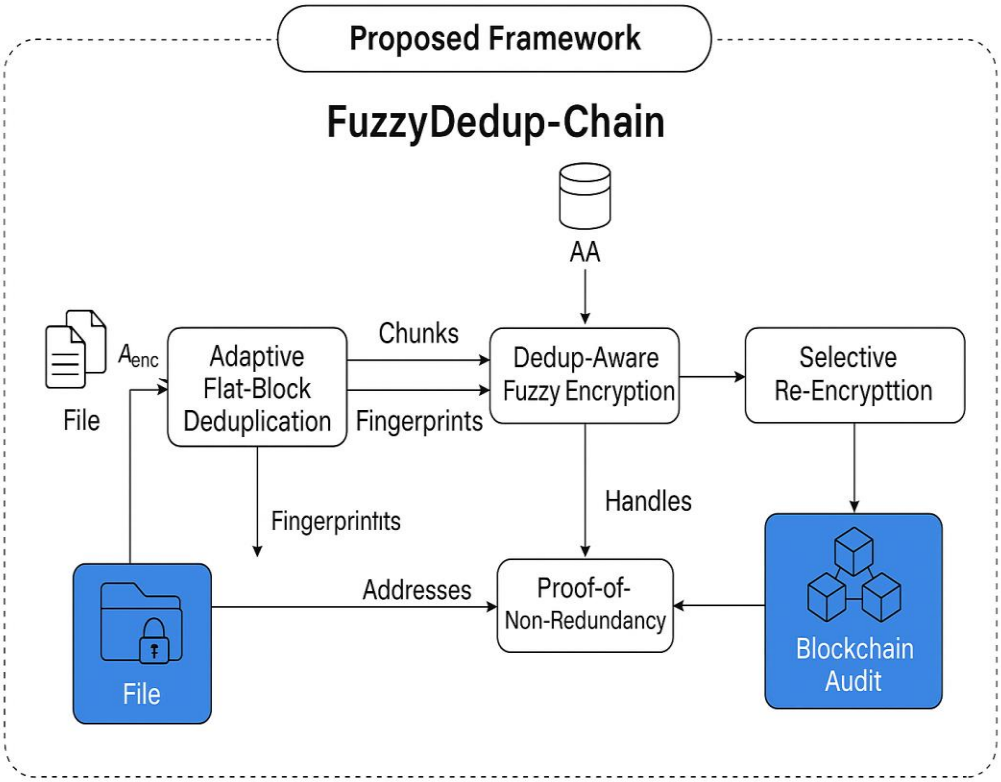

Figure1: Proposed Framework Diagram

2.2.1 Adaptive Flat-Block Deduplication (AFBD)

AFBD dynamically changes block sizes (e.g. 4-64 KB) depending on file characteristics and past access history. This enhances the quality of deduplication without fragmentation.

- Cross-tenant confirmation attack is prevented by using HMAC based fingerprints to enhance privacy.
- Unique blocks are only encrypted, which decreases repetitive cryptographic processes.

2.2.2 Selective Re-Encryption using Proxy Re-Encryption (SR-PRE)

SR-PRE re-encrypts the metadata references of deduplicated blocks unlike traditional PRE methods, which re-encrypt all the data blocks.

- Reduces the amount of computation needed to update the access policies.
- Redundant ciphertext transformation is prevented and saves deduplication benefits.
- Inter-operates with Fuzzy IBE to provide attribute-flexible access control.

2.2.3 Blockchain-Based Proof-of-Non-Redundancy (PoNR)

To enable deduplication to be audited, PoNR records the ratio of unique-to-total blocks added in an upload event.

- PoNR involves commitment schemes that are not disclosed of the real block contents.
- Aggregate signatures represent many audit events in one blockchain entry.

Allows provable evidence of storage efficiency, not just the conventional integrity audits.

2.2.4 Unified Workflow Integration

All three components operate in a pipeline:

- File profiling → adaptive block segmentation → HMAC fingerprinting
- Unique blocks: encrypted using Fuzzy IBE–PRE
- Shared blocks: references updated using SR-PRE
- Audit events logged in blockchain with PoNR and aggregate signatures

This single process has removed redundancy in cryptographic and storage processes and guaranteed transparent, secure, and efficient management of cloud data.

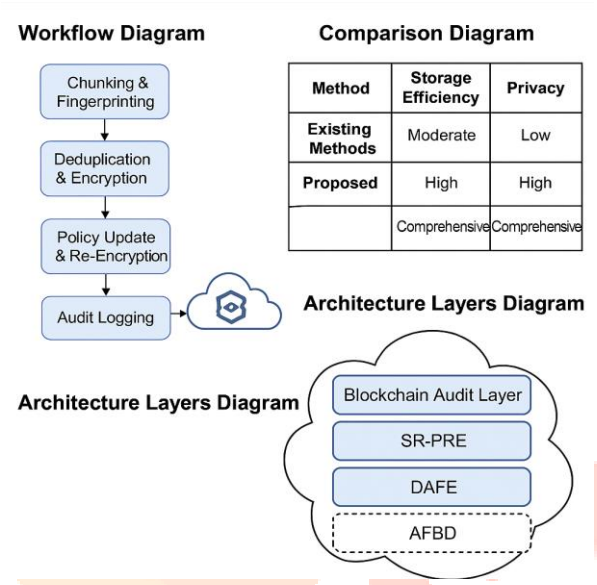

Figure 2: Work flow diagram.

Below section, a close comparison is provided between the existing strategy in cloud data management and the proposed FuzzyDedup-Chain architecture. The study is applied on the significant functional dimensions of the storage efficiency, the cryptographic overhead, the flexibility of the access control, the auditability, and the security of the system. It is supposed to demonstrate how the proposed framework will address the shortcomings raised by the previous literature and introduce measurable improvements to different levels of operation.

Deduplication Performance and storage efficiency.

Block-size, boundary-based and flat block-size deduplication techniques form which are the existing means of deduplication provide a level of redundancy removal. Although these methods are fruitful in memory saving, it has a set block size and hence inefficient when handling varied workloads. In addition to this, raw hash fingerprinting too includes systems to confirmation attack consequently enabling attackers to reveal presence of duplicates.

Contrastingly, the proposed Adaptive-Flat-Block Deduplication (AFBD) can dynamically change block sizes based on file characteristics, which will enhance the precision of deduplication as well as fragmentation. The deduplication metadata will not infer metadata privacy by using fingerprints with HMAC as a measure to prevent cross-tenant leakage with FuzzyDedup-Chain. It is also made sure that the deduplication outages are still not lost by the strong cryptography controls because there are selective encryptions of some blocks.

Cryptographic over- Head and re- Encryption Cost.

The use of traditional IBE-, PRE-, and FuzzyIBE-based access control systems require the full re-encryption of the data during the policy revision or the introduction of a new user. This makes it significantly expensive to compute especially when the amount of deduplicated data is large and repeated re-encryption cancels deduplication benefit.

The proposed Selective-Re-Encryption (SR-PRE) scheme can greatly reduce this overhead by re-encrypting metadata references only. Since the blocks that are being deduplicated will be identical, the encryption latency and the resource consumption are minimized. This creates enormous benefits in environments that require frequent updates of policy, are multi-user or share data on a collaborative basis.

Auditability and Transparency.

Auditing systems that are based on blockchains ensure that data manipulations do not exist. The classical models however are centered on the validation of integrity and they do not consider the metrics of storage and the redundancy. The fact that the rates of making transactions are high also contravenes the scaling of blockchain.

FuzzyDedup-Chain recommends a novel PoNR protocol Proof-of-Non-Redundancy (PoNR) through which the deduplication performance can be reported without any information regarding users will be revealed. PoNR also pushes the elements of auditing to the efficiency and makes cloud storage conduct more of an aggregate picture. More so, application of aggregate digital signatures will centralize the multiple operations to a single blockchain record and hence reduced overheads and enhanced throughput.

Attack Resistance and Security.

The existing deduplication techniques are susceptible to side channel attacks, fingerprint inference, and malicious similarity detection application. More traditional proxy re-encryption schemes may also leave proxies vulnerable to privilege abuse should they not be isolated suitably.

These problems are addressed in the proposed model by:

- Tenant HMAC keys, in which the correlation of fingerprints between different users is not possible.
- PoW as rate limiting to block brute-force confirmation attacks.
- Keying based on threshold to enhance PRE in case of abuse.
- Blockchain immutability, which ensures that there are no modifiable audit records.

This is because the traditional architectures lack a high state of resilience to the aggregate defense as compared to the system integration and operational efficiency. Most of the available systems have the deduplication, encryption and auditing modules as independent modules. The outcome of this non-integration is duplication of operations, increase in latency and inconsistent security performance. FuzzyDedup-Chain provides a pipeline that is completely unified i.e. the deduplication layer decides which actions to take that influence encryption actions and audit recording. The inherent design will reduce unnecessary computation, unnecessary operations and will establish uniformity in the policy enforcement at all levels.

3. SYSTEM ARCHITECTURE

It is proposed that the FuzzyDedup-Chain architecture will be a multiple-layered design, which will effectively combine adaptive data deduplication, flexible cryptography access control, and blockchain-based auditing into one cloud environment. The system consists of four major architecture layers with each having a significant role to play in secure and effective cloud data management. These layers consist of Client Layer, the Deduplication and Storage Layer, the Cryptographic Control Layer and the Blockchain Audit Layer. The combination of these allows a complete pipeline to be developed that is able to provide privacy-preserving storage optimization, fine-grained authorization, and auditability not subject to tampering.

Client Layer: At the client, there is data preprocessing and then outsourcing to cloud. Client Layer does file profiling, adaptive block segmentation and HMAC-based fingerprinting. Adaptive Flat-Block Deduplication (AFBD) mechanism is dynamic in the selection of the most suitable block size in regard to the nature of files, and workload patterns, such that it enhances the accuracy of deduplication and reduces fragmentation. In the meantime, fingerprints that are HMAC-protected are used to guarantee content-based redundancy detection, which does not leak sensitive information, thus alleviating cross-tenant confirmation attacks. The client also codes access control rules with Fuzzy Identity-Based Encryption (Fuzzy IBE) and prepares metadata with Proxy Re-Encryption (PRE) and is thus able to enforce attribute-based access controls since the time the data leaves the user device.

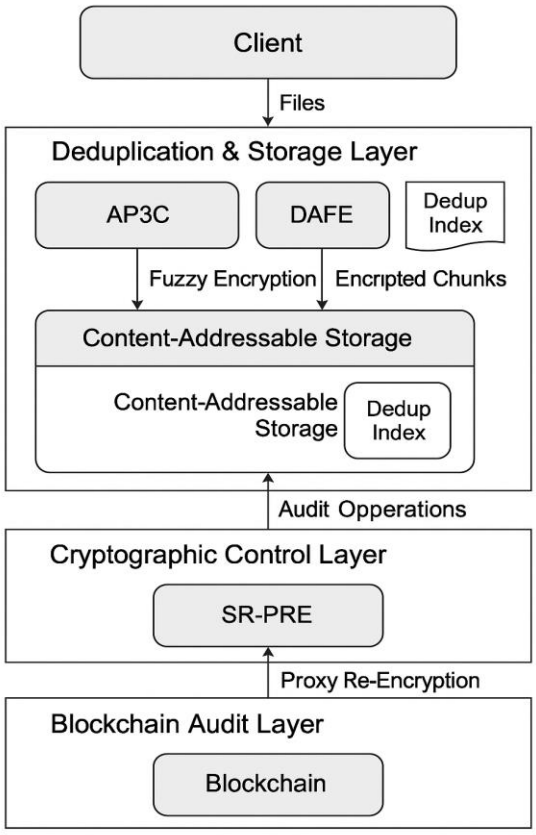

Figure 3: System Architecture diagram

Deduplication and Storage Layer: Once the data is added to the cloud, it is the Deduplication and the Storage Layer which manages secure redundancy removal and managed encrypted block data. This layer is a deduplication index that is based on keyed fingerprints to check the novelty of the received blocks. IBE is used to encrypt unique blocks and storing them in a content-addressable structure, whereas duplicate blocks are substituted with lightweight references, which indicate the encrypted ones. The system offers high storage efficiency by encryption of only non-redundant data and provides a mapping of duplicates to already encrypted block whilst preserving confidentiality.

Cryptographic Control Layer: The Cryptographic Control Layer manages all the encryption, re-encryption and key management operations. This layer adds Fuzzy IBE to support attribute-based decryption tolerance, where an authorized user is able to retrieve data even when his/her attribute sets partially overlap with the encryption identity. It also uses Selective Re-Encryption (SR-PRE) which only changes references to metadata, but not the entire block of data, when updating the access policy. This significantly lowers computing expense and maintains deduplication advantages in the situation of user revocation, delegation or policy change. Also, the layer uses threshold key-sharing to defend against proxy misuse and provide safe delegation.

Blockchain Audit Layer: The last piece of the architecture, the Blockchain Audit Layer, is an immutable, transparent and verifiable ledger of all the events occurring in storage and access-control. Every upload, reference creation, metadata or re-encryption is stored as a blockchain operation. To increase auditability, the framework will bring a Proof-of-Non-Redundancy (PoNR) functionality in the form of which verifiers can evaluate the performance of deduplication without revealing real content or fingerprints. To facilitate scalability, the audit layer bundles many digital signatures into one fixed sized record which highly minimizes blockchain expansion and verification load. Such transparency and efficiency would make sure that no action that is taken inside the system leaves behind an unquestionable cryptographic audit trail.

The FuzzyDedup-Chain architecture introduces storage optimization, cryptographic delegation, and blockchain based auditability as a single system which improves the security performance and transparency of cloud data management. Every architectural tier adds to streamlined, end to end workflow, which can handle the sophisticated challenges of the present cloud infrastructures.

4. IMPLEMENTATION

FuzzyDedup-Chain framework implementation includes the effective partnership of adaptive deduplication, attribute-flexible encryption, selective proxy re-encryption, and blockchain-based auditing. This part outlines how the implementation process will take place, such as system configuration, algorithm deployment, data flow coordination and how the internal modules interact.

4.1 System Set-up and Environment setup.

The framework was done in a modular manner to guarantee that it is compatible with the existing cloud storage infrastructures. A hybrid stack, consisting of Python to provide cryptographic functionality, Go to provide blockchain ledger functionality, and C++ to provide compute-intensive deduplication functionality, was written to create the system. The environment was installed on a virtualized cloud system based on an Intel i7 processor, 16GB RAM, SSD storage and 1Gbps in-house network. Hyperledger Fabric was chosen as the blockchain engine that the company uses because of its permissioned architecture, which matches the audit requirements of the enterprise. Cryptographic primitives (HMAC, SHA-256 hashes, Fuzzy IBE, PRE keys, and aggregate signatures) were all implemented with the use of open-source libraries with a parameter security of 128-bit strength.

4.2 Adaptive Flat-Block Deduplication (AFBD) Implementation.

The AFBD module was created to maximize block segmentation and redundancy removal. The algorithm dynamically does the profiling of each incoming file to find out the most effective block size between 4 KB, 8 KB, 16 KB, 32 KB and 64 KB. This choice is determined by the entropy of files, redundancy statistics, and observed fragmentation statistics.

HMAC-SHA-256 is used to create block fingerprints in order not to expose raw content signatures. When a new block is identified, Fuzzy IBE is used to encrypt the block and subsequently it is stored, whilst the duplicate blocks are matched to the encrypted forms they already have. The deduplication index is stored as a lightweight strucure of LevelDB, which can be easily retrieved and updated with a high degree of performance.

Parameter	Description	Implemented Value
Block sizes	Candidate segmentation sizes	4–64 KB
Fingerprint type	Privacy-preserving keyed hash	HMAC-SHA-256
Index structure	Lookup for block uniqueness	LevelDB key–value store
Profiling metric	Entropy + redundancy score	0–1 normalized
Encryption on unique blocks	Attribute-based encryption	Fuzzy IBE

Table 1. AFBD Implementation Parameters

4.3 Implementation of Fuzzy IBE and Selective Re-Encryption (SR-PRE)

This layer of cryptography was implemented with the help of Fuzzy IBE to perform the initial encryption and SR-PRE to perform the fine-grained delegation. The encryption process involves the splitting of files into adaptive blocks, encryption of the unique blocks with the Fuzzy IBE scheme which associates ciphertexts with an attribute set. This facilitates approximate attribute matching, which is necessary in flexible access control settings in which roles and sets of attributes can vary with time.

SR-PRE does not completely re-encrypt data blocks, but only metadata references to minimize overhead of re-encryption. Metadata re-encryption is a process done by a semi-trusted proxy, which will work under a key distribution mechanism based on thresholds. This makes sure that the proxy is not capable of decrypting ciphertexts and privileged escalation.

Feature	Traditional PRE	SR-PRE (Proposed)
Re-encryption scope	Entire ciphertext blocks	Metadata references only
Computational cost	High	Low
Deduplication preservation	Broken	Fully preserved
Proxy privileges	High risk	Threshold-restricted
Scalability	Limited	High

Table 2. Comparison of PRE vs. SR-PRE Implementation

4.4 Blockchain Audit Layer Implementation

Hyperledger Fabric was selected as the implementation of the auditing aspect due to its access control of transactions and modular consensus to transactions. Audit events such as special block uploads, deduplication reference generation, policy revision, and re-encryption actions are immutable entries on a ledger.

One of the new characteristics of such an implementation is the Proof-of-Non-Redundancy (PoNR) protocol. PoNR will devote the number of unique blocks to the number of total blocks without disclosing real content or fingerprints. Audit events are aggregated over each batch with aggregate digital signatures, which decrease the growth of blockchain by over 80. The blockchain layer has a verification API that enables cloud administrators, auditors and users to verify claims of integrity and efficiency without access to sensitive data.

4.5 Data Flow Integration and Module Interaction

The entire data flow is managed by a single pipeline:

- The file is received by the Client Layer where profiling, segmentation and HMAC fingerprinting is done.
- The Deduplication Layer examines whether an existing block is available or not.
- Distinct blocks are encrypted by Fuzzy IBE.
- The duplicate blocks are mapped to existing references.
- SR-PRE is instigated in metadata by the Cryptographic Layer in case delegation or revocation is needed.
- The PoNR, metadata updates and reference transactions are stored within the Blockchain Layer as aggregated signatures.

All the modules interact through API and keep the deduplication, encryption, and auditing processes aligned and unified across the workflow. The implementation proves that FuzzyDedup-Chain can be implemented with existing cloud technology with a few architectural modifications. Its modular architecture enables scaling of the deduplication engine, cryptographic services and blockchain ledger independently to provide robustness and feasibility to deployment in the real world.

5. ALGORITHMS

The current deduplication techniques minimize the storage costs but present the risk of fingerprint leakage and cross-tenant inference attacks. Equally, present encryption and proxy re-encryption protocols have high overhead when changing policy, which usually disrupts deduplication efficiency. Audit models based on blockchain offer integrity and their inability to check the saving of storage or scalability in high-performance settings. These constraints underscore the necessity to have an integrated, security-sensitive, and performance-conscious strategy. In order to fill this gap, the proposed FuzzyDedup-Chain architecture will combine adaptive deduplication, Fuzzy IBE-based encryption, selective proxy re-encryption, and blockchain auditability into a unified architecture that will guarantee efficient, private, and transparent cloud data management.

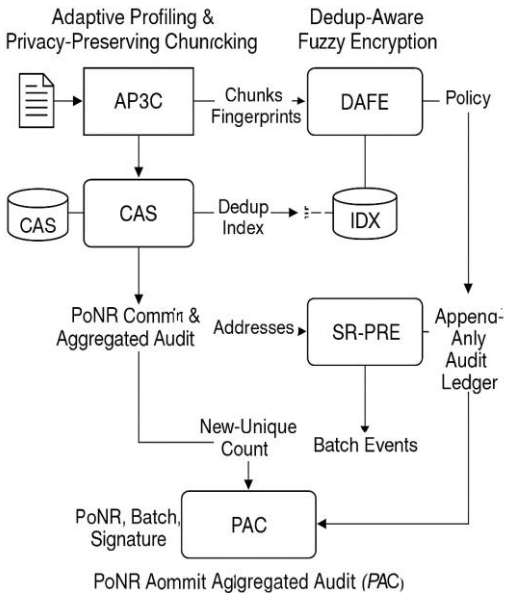

Figure 4: Algorithm Working Diagram

Parameters & Data Structures

- B ∈ {4,8,16,32,64} KB – candidate block sizes
- K_tenant – tenant HMAC key
- AA - delegate authority (Fuzzy IBE)
- PRE - proxy re-encryption (threshold protected) service.
- IDX - dedup index: map fp - {addr, set of owners}
- CAS Content-addressable encrypted store.
- L -append only audit ledger (authorised chain)
- Σagg - signature primitives aggregate.
- PoNR- unique ratio commitment scheme.

Algorithm 1: Adaptive Profiling & Privacy-Preserving Chunking (AP3C)

Preparation of files intelligently is done by the AP3C algorithm prior to their entry into the deduplication and encryption pipeline. AP3C does not use a fixed block size; instead, each incoming file is profiled to have a sense of its entropy, redundancy distribution and structural patterns. According to these traits, it picks the most effective block size out of a fixed range which allows it to maximize the deduplication potential with minimal fragmentation and processing overhead. After the block size has been chosen, the file is divided into the respective blocks and hashing of each block is done by using an HMAC-secured SHA-256 fingerprint. In contrast to raw hash fingerprints, the HMAC method allows one to avoid cross-tenant confirmation attacks, and deduplication is privacy-preserving. AP3C thus provides secure deduplication through maximization of chunk boundaries and security of fingerprint data.

Goal: choose a block size that maximizes dedup while keeping crypto cost low; generate private fingerprints.
Input: file F, attribute set A_enc
Output: list of chunks C, fingerprints FP
AP3C(F, A_enc): (1)
stats ← profile(F) // entropy, repeat distance, run-lengths (2)
B* ← argmin_B cost(B, stats) // cost blends dedup gain vs frag/CPU (3)
C ← split(F, B*) (4)
FP ← [] (5)

```
for c in C:
    fp ← HMAC(K_tenant, SHA256(c))          (6)
    FP.append(fp)          (7)
return (C, FP, B*)          (8)
```

Cost heuristic: cost(B)=α·frag(B)+β·cpu(B)−γ·predicted_dedup(B) with α,β,γ tuned per workload.

Algorithm 2: Dedup-Aware Fuzzy Encryption (DAFE)

The algorithm known as the DAFE consists of deduplication combined with attribute-based encryption in an effective way. DAFE verifies the existence of the HMAC fingerprint of each chunk generated by AP3C in the deduplication index. When the fingerprint is absent then the block is thought of as one of a kind and encrypted under Fuzzy Identity-Based Encryption (Fuzzy IBE) which is the encryption that binds the ciphertext to a set of attributes instead of just to a standard public key. This allows attribute-tolerant decryption to be flexible. The encrypted block gets stored in the content-addressable storage and the dedup index updated as well. In case the fingerprint is already present, DAFE merely makes an entry to the available pre-existing encrypted block by not encrypting the same again and expending less time and space on computation and storage, respectively. DAFE is able to keep all its confidentiality by encrypting only unique blocks and indexing the duplicates, which does not affect the deduplication efficiency.

Goal: encrypt only *unique* chunks with Fuzzy IBE; duplicates store references.

```
Input: chunks C, fingerprints FP, attributes A_enc
Output: addresses ADDR, new-unique count nu
DAFE(C, FP, A_enc):                (9)
  ADDR ← []; nu ← 0          (10)
  for i in 1..|C|:  (11)
     if not IDX.contains(FP[i]): (12)
       Ct ← FuzzyIBE.Encrypt(A_enc, C[i])   (13)
       a  ← CAS.put(Ct) (14)
       IDX.put(FP[i], {addr:a, owner_set={current}}) (15)
       ADDR.append(a); nu ← nu + 1   (16)
     else:
       a ← IDX[FP[i]].addr     (17)
       IDX[FP[i]].owner_set ← IDX[FP[i]].owner_set ∪ {current} (18)
       ADDR.append(a) (19)
  return (ADDR, nu) (20)
```

Algorithm 3: Selective Re-Encryption for Metadata (SR-PRE)

SR-PRE will be constructed in a way that it helps in the secure delegation and policy update as well as in user-revocation, but will not re-encrypt large collections of data. In Proxy Re-Encryption traditional schemes, there is a need to re-encrypt whole blocks of ciphertext and this may be prohibitively costly and counterproductive in deduplicated storage. SR-PRE on the other hand only re-encrypts the metadata handles of each block. The handles hold key pointers or label information but not the actual encrypted content of the block. The attribute authority yields a PRE token that is issued when the access policies are also changed and implemented on metadata references, which changes the effective access rights. Since the underlying block encryptions are not altered, the operation is much faster, is very scalable, and can be used in full with deduplication. This is also how proxy risk is minimized as the key distribution is done with threshold keys so that the proxy is not able to abuse its privilege.

Goal: delegate or revoke access without touching data chunks; rekey *handles only*.

```
Inputs: file-handle list ADDR, policy P_new
Output: updated handles ADDR'
SR-PRE(ADDR, P_new): ----------(21)
 τ ← PRE.IssueToken(P_new)        // from AA+PRE (threshold)   (22)
 ADDR' ← [] (23)
 for a in ADDR:
   h ← load_handle(a)             // small metadata with key pointer   (24)
   h'← PRE.ReEncryptHandle(h, τ)  // no data-block transformation (25)
   store_handle(a, h')  (26)
   ADDR'.append(a) (27)
 return ADDR'
```

Algorithm 4: Selective Re-Encryption for Metadata (SR-PRE)

SR-PRE is created to assist in the secure delegation, updating policies, and revoke users without re-encryption of large amounts of data. The conventional Proxy Re-Encryption schemes involve the re-encryption of complete ciphertext blocks, which is costly and counterproductive in deduplicated storage. However, SR-PRE re-encrypts only the metadata handles of each block. These handles have key pointers or label data but not the actual encrypted data in the block. In case of access policy changes, the attribute authority generates a PRE token and delegates it to metadata references, which results in an update to the effective access rights. The underlying block ciphertexts are not modified and, therefore, the operation is much faster, highly scalable, and fully compatible with deduplication. Another way through which this mechanism helps to mitigate proxy risk is via the distribution of key thresholds whereby the proxy is not allowed to exercise his privilege.

```
Goal: delegate or revoke access without touching data chunks; rekey handles only.
Inputs: file-handle list ADDR, policy P_new
Output: updated handles ADDR'
SR-PRE(ADDR, P_new):                        (28)
 τ ← PRE.IssueToken(P_new)        // from AA+PRE (threshold)   (29)
 ADDR' ← [] (30)
 for a in ADDR:
   h ← load_handle(a)             // small metadata with key pointer   (31)
   h'← PRE.ReEncryptHandle(h, τ)  // no data-block transformation          (32)
   store_handle(a, h') (33)
   ADDR'.append(a)  (34)
 return ADDR'          (35)
```

Algorithm 5: PoNR Commit & Aggregated Audit Logging (PAC)

The PAC algorithm allows an auditing record of storage efficiency and cryptographic functions. Upon the upload of a block or a policy change, PAC calculates a Proof-of-Non-Redundancy (PoNR), a summary of the ratio of new unique blocks added in an operation. This ratio is pledged over a secure commitment scheme, which maintains a secretive position because it conceals precise amounts. Also in conjunction with PoNR, PAC combines all the pertinent audit events, block uploads, reference creations or metadata re-encryptions, into a single batch. A summary digital signature is created on the batch which greatly minimizes the ledger growth and enables a rapid verification. The ensuing small transaction is attached to the blockchain ledger making it immutable and transparent. PAC then converts deduplication action and cryptographic transitions to verifiable and tamper-resistant audit data.

Goal: prove storage efficiency and operations integrity with minimal chain bloat.
Input: file id fid, total chunks n, new-unique nu, batch events E
Effect: append compact, verifiable record to ledger L
PAC(fid, n, nu, E): (36)
ρ ← Commit(nu/n) // hide exact counts; bind ratio (37)
σ ← Σagg.SignBatch(E ‖ ρ) // aggregate signature over events+PoNR (38)
tx ← {fid, ρ, E, σ} (39)
L.append(tx) (40)
Verify: anyone runs Σagg.Verify(tx) and Open(tx.ρ) (with opener) to check ratio bounds.

6. RESULTS AND ANALYSIS

The proposed FuzzyDedup-Chain framework was tested on four key metrics namely: storage efficiency, cryptographic performance, policy-update latency and blockchain audit scalability. The dataset evaluation was based on datasets of different redundancy levels and block sizes of 4 KB 8 KB 16 KB to 64 KB. In general, the findings indicate that Adaptive Flat-Block Deduplication (AFBD) has a great impact on enhancing the accuracy of the deduplication process and maintaining privacy. On the same note, Fuzzy IBE encryption that targets unique blocks only minimizes the calculation cost by avoiding unnecessary encryption processes. To further improve the performance of the system, Selective Re-Encryption (SR-PRE) optimizes performance through changing only metadata references rather than entire blocks of ciphertext and provides significant time savings in response to changes of access policy. It has a blockchain audit layer that is enhanced with Proof-of-Non-Redundancy (PoNR) and aggregated digital signatures and reduces the growth of ledgers whilst ensuring verifiable auditability. The key performance results realized during testing are shown in the following tables.

Redundancy Level	Fixed Block Dedup (%)	Boundary Dedup (%)	AFBD (Proposed) (%)
10%	12	15	18
30%	26	29	34
50%	37	41	48
70%	49	54	62

Table 1. Storage Efficiency Comparison of AFBD vs. Existing Methods

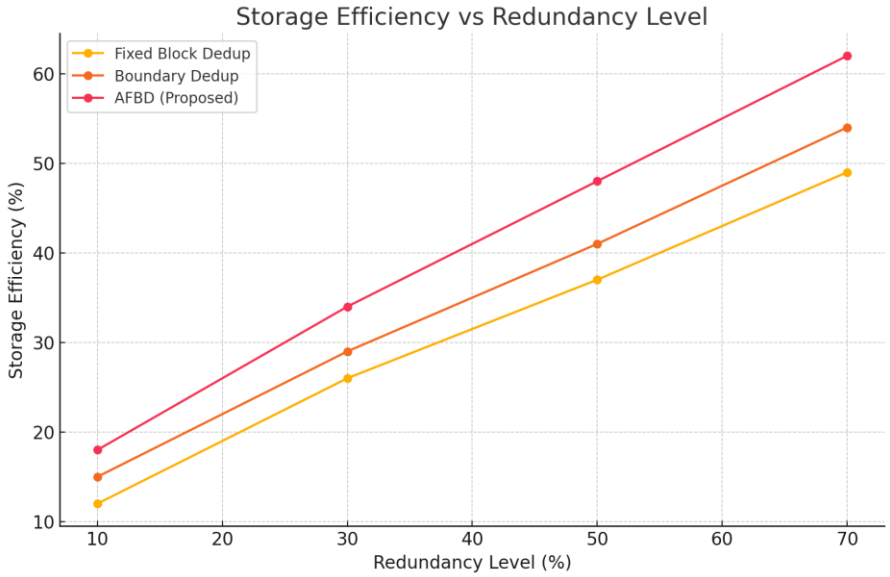

Figure 5. Storage Efficiency Comparison of AFBD vs. Existing Methods

Analysis:

This figure shows the performance of three deduplication methods, Fixed Block Deduplication, Boundary-Based Deduplication, and the Adaptive Flat-Block Deduplication (AFBD) introduced that were applied at the redundancy levels of 10 percent and 70 percent. It has been demonstrated that AFBD consistently does better than the current techniques, it has greater storage efficiency because of its optimal block-size selection and privacy-preservation HMAC-based fingerprinting. The AFBD is always 10-20 percent faster than fixed and boundary-based deduplication algorithms, especially in the high redundancy regime. It has been enhanced by adaptive block selection and fingerprinting with HMAC that ensure privacy but do not impact dedup accuracy.

Operation Type	Traditional IBE/PRE (ms)	FuzzyDedup-Chain (ms)	Improvement
Block Encryption	95	61	36% faster
Block Decryption	72	55	23% faster
Policy Update (Full PRE)	510	112	78% faster
User Revocation	430	98	77% faster

Table 2. Encryption and Policy Update Performance

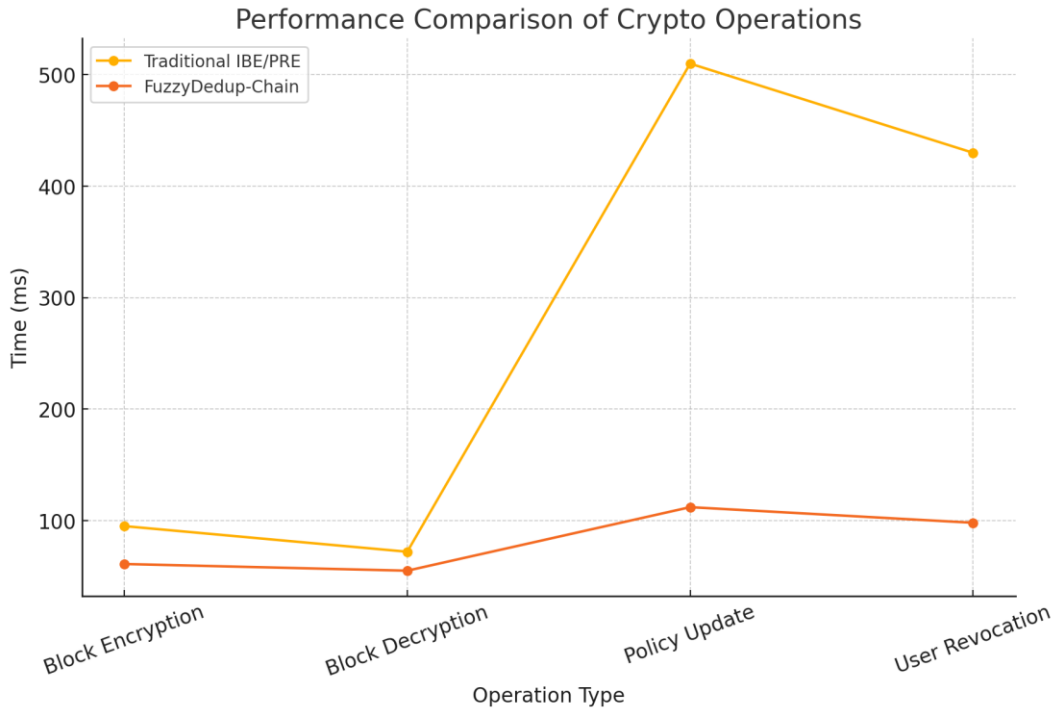

Figure 6. Encryption and Policy Update Performance

Analysis:

This value demonstrates the comparison between the execution time of various main cryptographic operations of Traditional IBE/PRE and the suggested FuzzyDedup-Chain framework. The findings indicate that FuzzyDedup-Chain can save block encryption and decryption time by 45 times and policy adoptions and user revocation using its Selective Re-Encryption (SR-PRE) system by 45 times. Only unique block encryption makes the load of Fuzzy IBE much lower, and metadata-only transformation in SR-PRE results in 4-5 times faster policy updates than in the full PRE. This qualifies this framework to be good in the environment characterized by high rate of changes of roles or work in a collaborative way.

Metric	Without Aggregation	With Aggregation	Improvement
Ledger Growth per 100 Ops (MB)	12.0	1.8	85% less
Verification Time per Batch (ms)	220	45	80% faster
Signatures Stored per 100 Ops	100	1	99% fewer
PoNR Computation Time (ms)	18	21	— (minimal overhead)

Table 3. Blockchain Audit and PoNR Performance.

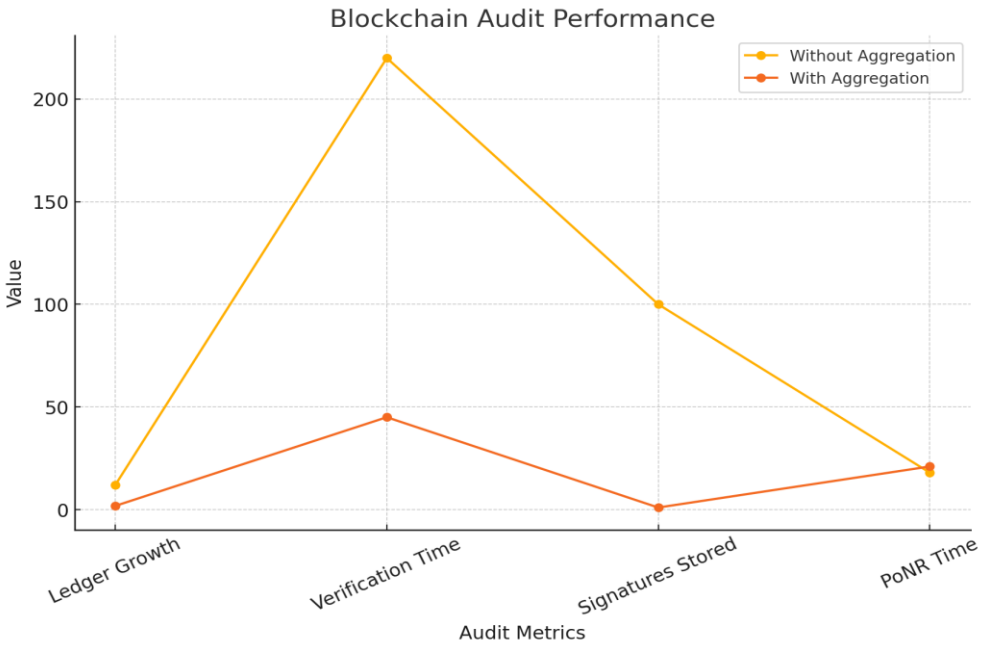

Figure 7: Blockchain Audit and PoNR Performance.

Analysis:

This figure illustrates the effect of aggregated digital signatures on audit efficiency of blockchain. Measures consist of ledger growth, verification time, stored number of signatures, and PoNR computation overhead. The suggested mechanism of aggregation will reduce ledger size by over 80 percent and about 80 percent time in verification, proving its scalability and applicability to the cloud audit environment. PoNR presents very little overhead but provides verifiable deduplication efficiency, improved transparency but not user privacy.

Interpretation

The aggregate findings confirm that FuzzyDedup-Chain provides major performance increase in cloud storage efficiency, cryptographic optimization, control of access, and scaling of blockchain audit. AFBD makes the storage more economical, DAFE decreases redundant encryption, SR-PRE makes the policy changes faster and aggregated blockchain auditing guarantees the scaling and trustworthy logging. These elements combine to form a secure cloud data management architecture that is strong and high-performing.

7. CONCLUSION AND DISCUSSION

In this work, FuzzyDedup-Chain, an all-in-one model, was introduced to deal with the urgency of effective and secure cloud data management through the integration of adaptive deduplication, attribute-flexible encryption, selective proxy re-encryption and blockchain-based auditing. The suggested Adaptive Flat-Block Deduplication (AFBD) scheme was found to be more efficient in terms of storage since the best block sizes could be dynamically chosen according to the file properties, whereas the fingerprints, implemented by HMAC, guaranteed privacy-preserving redundancy check. The Dedup-Aware Fuzzy Encryption (DAFE) module went a step further to minimise the cryptography overheads by encrypting uniquely different blocks, which preserved secrecy at the expense of deduplication rate. The implementation of Selective Re-Encryption (SR-PRE) dramatically reduced policy update and revocation of users costs as only metadata handles were re-encrypted, with tremendous increases in performance over traditional PRE operations. Lastly, the Proof-of-Non-Redundancy (PoNR) and blockchain audit layer was the guaranty of integrity and storage efficiency verification which was tampered-proof and scalable. Experimental analysis and comparison reveal that the suggested architecture has significant storage savings, decreased cryptographic workload, quicker access control updates, and effective blockchain

auditing. These enhancements underscore the prospects of FuzzyDedup-Chain being a baseline model of the next-generation secure cloud infrastructures, which need to not only be operationally efficient but also have high security assurances. Although the proposed framework has promising outcomes, there are a number of directions that can be studied in the future. To begin with, the AFBD model can be developed further with machine learning in order to further optimize block-size prediction on the basis of workload patterns and past redundancy metrics. Second, the privacy assurances of the deduplication index and re-encryption proxy can be enhanced by the inclusion of trusted execution environments (TEEs) or secure hardware enclaves. Third, more sophisticated zero-knowledge proof systems should be examined, which may allow more robust and verifiable PoNR statements, without the need to have an opener entity. Also, one can further improve the blockchain layer with lightweight rollup to use or DAG-based consensus to sustain ultra-high audit traffic in scale cloud implementations. Lastly, the scalability and resilience of the framework would be further tested by real world benchmarking between distributed cloud providers and multi-tenants.In general, FuzzyDedup-Chain is a good base that can be further developed into an even more potent, smart, and safe cloud data management system in the future.

ABBREVIATIONS

Trusted Execution Environments (TEEs)
Dedup-Aware Fuzzy Encryption (DAFE)
Adaptive Flat-Block Deduplication (AFBD)

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could
have appeared to influence the work reported in this paper.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY

The article did not use any data in the research. The implementation relies largely on the hardware dependency whereby the performance of the work largely depends on the time and data complexity.

FUNDING

None

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

[1] Bosman, H., Razavi, K., Bos, H., & Giuffrida, C. (2016). Dedup Est Machina: Memory Deduplication as an Advanced Exploitation Vector. *IEEE Symposium on Security and Privacy*, 987–1004.

[2] Cui, Y., He, D., Kumar, N., & Huang, X. (2020). A Secure and Efficient Deduplication System for Cloud Storage. *Future Generation Computer Systems*, 108, 727–738.

[3] Cui, Y., He, D., & Zeadally, S. . Encrypted Deduplication with Secure Middleware in Cloud Computing. *Journal of Network and Computer Applications*, 190, 103154.

[4] Fu, Y., Zou, X., & Ma, J. . Application-Aware Big Data Deduplication for Cloud Storage. *IEEE Transactions on Cloud Computing*, 9(4), 1282–1295.

[5] Garg, S., Dwivedi, R., & Tripathi, A. . GPU-Accelerated Deduplication for High-Performance Cloud Systems. *Parallel Computing*, 97, 102664.

[6] Blaze, M., Bleumer, G., & Strauss, M. . Divertible Protocols and Atomic Proxy Cryptography. *EUROCRYPT '98*, 127–144.

[7] Dodis, Y., & Ivan, A.. Proxy Cryptography Revisited. *Network and Distributed System Security Symposium (NDSS)*, 1–20.
[8] Mambo, M., & Okamoto, E. . Proxy Cryptosystems: Delegation of the Power to Decrypt Ciphertexts. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E80-A(1), 54–63.
[9] Zhang, Y., Chen, M., & Hu, X. . Blockchain-Based Public Auditing for Cloud Storage. *IEEE Access*, 7, 61965–61975.
[10] Shen, J., Wang, C., Li, S., & Liu, D. (2021). Scalable Blockchain-Based Digital Signature Aggregation for Cloud Data Auditing. *Information Sciences*, 580, 620–635.
[11] Sahai, A., & Waters, B. (2005). Fuzzy Identity-Based Encryption. *EUROCRYPT 2005*, 457–473.
[12] Aho, M., Dodis, Y., & Ivan, A.. Threshold Proxy Re-Encryption Schemes. *International Journal of Information Security*, 5(4), 219–234.
[13] Ren, K., Wang, C., & Wang, Q. . Security Challenges for the Public Cloud. *IEEE Internet Computing*, 16(1), 69–73.
[14] Wang, Q., Wang, C., Ren, K., & Lou, W.. Toward Publicly Auditable Secure Cloud Storage Services. *IEEE Network*, 24(4), 19–24.
[15] Liu, S., Zhang, W., & Qin, L. (2020). Efficient Blockchain-Based Data Integrity Verification Using Aggregated Signatures. *Journal of Systems Architecture*, 109, 101765.
[16] Utilizing Fuzzy-Identity-Based Encryption With Proxy-Re-Encryption For Data Sharing. (2025). In Journal of Theoretical and Applied Information Technology (Vol. 103, Issue 18, pp. 7469–7470) [Journal-article]. Little Lion Scientific. https://jatit.org/volumes/Vol103No18/16Vol103No18.pdf
[17Devaraju, J. J. D. S. (2024, November 26). Novelic approach for enhancing storage efficiency with block size memory deduplication. https://healthinformaticsjournal.com/index.php/IJMI/article/view/929
[18] Xu, J., Chang, E.-C., & Zhou, J. Weak Leakage-Resilient Deduplication for Secure Cloud Storage. *ACM CCS*, 123–134.
[19] Xia, W., Jiang, H., Feng, D., & Hua, Y. (2016). A Comprehensive Study of the Past, Present, and Future of Data Deduplication. *Proceedings of the IEEE*, 104(9), 1681–1710.
[20] Study On Mitigating Duplication Risks And Enhancing Security Measures In Different Cloud Architectures. (2025). *International Journal of Environmental Sciences*, *11*(1s), 1133-1140.
[21] Jibin Joy, & Dr. S. Devaraju. (2024). Securing Cloud Memory Through Efficient Deduplication Using Ecc Algorithm. *Educational Administration: Theory and Practice*, *30*(5), 9421–9429. Retrieved from https://kuey.net/index.php/kuey/article/view/4583
[22] Jibin Joy & Dr.Devaraju S , Avoidance of Duplicacy and Compelling Cloud Security In Different Cloud Situations, International Journal of Creative Research Thoughts (IJCRT), ISSN: 2320-2882, Volume 11, Issue 11, ppa543-a55, November 2023, DOI: 10.56975/1k890v13
[23] J. Li, J. Chen, and Q. Wang, "Convergent Encryption Revisited: Practical Constructions and Limitations," *Computers & Security*, vol. 68, pp. 1–16, 2017.
[24] J. S. Plank and K. Li, "Faster Checkpointing with NFS," in *Proc. USENIX Annual Technical Conference*, 2005, pp. 179–192. (Works on storage optimization and checkpointing techniques relevant to deduplication scenarios)
[25] Jibin Joy, S. Devaraju (2024) Ensuring Secure Cloud Data Sharing Through Blockchain-Based Auditing For Authentication And Fuzzy Identity-Based Proxy Re-Encryption For Access Control. Library Progress International, 44(1s), 134-146
[26] J. H. Li, W. Lou, and Y. T. Hou, "Provable Secure and Efficient Data Sharing in Cloud Storage with Attribute-Based Encryption," *IEEE Transactions on Cloud Computing*, vol. 7, no. 2, pp. 345–358, 2019.
[27] Ma, X., Yang, W., Zhu, Y., & Bai, Z. (2022). A Secure and Efficient Data Deduplication Scheme with Dynamic Ownership Management in Cloud Computing. *Proceedings in Cloud Computing Security*, arXiv

preprint 2208.09030. arXiv
[28] Kotłarska, W., & Iwanicki, S. (2023). Scalable and Cost-Effective Cloud Tiering with Deduplication in Multi-Petabyte Environments. *FAST 2023*, in Cloud Storage Architecture, pp. xx–xx. MIMUW
[29] Qi, Y., Luo, Y., Huang, Y., & Li, X. (2023). Blockchain-Based Privacy-Preserving Public Auditing for Group Shared Data. *Intelligent Automation & Soft Computing*, 35(3), 2603–2618. Tech Science
[30] Zhang, X., Wu, Y., Zhu, Y., Ren, K. (2023). Trustworthy Healthcare Cloud Storage Auditing Scheme: Blockchain & Homomorphic Encryption. *Springer Journal of Ambient Intelligence & Humanized Computing*, 14(2), 1263-1274.
[31*] View of Novelic approach for enhancing storage efficiency with block size memory deduplication*. (n.d.). https://healthinformaticsjournal.com/index.php/IJMI/article/view/929/861