Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2025.060209
ISSN 2790-0932 Vol. 6, Issue 2: 84-93

Identity Authentication and Data Protection System and Deployment Optimization Framework Driven by Layered Security Architecture in Cloud Service System

Linghong Cheng*

Security Org, Microsoft, Redmond, 98052, WA, US

*Corresponding author

Abstract: The popularity of cloud computing has made cloud storage mainstream, but its centralized nature has led to security issues such as loss of data control, integrity breaches, and privacy breaches. Traditional full verification is not suitable for resource constrained users due to its high cost. Although third-party auditing (TPA) alleviates trust issues, it still has high costs, privacy breaches, and collusion risks. The lack of data damage location and recovery mechanisms, insufficient support for dynamic updates, high cost of dynamic revocation for enterprise users, and poor real-time performance in multi cloud migration scenarios further limit the reliability of cloud storage.In response to challenges, this article proposes a hierarchical security architecture: firstly, a dynamic verification framework based on hierarchical authentication is designed, combined with error correction code technology to achieve damaged data localization and recovery, and reduce redundant authentication overhead through hierarchical verification; Secondly, a real-time revocation scheme based on proxy re signature and Trusted Execution Environment (TEE) is proposed, which supports dynamic updates of administrator attributes and remote signature replacement to avoid privacy leakage.Experiments have shown that hierarchical authentication significantly reduces authentication time overhead and optimizes data migration performance; Proxy re signing combined with TEE enables real-time revocation of user permissions, with controllable computational and communication costs. This article constructs a full chain protection system covering "verification positioning recovery revocation", systematically solving the contradiction between efficiency, functionality, and security of cloud storage data integrity verification. Future research will focus on discontinuous storage optimization, lightweight re signature mechanism and multi-dimensional security fusion (such as zero trust architecture and AI anomaly detection) to promote the transformation of the scheme to edge computing and enterprise level complex scenarios.

Keywords: Cloud storage security, layered security architecture, data integrity verification, user dynamic revocation, proxy re signature technology

1 Introduction

This study innovatively explores two core scenarios: cloud storage security [1] and intelligent code search [2]. In the field of cloud storage, with the popularization of cloud computing, although cloud storage alleviates the scalability pressure of local storage, it faces challenges such as low efficiency of data integrity verification, third-party auditing (TPA)[4]privacy leakage risks, and lack of damage recovery mechanisms in multi cloud migration. This article proposes a layered data authentication mode that combines error correction code technology[5] to achieve fast damage localization and efficient recovery in multi cloud migration scenarios; At the same time, a real-time revocable verification scheme based on proxy re signature technology[6] and Trusted Execution Environment (TEE)[8]was designed, which supports dynamic updates of administrator attributes[9]to ensure information security during the user revocation process. On the other hand, Chen Anyi's research [11]focus is on code intelligent search technology, aimed at meeting the needs of software scale expansion and code reuse. Through deep learning, a code semantic representation system was constructed, combining Transformer architecture, Graph Neural Network (GNN)[11], and multimodal fusion technology to break through the semantic gap and achieve efficient cross modal retrieval. The proposed dual tower architecture and industrial deployment optimization strategy provide a technical paradigm for intelligent code service platforms, promoting the intelligent development of software engineering. This study validated the efficiency and security of the proposed solution through security analysis and performance evaluation, providing theoretical support for improving the reliability of cloud storage systems and enterprise level applications.

2. Correlation theory

2.1. Research on Optimization Strategies for Cloud Storage Technology

Cloud storage, as a new type of service in cloud computing, integrates storage resources from different devices in the network to provide users with comprehensive services for storing, managing, and computing data[10]. It has the characteristics of high scalability, data security and storage stability, portability, etc. Its system architecture consists of four layers: the storage layer is the foundation, which realizes logical virtualization management by interconnecting various storage devices and physical resources; The basic management layer is at the core, integrating storage devices to work together, and ensuring data security and stability through measures such as data encryption, redundant backup, and disaster recovery; The application interface layer is the most flexible, supporting cloud service providers to develop customized services such as video on demand and data backup according to business needs; The access layer provides users with differentiated data access interfaces[11]. The data storage methods are divided into two categories: file storage (parallel storage, downplaying folder hierarchy) and database storage (generally using high scalability, low-cost NoSQL databases). Although cloud storage has become the mainstream data storage model due to its advantages of low cost and high scalability[12], its centralized storage characteristics have caused security challenges such as loss of data control, privacy breaches, and malicious attacks[13]. To address these issues, cloud storage adopts multiple security mechanisms: encryption technology (server-side/client-side encryption) to ensure data confidentiality; Identity authentication (combining multiple modes such as single point and collaboration) to prevent unauthorized access[14]; Security access policies (data isolation, permission grading) strengthen user data independence; Secure transmission protocols (VPN, SSL, etc.) protect the data transmission process; Firewalls enhance transmission security through traffic monitoring and filtering[15], with the ultimate goal of achieving data confidentiality, integrity, and availability. Its advantages lie in breaking through the capacity limitations of traditional storage, achieving flexible resource allocation through architecture layering, supporting multi cloud migration and business customization, significantly reducing user storage costs and management complexity; Limitations

include security risks caused by the transfer of data control rights, high costs and privacy risks of third-party audits, lack of data damage location and recovery mechanisms in multi cloud migration scenarios, insufficient support for dynamic data updates, and high costs and poor real-time performance of user dynamic revocation caused by frequent administrator changes in enterprise scenarios. In the future, it is necessary to deepen research on optimization of non continuous storage, lightweight security mechanisms, and multidimensional security integration (such as combining zero trust architecture with AI detection) to promote the development of cloud storage towards a more intelligent and efficient direction.

2.2. Basic cryptographic functions

Cryptography, as the core technology of information security, covers multiple key mathematical concepts in its theoretical foundation. A cyclic group is an important structure in group theory, defined as a group G with a generator g, where each element can be represented as a power of g. Its basic characteristics include: the existence of a unit element e satisfies; Satisfy the law of union; And each element a has an inverse element, satisfying. Bilinear mapping (bilinear pairing) is a key tool for constructing modern cryptographic schemes, defined as the mapping between two multiplicative cyclic groups G and GT of order p (with generator g), which must satisfy the following properties: computability (effective algorithm exists to calculate the value of e (R, S)), bilinear property (for any a, b ∈ Zp and R, S ∈ G, e (Ra, Sb)=e (R, S) ab), non degeneracy (existence of R, S ∈ G such that it is a unit element), and commutativity (for any R1, R2, S∈G, There is e (R1 · R2, S)=e (R1, S) · e (R2, S)）。 A hash function is a compression function that maps any length input to a fixed length output, expressed mathematically as h=H (m), where h is the hash value, H is the hash function, and m is the input message. Its core properties include: availability (for any input m, H (m) can be efficiently calculated), unidirectionality (given a hash value h, it is impossible to find the original image m in polynomial time), weak collision resistance (for any m, finding such that it is computationally infeasible), and strong collision resistance (finding any pair of m and m 'that satisfies H (m)=H (m') is computationally infeasible). Hash functions are widely used in cryptography for data integrity verification, digital signatures, and other scenarios. Typical algorithms include MD5 and SHA series, which often construct data block labels or authentication structures through hash values to achieve efficient verification.

3 Research methods

3.1. Research on Integrity and Security Protection of Cloud Storage Data

Data integrity is a core requirement for cloud storage security, aimed at ensuring that data is not tampered with, forged, or damaged during storage and transmission, and is a key means of verifying data authenticity. Compared to data encryption that focuses on confidentiality (such as asymmetric encryption), integrity verification generates a unique identifier through hash functions such as MD5, which can detect data anomalies in a timely manner. The centralized nature of cloud storage poses unique challenges: traditional local verification methods are difficult to adapt to cloud based decentralized scenarios, and efficient protocols (such as POR and PDP) need to be designed to balance computing and communication overhead; At the same time, scalability should be taken into account, supporting damaged data location and recovery, as well as real-time permission revocation for enterprise users.

Among the existing integrity detection technologies, digital signatures (asymmetric encryption to generate unique identifiers) have high security but slow computation speed, MAC (shared key

computation) has high efficiency but requires key management, and digital watermarking is mainly suitable for the multimedia field, with limited applications for text data. The cryptographic foundations (cyclic groups, bilinear mappings, hash functions) provide theoretical support for mechanism design, among which the unidirectionality and collision resistance of hash functions are crucial in signature and verification. Error correction code techniques (such as erasure codes and RS codes) generate redundant data blocks through encoding, supporting the use of remaining intact blocks to recover the original data in case of data corruption (such as (n, m) erasure codes that can tolerate damage to n-m data blocks). However, traditional methods such as Vandermonde matrix encoding have high computational overhead and need to be optimized. Trusted Execution Environment (TEE) serves as a secure area for hardware isolation, protecting sensitive data (such as re signature calculations) through independent memory and authorized access interfaces, avoiding malicious attacks from Rich Execution Environment (REE), and providing a trusted computing environment for dynamic updates of enterprise user permissions.

The current method still needs to balance efficiency and security (such as slow computation of digital signatures and key management required for MAC), and the computation cost of error correction code encoding and decoding is relatively high, resulting in high deployment costs for TEE. In addition, the damage location and recovery mechanism for non continuous storage in multi cloud scenarios still needs to be improved. In the future, it is necessary to combine lightweight algorithms (such as optimizing RS code encoding), zero trust architecture, and AI detection technology to promote the development of cloud storage towards higher efficiency and intelligence.

3.2. Scene construction and spatial design of metaverse banking financial community

This chapter focuses on the gray box attack scenario and systematically explores for the first time the robustness evaluation method of information retrieval models in this scenario. In gray box scenarios, although attackers cannot obtain model architecture, parameters, or training data, they can obtain ranking results and candidate document relevance scores by calling the target model infinitely, and use this information to construct attacks. The attack target is defined as reversing the original sorting result (such as changing from $s(q, pi) > s(q, pj)$ to $s(q, pi) < s(q, pj)$) by inserting a trigger containing key target information (to maintain semantic coherence to avoid detection), while ensuring the concealment of the trigger. The robustness evaluation is based on the criterion of "adversarial failure": if the model can still maintain the original correct ranking after the attack ($s(q, pi) > s(q, pj)$), it is considered robust. The experiment used the MiniLM-L-12 model fine tuned with MS MARCO and TREC DL datasets as the evaluation object. This model was selected as a representative paragraph sorting model due to its excellent performance in public rankings. On an ethical level, the research follows the ACM Code of Ethics to avoid attacks on real systems (such as Wikipedia) and only conducts controllable testing on local models. The aim is to promote the development of defense algorithms through public vulnerabilities, rather than causing actual harm. This process is analogous to the vulnerability disclosure mechanism of white hat hackers.

3.3. Research on Integrity Verification and Fast Recovery Mechanism for Multi Cloud Data Migration

As enterprises and users widely outsource their data to multi cloud environments, the need for integrity verification and rapid recovery during data migration is becoming increasingly prominent. The incomplete trustworthiness of cloud service providers may lead to the risk of data corruption, loss, or incomplete deletion during the migration process, while traditional full validation methods incur high computational and communication overhead when dealing with large-scale data.

Therefore, this study proposes a multi cloud migration verification scheme that supports rapid damage localization and recovery (as shown in Figure1),

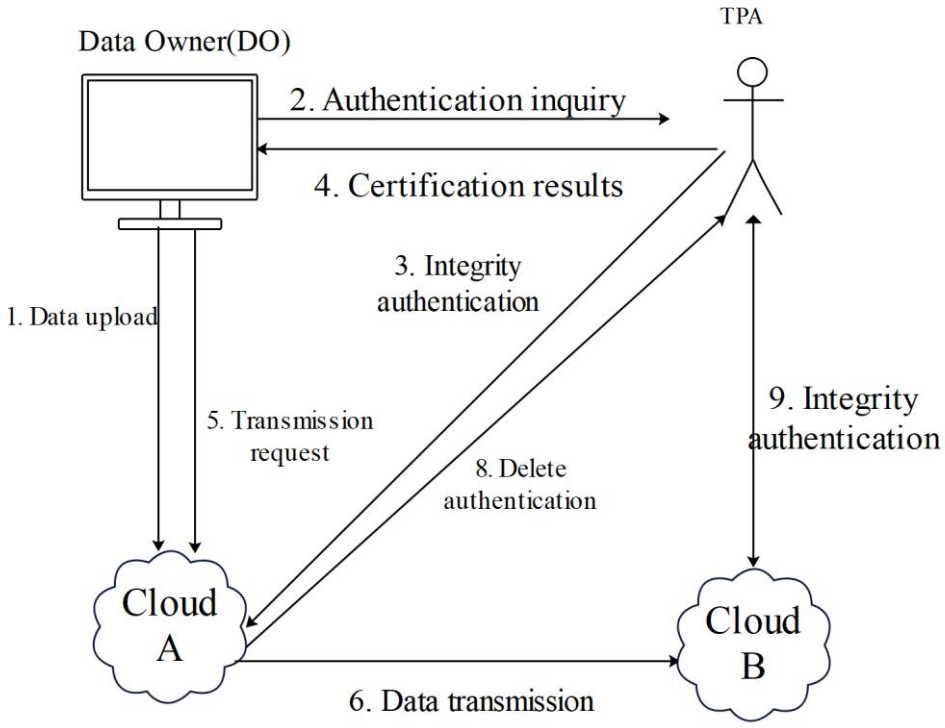

Figure 1. Can Prove the Data Migration Model Diagram

Its core model includes four entities: data owner (DO), cloud server A/B, and third-party auditor (TPA): DO is responsible for initiating migration requests and utilizing TPA's strong computing power for verification; Cloud servers A/B provide storage and transmission services; TPA, as a trusted third party, replaces DO in performing integrity verification, deletion proof verification, and migration result confirmation.

The scheme architecture is based on the extended PDP (provable data holding) pattern, and five core algorithms are designed: KeyGen generates public-private key pairs to support asymmetric verification; The Store algorithm generates data labels and root node signatures by partitioning, inserting probe blocks, and constructing a level based Merkle Hash Tree (RMHT); Transfer ensures that data is correctly transferred to the target cloud during the migration process; DeletCheck verifies whether the source cloud has completely deleted the migrated data; IntegCheck achieves continuous integrity verification of cloud data through a challenge response mechanism. To address the issue of validation efficiency for large-scale continuous data, this study introduces a hierarchical authentication mode. The basic idea is to quickly locate damaged data blocks through hierarchical validation: first, perform initial validation with coarse granularity (such as equidistant data blocks), and if failed, refine the validation scope layer by layer until the specific location of the damage is determined at the minimum granularity. This mode significantly reduces the cost of full validation and supports fast recovery of damaged data. Innovation is reflected in three aspects: firstly, the independent operation of public verification and DO through TPA, combined with RMHT and HVM modes, optimizes verification efficiency while ensuring security; The second is to remove the verification algorithm (DeletCheck) to ensure the thorough removal of source cloud data after migration, meeting compliance requirements; The third is the hierarchical authentication mode

(HVM), which dynamically adjusts the verification granularity to achieve rapid localization and recovery of damaged data. The limitations include that the current solution still needs to improve the non continuous storage damage localization in multi cloud scenarios, and the layering strategy of HVM mode may increase slight latency. In the future, AI anomaly detection and lightweight encoding technology can be combined to further enhance the real-time and reliability of multi cloud migration.

4. Results and discussion

4.1. Research on Hierarchical Verification Scheme for Damaged Data Localization and Recovery in Multi Cloud Data Migration

This article proposes a scheme for locating and recovering damaged data during cloud data migration, which includes three core algorithms: hierarchical provable integrity verification algorithm (H-IntegCheck), hierarchical provable data migration algorithm (H-Transfer), and recovery mechanism based on Reed Solomon error correction code. H-IntegCheck achieves data integrity verification through hierarchical detection granularity adjustment. The initial granularity X determines the detection block range, and TPA generates a set of random numbers to challenge CSP. CSP returns a proof containing the signature, auxiliary authentication information, and polynomial calculation results. TPA verifies the root node signature and the validity of the equation. If it fails, the granularity recursive detection is reduced until the damaged block is located. H-Transfer verifies the integrity of the source cloud data before data migration. By generating migration requests and new labels, CSP transfers data blocks, labels, and auxiliary authentication information. TPA verifies the correctness of the root node to ensure successful transmission. If damage is detected, it recursively adjusts the detection granularity to locate the problem block. The recovery mechanism based on Reed Solomon error correction codes generates redundant codes through encoding, and during decoding, the data is recovered by reverse encoding based on the damaged block number. Security analysis shows that this scheme has provable data ownership (signature unforgeability ensures CSP cannot forge valid proofs) and provable data transmission (ensures complete transmission by verifying migration request signatures and root node correctness), and H-Transfer has an average dropout rate reduced by 44% and 73% compared to traditional algorithms at correlation degrees of 2 and 5, respectively. Performance evaluation shows that the validation time of H-Transfer is significantly lower with increasing data volume compared to traditional algorithms (with an average improvement of 36.6% under changes in the number of data blocks), and it still maintains its advantage with increasing damage ratio (with an average decrease of 42.2% under changes in damage ratio). The experimental environment is Intel i5/4G memory, implemented based on PBC library, and the results verify the advantages of the scheme in terms of efficiency and accuracy.

4.2. Model experiment

This scheme ensures the security of user revocable data integrity verification through multi-dimensional cryptographic mechanisms. Firstly, correctness verification is implemented through triple equality verification in the audit stage - bilinear mapping equation verification of the association between data block labels and original data, chain equation and verification of the recursive consistency of public key parameter chains. Mathematical derivation shows that all equations strictly hold under legal operations, proving the logical consistency of the scheme. Secondly, the defense against forgery attacks is achieved through the key verification mechanism of the proxy server: if a malicious user forges a re signed key, the proxy server will verify the equation,

and since the forged key cannot satisfy the equation, the attack is effectively blocked. Furthermore, the defense against replay attacks relies on a dual privacy protection mechanism: CSP hides the linear combination μ 'through random masking techniques (and), and based on the assumption of discrete logarithm (DL) difficulty, TPA cannot solve for the random number r. At the same time, the calculation of label σ depends on the private keys of previous users, and TPA cannot extract μ ' from σ due to the lack of a private key, ensuring data privacy. Finally, collusion attacks can be classified into two scenarios: when a revoked user colludes with a CSP, the re signature key is generated from the revoked user's public key and the new user's private key. Even if the CSP obtains it, it cannot be deduced, and the re signature calculation is isolated in the trusted execution environment of the proxy server; When a revoked user colludes with TPA, TPA cannot infer data through the label σ or mask μ (relying on the DL difficulty assumption). Even if the revoked user's private key (such as) is obtained, it cannot be cracked to obtain μ 'and ensure data confidentiality. In summary, this scheme effectively resists forgery, replay, and collusion attacks through mathematical equation verification, key isolation mechanism, random masking technique, and cryptographic difficulty assumption, ensuring cloud data security. The algorithm flow is shown in Figure 2.

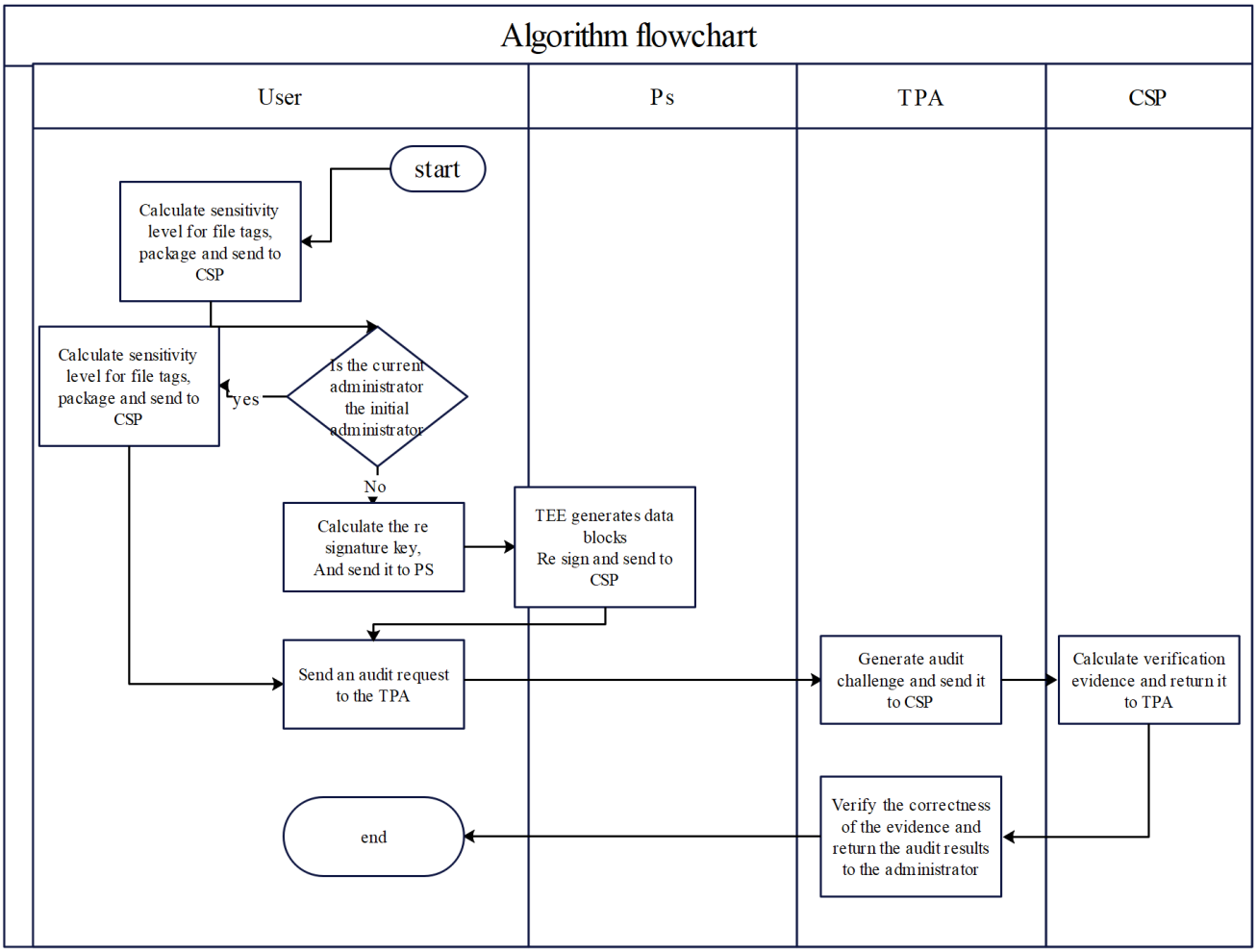

Figure 2. Algorithm Flowchart

4.3. Effect analysis

This study proposes a cloud storage data verification scheme that supports real-time user revocation, and its effectiveness has been verified through security analysis and performance evaluation. The scheme ensures data integrity through a multi-level verification mechanism: the audit stage combines initial equation verification and extended equation verification (including equation 4-3), uses bilinear pairing operation to ensure the validity of the re signed key, and resists forgery attacks based on the difficulty of discrete logarithm problem (DL); Using random masking techniques to hide sensitive information (such as μ, σ) in audit evidence, combined with DL assumptions to prevent third-party auditors (TPA) from deriving raw data, effectively responding to replay attacks; By implementing key isolation design (such as revoking the user's private key independently from the new user's private key) and introducing a trusted execution environment (TEE), we can prevent collusion between revoking users and cloud service providers (CSPs) or TPAs, ensuring key and data privacy. This solution achieves comprehensive optimization in real-time, security, and anti attack capabilities, providing an innovative solution for trusted verification of cloud storage data.

Performance analysis shows that the proposed solution outperforms traditional methods in terms of communication and computational overhead. The communication overhead mainly includes re signature transmission and audit interaction (TPA challenge), CSP return proof, and the total overhead is; The computational cost is focused on re signature generation and audit verification, with a total cost of. Comparison with existing solutions (Table 2)

Table2. Comparison of Scheme Performance and Security Attributes

Scheme	Resistance to Collusion between Revoked Users and CSP	Resistance to Collusion between Revoked Users and TPA	Real-Time Dynamic User Revocation	Data Privacy Protection	Resignature Key Security	Communication Overhead	Computational Overhead
[39]	√	×	×	√	×	$(2d+1)	G
[40]	√	×	×	√	×	$4(c+1)	G
[41]	√	√	×	√	×	$(2 + 2d)	G
Our Scheme	√	√	√	√	√	$(1 + 2d + 3n)	G

This scheme performs excellently in supporting real-time dynamic revocation, resisting collusion attacks, re signing key security, and data privacy protection. Although the communication overhead slightly increases (due to the introduction of proxy servers), the computational efficiency is significantly improved (such as reducing the number of bilinear operations).

The experimental verification further supports the theoretical analysis. In the Ubuntu environment configured with Intel i5-7200U CPU and 8GB memory, the Python implementation based on PBC library shows that as the number of data blocks n increases (c=100 when n=500, c=200 when n=1000), the advantages of this scheme in communication and computation costs gradually expand; The time cost of user revocation remains constant due to the proxy server's re signature mechanism, which is significantly better than the traditional approach where new users need to re download and re sign all data. In summary, the solution provides an efficient and reliable solution for cloud storage data verification by balancing security mechanisms with performance optimization.

5. Conclusion

In the digital age, although cloud storage technology has become the mainstream of data

management with the advantages of relieving local storage pressure and reducing costs, the centralized storage mode has caused security challenges such as loss of data control, integrity damage, and privacy leakage, making data integrity verification the core issue of cloud storage security. This article is driven by a layered security architecture and constructs an optimization framework for identity authentication and data protection in cloud service systems. Two core contributions are proposed: firstly, a dynamic authentication framework based on hierarchical authentication is designed to address the problems of low efficiency and lack of damage handling in traditional authentication schemes in massive data migration scenarios. The hierarchical authentication mechanism is used to reduce redundant authentication overhead, and error correction code technology is integrated to achieve synchronous localization and automatic recovery of damaged data, significantly reducing authentication time overhead and optimizing migration performance; Secondly, in response to the lack of flexibility in existing user revocation mechanisms and the risk of privacy leakage, a security scheme based on proxy re signature is proposed, which introduces administrator attributes representing user tenure to support real-time permission revocation. Combined with a trusted execution environment (TEE), the data confidentiality of the remote signature replacement process is guaranteed, ensuring security and feasibility while controlling computing and communication costs. This system innovatively integrates data integrity verification with user identity authentication, forming a full chain protection covering "verification location recovery revocation", breaking through the single function limitation of traditional solutions, realizing self-healing of damaged data and dynamic management of user permissions, and reducing security costs through algorithm optimization and trusted computing technology, improving the applicability of actual deployment. Future research will focus on the optimization of discontinuous storage scenarios, the design of lightweight and heavy signature mechanisms, and multi-dimensional security integration. Combining zero trust architecture and AI anomaly detection technology, it will promote the in-depth transformation of theory to resource constrained environments such as edge computing and enterprise level complex scenarios, and provide a more intelligent and efficient layered security solution for cloud service systems.

References

[1] *Zhu, Z. (2025). Application of Database Performance Optimization Technology in Large-Scale AI Infrastructure. European Journal of Engineering and Technologies, 1(1), 60-67.*

[2] *Yang D, Liu X. Collaborative Algorithm for User Trust and Data Security Based on Blockchain and Machine Learning[J]. Procedia Computer Science, 2025, 262: 757-765.*

[3] *Wu X, Bao W. Research on the Design of a Blockchain Logistics Information Platform Based on Reputation Proof Consensus Algorithm[J]. Procedia Computer Science, 2025, 262: 973-981.*

[4] *An, C. (2025). Exploration of Data-Driven Capital Market Investment Decision Support Model. European Journal of Business, Economics & Management, 1(3), 31-37.*

[5] *Zhang Y. Research on Optimization and Security Management of Database Access Technology in the Era of Big Data[J]. Academic Journal of Computing & Information Science, 2025, 8(1): 8-12*

[6] *Lai L. Data-Driven Credit Risk Assessment and Optimization Strategy Exploration[J]. European Journal of Business, Economics & Management, 2025, 1(3): 24-30.*

[7] *Tang X, Wu X, Bao W. Intelligent Prediction-Inventory-Scheduling Closed-Loop Nearshore Supply Chain Decision System[J]. Advances in Management and Intelligent Technologies, 2025, 1(4).*

[8] *Zhang, Jingtian. "Research on Worker Allocation Optimization Based on Real-Time Data in Cloud Computing." Frontiers in Science and Engineering 5.2 (2025): 119-125.*

[9] Xu, Yue. "Research on Maiustream Web Database Development Techncl ogy." Journal of Computer Science and Artificial Intelligence 2.2 (2025): 29-32.

[10] Information V F A , Saravana K E , Information V F A ,et al.Development of Trustworthiness for Cloud Service Providers Using DBN-Based Trust Model in Cloud Computing Environment[J]. 2024.

[11] Chen A. Research on Intelligent Code Search Technology Based on Deep Learning[J]. Pinnacle Academic Press Proceedings Series, 2025, 2: 137-143.

[12] Yaojia J , Bohao L , Jiankang X ,et al.Graph-CRISPR: a gene editing efficiency prediction model based on graph neural network with integrated sequence and secondary structure feature extraction[J].Briefings in Bioinformatics, 2025(4):4.DOI:10.1093/bib/bbaf410.

[13] Wang Z , Zhou Y .Analysis and Evaluation of Intel Software Guard Extension- Based Trusted Execution Environment Usage in Edge Intelligence and Internet of Things Scenarios[J].Future Internet, 2025, 17(1):32.DOI:10.3390/fi17010032.

[14] Pan, H. (2025). Development and Optimization of Social Network Systems on Machine Learning. European Journal of AI, Computing & Informatics, 1(2), 73-79.

[15] Zhang Y , Liang W , Xu W ,et al.Cost Minimization of Digital Twin Placements in Mobile Edge Computing[J].ACM Transactions on Sensor Networks, 2024, 20(3).DOI:10.1145/3658449.