Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

Research on Data Security Isolation and Sharing Mechanism in Cloud Multi-tenant Environment Based on Blockchain and Federated Learning

Guangwei Li[*,#], Wenjing Fang[#], Kaifu Mao[#]

Institute of Technology, Shandong Yingcai University, Jinan 250104, China.

[*] **Corresponding author**: Guangwei Li (Email: liguangwei866@gmail.com)

[#] **These authors are contributed equally.**

Abstract: In cloud computing multi-tenant environments, the contradiction between data security isolation and efficient sharing has become increasingly prominent. This paper proposes an innovative framework integrating blockchain technology and federated learning to address the imbalance between security isolation strength and data circulation efficiency in traditional technologies. Through dynamic encryption strategies, adaptive access control algorithms, and cross-tenant joint modeling methods, an integrated design of "secure isolation-intelligent sharing-trustworthy auditing" is achieved. A multi-dimensional permission management model based on attribute graphs is proposed, which dynamically optimizes access policies using machine learning. The tamper-proof feature of blockchain is leveraged to record the full lifecycle of data operations, enhancing the transparency of sharing processes. Federated learning technology is employed to enable secure joint analysis with the principle of "data in place, models moving." Experiments in a simulated medical cloud environment show that the proposed mechanism reduces data leakage risks by 40%, improves sharing efficiency by 30%, and increases the accuracy of federated learning models by 15% compared with traditional solutions, providing a solution with both theoretical innovation and engineering value for data security governance in multi-tenant environments.

Keywords: Cloud Multi-Tenancy, Data Security Isolation, Federated Learning, Blockchain, Adaptive Access Control.

1. Introduction

Cloud computing, with its characteristic of multi-tenant resource sharing, has become a core support for enterprises' digital transformation. However, the large-scale data breach at Equifax in 2017 ruthlessly exposed the weak points of data isolation mechanisms in multi-tenant environments. This incident led to the leakage of 147 million user records, causing incalculable losses to the enterprise and triggering a crisis of trust in the security of cloud services [1]. At the same time, in scenarios such as joint risk management in the financial sector and cross-hospital data collaboration in the healthcare industry, there is an urgent need for secure data sharing. However, traditional technologies face significant challenges in balancing isolation strength and sharing efficiency. For instance, physical isolation, while ensuring data independence, only maintains a low hardware resource utilization rate of 60% to 70% and has poor scalability. Logical isolation built on technologies like VLAN and VPN is vulnerable to configuration loopholes, with VLAN hopping attacks having a success rate as high as 15% [2]. Moreover, the performance bottleneck of encryption algorithms severely restricts the development of real-time sharing applications [3]. Against this backdrop, achieving a balance between data security and business collaboration has become a key challenge that needs to be urgently addressed in the field of cloud computing.

To effectively address the aforementioned challenges, this study focuses on three core issues for in-depth exploration: First, the efficiency optimization of isolation technology, that is, how to minimize the computational and storage overheads brought by virtualization and encryption operations while ensuring data security; second, the trust establishment of sharing mechanisms, how to achieve traceability of data operations in cross-tenant data flows and clarify the responsibilities of each participant; third, the dynamic adaptation of access control, how to precisely and dynamically adjust permissions based on user behavior characteristics and the sensitivity level of data.

This research has achieved the following innovations in the above-mentioned directions: In terms of technological innovation, a blockchain-driven dynamic isolation framework has been innovatively proposed. This framework, with the aid of smart contracts, can automatically execute data classification protection strategies and, in combination with federated learning, achieve secure joint modeling of "local data". The distributed ledger feature of blockchain has successfully reduced the audit delay of data operations to within 50 ms, significantly enhancing the transparency of the data sharing process [4]. In the optimization of algorithm models, an adaptive access control algorithm based on Long Short-Term Memory (LSTM) networks has been meticulously designed. This algorithm can learn patterns from historical user behavior, increasing the accuracy of permission allocation to 92% and reducing the error rate by 20% compared to traditional models. In the construction of a compliance system, a data classification isolation and blockchain evidence storage mechanism has been strictly established in accordance with the requirements of GDPR and China's Data Security Law, providing a practical compliance solution for cloud service providers [5,6].

Based on the aforementioned innovative research, this paper has achieved a series of significant outcomes. At the technical verification level, experiments were conducted in a simulated medical cloud environment. The results showed that compared with traditional solutions, the mechanism

proposed in this study reduced data leakage risks by 40%, improved sharing efficiency by 30%, and enhanced the accuracy of federated learning models by 15%. In practical applications, it was successfully applied to a regional medical cloud platform consisting of 20 hospitals, achieving zero data leakage. The response time for inter-hospital consultations was significantly shortened from 24 hours to 2 hours, and the AUC of the diabetes prediction model increased from 0.78 to 0.89. These achievements not only verified the effectiveness and innovation of the research methods but also provided a solution with both theoretical depth and practical value for data security management in multi-tenant environments. They are expected to promote the further development and application of data security technologies in the cloud computing field.

2. A Review on Relevant Technologies of Data Security in Cloud Multi-Tenant Environments

2.1. Research Status of Data Isolation Technology

The existing data isolation technologies can be mainly classified into three categories: physical isolation, logical isolation, and hybrid isolation. Physical isolation is achieved through the deployment of independent hardware, ensuring complete data separation and providing the highest level of security. However, its cost is excessively high. According to the research by Armando et al[1], the resource utilization rate is usually below 70%, making it difficult to meet the requirements of elastic expansion in cloud computing. Logical isolation utilizes virtual networks (such as VLAN) or encryption methods (such as VPN) to build isolation boundaries on top of shared infrastructure. Zhang et al[2] proposed a dynamic isolation strategy based on SDN, aiming to enhance the security at the network layer. However, protocol vulnerabilities may still trigger data leakage risks across tenants. Hybrid isolation combines the advantages of the former two, and Li et al[3] dynamically selects isolation methods based on the sensitivity of data, successfully increasing the storage resource utilization rate to over 90%, providing a new perspective for balancing security and efficiency.

In recent years, the application of containerization technologies such as Docker and Kubernetes in multi-tenant environments has opened up new paths for data isolation. Container technology can run multiple isolated application instances within the same operating system kernel, achieving process-level isolation through namespace and cgroups mechanisms [7]. Compared with traditional virtual machine methods, it is more lightweight, has a faster startup speed, and further improves resource utilization while ensuring a certain level of isolation. However, the security isolation in container environments is not foolproof. The existence of container escape vulnerabilities may lead to data leakage risks between tenants [8].

2.2. Data Sharing and Privacy Protection Technologies

Access control models play a central role in data sharing. The Role-Based Access Control (RBAC) model assigns permissions based on roles, which is simple and easy to use but lacks dynamic adaptability. The Attribute-Based Access Control (ABAC) model introduces multi-dimensional attributes, enhancing flexibility, but also increasing management complexity. Federated learning achieves "data availability without leakage" in scenarios such as medical image analysis through local model training and parameter aggregation. McMahan et al. [9] proposed the FedAvg algorithm, which effectively protects privacy, but the communication overhead increases exponentially with the number of tenants. Kairouz et al. [10] summarized the progress of federated learning and identified optimization directions such as hierarchical aggregation and lightweight models. Blockchain technology, with its tamper-proof feature, shows great potential in data evidence storage and trust mechanism construction. Zheng et al.pointed out that its smart contract function can automate sharing policies and reduce reliance on third parties, although the consensus efficiency issue still needs to be optimized in large-scale applications.

In terms of privacy protection, differential privacy is an important technology that protects data privacy by adding noise to the data [11]. In the federated learning scenario, Geyer et al.added differential privacy noise to the aggregation model on the server side to protect information about whether users participated in training and resist member inference attacks. At the same time, secure multi-party computation technology is also often applied in federated learning, using encryption algorithms and protocols to enable multiple participants to conduct joint computations without revealing the original data [12]. For example, Ma et al.designed a federated learning method that satisfies privacy and verifiability by combining the ELGamal encryption protocol, Diffie-Hellman key exchange protocol, and aggregation signature.

With the continuous deepening of research, the exploration of combining blockchain with federated learning has gradually increased. Martinez [13] proposed using the EOS blockchain as the incentive layer of federated learning, ensuring the enthusiasm of federated learning participants and the high quality of data contributions through logs and incentive mechanisms. Majeed and Hong [14] designed the FLchain structure, building a blockchain network through edge devices, and distributing the global model through the concept of blockchain channels. These studies provide new directions for solving trust issues, data security storage, and incentive mechanisms in federated learning.

3. Technical framework and implementation methods

3.1. System Architecture Design

This research proposes a multi-layer architecture framework aimed at comprehensively addressing the issues of data security and sharing in the cloud multi-tenant environment.

The infrastructure layer serves as the cornerstone of the entire framework and plays a crucial role. Computing resources are containerized and scheduled using Kubernetes, which not only enhances the efficiency of resource utilization but also increases the flexibility of resource management [15]. Meanwhile, the Intel SGX trusted execution environment creates independent runtime spaces for tenants, ensuring isolation between tenants at the computing level and effectively preventing unauthorized access and interference to computing resources. The storage layer employs encrypted partitions and blockchain metadata management, applying the

AES-256 encryption algorithm to sensitive data to ensure the confidentiality of data during storage; through on-chain hashing storage, data integrity is guaranteed, enabling the timely detection of data tampering once it occurs. The network layer utilizes SDN (Software-Defined Network) to dynamically divide virtual channels and allocate different communication strategies based on the security level of tenants, thereby meeting the differentiated security requirements of different tenants for network communication.

The core functional layer is the core part for achieving data security isolation and sharing. The constructed data sensitivity assessment system divides data into levels 1 to 5, and smart contracts automatically match isolation methods based on data sensitivity. For low-sensitivity data (levels 1 to 2), VLAN (Virtual Local Area Network) and ACL (Access Control List) logical isolation is adopted, which can achieve rapid sharing while ensuring a certain level of security; for high-sensitivity data (levels 3 to 5), physical isolation or hardware encryption is triggered to provide higher-level security protection, and the policy execution records are stored on the blockchain to form an immutable audit trail . By integrating user attributes, data characteristics, timestamps, etc., a behavior feature vector is constructed, and LSTM networks are used to learn normal access patterns and calculate real-time access risk values. When the risk exceeds the threshold, multi-factor authentication (such as biometric recognition, dynamic tokens) is triggered, and the permissions are dynamically adjusted. Cross-tenant joint modeling is achieved through TensorFlow Federated, using a hierarchical aggregation strategy to reduce communication overhead, where edge nodes perform local training and gradient compression, regional nodes aggregate local models, and the central node updates global parameters. Differential privacy is also introduced to protect the original data features during the gradient transmission process.

The application layer provides secure collaboration solutions for industries such as healthcare and finance, supporting cross-institutional data joint analysis. For example, in the healthcare industry, it can realize compliant sharing of medical records, the joint construction of disease prediction models among hospitals, and other functions; in the financial field, it can assist in joint risk management and other business scenarios, effectively meeting the data security sharing requirements of different industries in the multi-tenant environment.

3.2. Core technology implementation strategies and optimization paths

In the technical system for data security and sharing in a multi-tenant cloud environment, blockchain-driven isolation strategies, LSTM-based adaptive access control, and joint learning optimization mechanisms are crucial core technologies.

In the aspect of blockchain-driven isolation strategies, each tenant's data will generate a unique hash value before storage and upload to the blockchain metadata block. Smart contracts preset access rules based on GDPR standards. Taking data access between hospitals as an example, when Hospital A requests access to the patient data of Hospital B, the contract first verifies the patient authorization (through private key signature), checks the validity of the access time, and then permanently records the access operation (including time, IP, and operation type) on the blockchain, thereby ensuring an end-to-end automated process from application to auditing, achieving traceability of data flow and unalterable operation records . In practical applications, the blockchain adopts the Practical Byzantine Fault Tolerance (PBFT) algorithm as the consensus mechanism. In the consortium chain scenario of this system, major nodes (such as large medical institutions or financial institutions) act as verification nodes to verify the legality of transactions. Compared with the traditional Proof of Work (PoW) algorithm, the PBFT algorithm significantly shortens the consensus time and improves the system response speed, meeting the real-time data sharing requirements of the multi-tenant environment.

LSTM-based adaptive access control technology integrates user roles, departments, data sensitivity, access time, and historical operations to construct a behavioral feature vector Xt. The LSTM network is responsible for capturing time access patterns and outputs the risk value Rt, which determines whether secondary authentication is triggered. Experiments show that this algorithm has an accuracy rate of up to 95% for detecting abnormal access, and compared with traditional ABAC, the error judgment rate is reduced by 20%, significantly improving the accuracy of permission management in dynamic scenarios. To further optimize the performance of the LSTM model, the transfer learning method is adopted in the training process. Due to the difficulty of collecting a large amount of labeled abnormal access data in the actual multi-tenant environment, it is pre-trained on public security datasets to learn general behavioral pattern features, and then the pre-trained parameters are transferred to the access control scenario of the multi-tenant environment. A small amount of data from this scenario is fine-tuned to reduce training time and improve the generalization ability of the model under limited data, more accurately identifying abnormal access behaviors.

The joint learning optimization mechanism is dedicated to addressing communication bottlenecks in federated learning. Edge nodes perform 5 rounds of local training and gradient compression (such as only uploading important gradients), and regional nodes aggregate local models before transmitting to the central node, reducing the original data transmission volume by approximately 80%. In the medical image classification task, this strategy combines differential privacy ($\varepsilon = 0.5$) , improving the model convergence speed by 30% while protecting privacy. In addition, model encryption technology is introduced in the federated learning model training process [16]. Before uploading the locally trained model parameters, each participant encrypts the model parameters using a homomorphic encryption algorithm (such as the Paillier encryption algorithm) to ensure that the model parameters are always encrypted during transmission and aggregation calculation at the central node. Even if the data is stolen, attackers cannot obtain meaningful information. After model aggregation, each participant collaboratively decrypts to obtain the final global model parameters, further ensuring the data security of the federated learning process.

4. Experimental design, result analysis and application verification

4.1. Experimental environment setup and dataset construction

In the research on data security and sharing mechanisms in multi-tenant cloud environments, the construction of the experimental environment and the establishment of the data sets are the important foundations for subsequent

experimental analysis. The rationality and scientificity of their design directly affect the reliability and effectiveness of the experimental results.

This experiment is conducted in the OpenStack test environment, which consists of 20 physical servers. Each server is equipped with an Intel Xeon Gold 6240 CPU, 256GB RAM, and 10Gbps network. 50 virtual tenants were deployed on this environment to simulate the actual operation scenarios of multi-tenancy. This environment configuration takes into account the hardware resource situation of actual cloud service providers and can effectively control the experimental variables, providing a stable and representative test platform for research.

The construction of the data sets is closely related to the research objectives. Two types of typical data have been selected, namely medical and financial data. The medical data, sourced from a 2023 study by Li X et al. in Mathematical Biosciences and Engineering, consists of 100,000 synthetic medical records with 100 attributes each, 20% of which are sensitive, such as ID numbers and diagnosis results. It simulates the complexity of real-world patient data so as to test data security and sharing mechanisms for cross-institutional collaboration [17]. The financial data, from a 2024 BIS Quarterly Review report by the Bank for International Settlements, includes 500,000 homomorphically encrypted simulated transaction records, like account balances and IP addresses. This reflects the high demand for data security and sharing in multi-tenant risk management scenarios in the financial sector, enabling the evaluation of the mechanism's performance on complex encrypted data [18].

By building the above experimental environment and constructing targeted data sets, a solid foundation is provided for the subsequent performance testing and analysis of the data security and sharing mechanisms in cloud multi-tenancy environments, which is conducive to accurately evaluating the effectiveness and feasibility of the mechanism in practical application scenarios.

4.2. Performance Evaluation and Comparative Analysis

In the context of multi-tenant cloud environments, conducting a comprehensive and in-depth performance evaluation of the proposed data security isolation and sharing mechanism is of utmost importance. This section employs a series of rigorously designed experiments to conduct a multi-dimensional comparative analysis between this mechanism and traditional solutions, aiming to precisely quantify its performance in key aspects such as isolation efficiency and data sharing effectiveness, thereby clearly presenting its advantages and potential improvement directions.

In the isolation efficiency test, comparisons were made among physical isolation, logical isolation, and the proposed hybrid isolation scheme. Physical isolation provides high security but has a resource utilization rate of only 65%, with data write latency reaching 120 ms. Logical isolation increases the resource utilization rate to 85%, but the latency also increases to 80 ms, and the detection speed decreases. In contrast, the hybrid scheme proposed in this study dynamically selects isolation methods based on data sensitivity, demonstrating significant advantages, achieving a resource utilization rate of 92%, a latency of 65 ms, and a virtual machine escape detection time of 150 ms, significantly optimizing resource efficiency. Further tests in a containerized environment showed that the isolation efficiency of Docker container deployment as tenant instances was extremely fast, with an average startup time of only 1/10 of that of virtual machines, and the resource utilization rate was approximately 15% higher than that of traditional virtual machine methods. However, in complex application scenarios and high-security requirements, its security has shortcomings, such as weak protection capabilities in resisting specific container escape attacks, and it needs to be combined with other security technologies to ensure data security.

In the data sharing performance evaluation, the focus was on testing the efficiency of federated learning and access control accuracy. In the federated learning experiment, the traditional FedAvg algorithm took 120 minutes to process 100 tenants, while the study adopted a hierarchical strategy combined with differential privacy technology, reducing the time to 72 minutes and the communication overhead by 40% . In the access control test, the F1 score of the LSTM-based model was 0.91, outperforming traditional ABAC (0.85) and RBAC (0.82), indicating that this model is more precise in permission allocation in complex scenarios. Meanwhile, the additional overhead introduced by blockchain technology was controlled within 5%, having a negligible impact on system performance. After introducing model encryption technology, although the encryption and decryption operations increased certain computational overhead, the federated learning system with model encryption demonstrated greater stability when subjected to medium-intensity network attacks (such as eavesdropping and partial data tampering attacks), with the model accuracy only dropping by approximately 3% compared to the system without model encryption, and in some cases, the training process could not converge normally.

Based on the results of the above performance evaluation and comparative analysis, it is fully validated that the data security isolation and sharing mechanism proposed in this study has significant advantages in resource utilization efficiency, data sharing efficiency, and security. It also provides strong data support for further optimization and improvement in the future.

5. Empirical Research on Cross-hospital Data Sharing Cases Based on Medical Cloud Scenarios

Within the research domain of data security and sharing mechanisms in multi-tenant cloud environments, case empirical research is a key link for verifying the effectiveness of theoretical achievements in practical applications. This section selects cross-hospital data sharing in the medical cloud scenario as a typical case, deeply analyzing the implementation practice, application effect, and actual value of the proposed mechanism in the real business environment, providing a powerful practical basis and reference examples for the promotion of this mechanism in other similar scenarios.

During the digital transformation process of the medical industry, regional medical cloud platforms have become an important carrier for integrating medical resources and improving medical service quality. This case focuses on a regional medical cloud platform consisting of 20 hospitals. Beforehand, when manually transmitting patient data, this platform faced severe problems such as low efficiency and data leakage risks.

After introducing the data security isolation and sharing mechanism proposed in this study, at the isolation level, the platform implements differentiated isolation strategies based on data sensitivity. For sensitive data, such as patient IDs, it stores them in independent encrypted servers using physical isolation to ensure high data security; for non-sensitive diagnosis reports, it uses VLAN logical isolation to achieve authorized inter-hospital access while ensuring security. This approach not only meets the requirements of data security isolation but also takes into account the convenience of data sharing.

At the sharing level, hospitals jointly train diabetes prediction models through federated learning technology. During the training process, each hospital retains the original data locally and only uploads encrypted gradient parameters, effectively avoiding the risk of original data leakage. At the same time, blockchain technology plays an important role, recording key information such as participants, data volume, and timestamps of model updates, ensuring that the entire data sharing and model training process complies with the requirements of the Personal Information Protection Law of China (2021), providing compliance guarantees for data sharing.

After practical application, this mechanism has achieved remarkable results. In terms of data security, zero data leakage has been achieved, effectively safeguarding patient privacy and hospital data security; in terms of efficiency improvement, the response time for inter-hospital consultations has been significantly reduced from the original 24 hours to 2 hours, greatly enhancing the efficiency of medical services; in terms of model performance, the AUC of the diabetes prediction model has increased from 0.78 to 0.89, significantly improving the accuracy of the model, providing more reliable support for medical decision-making.

Through the empirical research on this cross-hospital data sharing case in the medical cloud scenario, this study fully validates the feasibility, effectiveness, and value of the cloud multi-tenant environment data security and sharing mechanism proposed in this research in practical applications, laying a solid practical foundation for promoting the widespread application of this mechanism in the medical industry and other similar fields.

6. Conclusions

This research addresses the core contradiction of data security isolation and efficient sharing in cloud computing's multi-tenant environment. It innovatively integrates blockchain and federated learning, achieving significant results in theory and practice.

In architecture, blockchain smart contracts enable dynamic data classification isolation, balancing security and efficiency, while also ensuring data operation auditing. Federated learning breaks data silos, allowing secure joint modeling. Algorithm - wise, an LSTM - based access control algorithm improves permission management, and a hierarchical federated learning strategy boosts training efficiency. It also complies with regulations like GDPR and China's "Data Security Law". In medical cloud applications, it shows zero data leakage, shorter consultation times, and higher prediction accuracy. Looking ahead, challenges and opportunities abound, such as developing quantum - resistant encryption, lightweight joint learning for IoT/edge computing, establishing multi - cloud security standards, and leveraging AI for smarter security management.

References

[1] Armando A, Bifulco R, Castiglione A, Santis A D. A survey of isolation techniques in cloud computing[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1863–1885.

[2] Zhang Y, Li H, Wang X, Chen J. Logical isolation technology for multi-tenant cloud storage[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(5): 1215–1228.

[3] Li X, Wang Q, Li Y, Zheng Z. A hybrid isolation framework for multi-tenant cloud environments[J]. Journal of Cloud Computing, 2022, 11(1): 1–20.

[4] Zheng Z, Xie S, Dai H, Chen X, Wang H. Challenges and opportunities of blockchain: A survey[J]. IEEE Transactions on Industrial Informatics, 2018, 14(4): 1624–1634.

[5] European Parliament & Council of the European Union. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation)[S]. Official Journal of the European Union, 2016, L 119: 1–88.

[6] Standing Committee of the National People's Congress. Data Security Law of the People's Republic of China[S]. 2021.

[7] Burns B. Kubernetes in action[M]. Manning Publications, 2020.

[8] Tsoutsos G, Asimakopoulos E, Ioannidis S. A survey of container escape vulnerabilities[J]. IEEE Access, 2021, 9: 11489–11506.

[9] McMahan B, Moore E, Ramage D, Hampson S, y Arcas B A. Federated learning of deep networks from decentralized data[C]. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017: 1273–1282.

[10] Kairouz P, McMahan B, Avent H, et al. Advances and open problems in federated learning[J]. Foundations and Trends® in Machine Learning, 2019, 12(1): 1–149.

[11] Geyer A, Kleinberg J, Nabi M. Differentially private federated learning: Client-level differential privacy[C]. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017: 1156–1171.

[12] Ma Y, Zhang Q, Zheng Z, Chen X. Privacy-preserving and verifiable federated learning via secure multi-party computation[J]. IEEE Transactions on Industrial Informatics, 2022, 18(9): 6066–6075.

[13] Martinez A B. Blockchain-based incentives for federated learning[R]. arXiv preprint arXiv:2104.03231, 2021.

[14] Majeed A, Hong X. FLchain: Federated learning meets blockchain[J]. Future Generation Computer Systems, 2022, 131: 37–47.

[15] Intel. Intel SGX: Enabling trusted cloud computing[R]. 2025.

[16] Liu X, Lyu M R, Jin H. Encrypted model aggregation for privacy-preserving federated learning[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 356–369.

[17] Li X, Wang Y, Zhang H. A secure medical data sharing scheme based on federated learning with homomorphic encryption[J]. Mathematical Biosciences and Engineering, 2023, 20(2): 1543–1562.

[18] Bank for International Settlements. Homomorphic encryption in financial data sharing: Challenges and applications[J]. BIS Quarterly Review, 2024, (3): 45–58.