ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

Theoretical Analysis, Practical Defects and Optimization Paths of Coupling Structures Between Blockchain and Federated Learning

Xukai Cui [1, a] , Hui Wang [1, b] *

[1] Nanjing University of Science and Technology, Nanjing City, Jiangsu Province, 210094, China

[a] 125127224238@njust.edu.cn

[b] 29929309@qq.com

Abstract. Federated learning, featuring "data available but not visible", enables cross-entity collaborative modeling while protecting data privacy. However, its traditional centralized architecture suffers from inherent drawbacks, including single-point failure, lack of trust, and insufficient incentive mechanisms. Blockchain technology, with its decentralization, immutability, and smart contract automation features, provides a targeted technical solution to address these issues. Recent studies have proposed three core coupling structures for the integration of blockchain and federated learning, namely the fully decoupled type, the overlapping type, and the overlapping transformable type, laying a solid framework for the in-depth integration of the two technologies. Taking these three architectures as the research object, this paper systematically analyzes their core connotations and theoretical adaptability by drawing on practical literatures from multiple fields, and focuses on exploring their prominent problems and underlying technical causes in practical deployment. On this basis, it puts forward operable optimization schemes from three dimensions: structure-specific optimization, common technical breakthroughs, and scenario-based adaptation, clarifies key directions such as lightweight integration and dynamic mechanism design, and provides technical references for the practical deployment and performance improvement of BlockFed architectures.

Keywords: Blockchain; Federated Learning; Coupling Structure; Optimization Path

1. Introduction

1.1 Research Background

The contradiction between data element value release and privacy protection is a key digital economy bottleneck. Sensitive data (e.g., medical records, financial data) is barred from direct sharing by privacy rules. Federated learning (FL) — which enables "data availability without exposure" via "local edge training → server aggregation → global model updates"—avoids raw data risks and is widely used in healthcare and IoT [1]. Yet traditional FL relies on centralized servers, facing single-point failure, trust deficits, and weak incentives that hinder large-scale deployment.

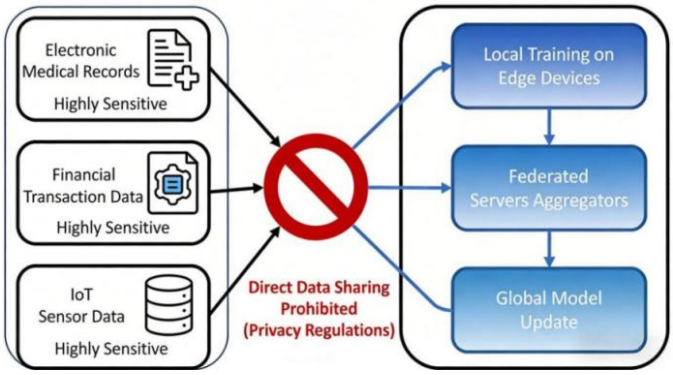


Fig. 1 Schematic Diagram of Sensitive Data Sharing Restrictions Under Privacy Regulations and Federated Learning Collaborative Modeling Process

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

The decentralization, immutability, and traceability of blockchain technology address the core drawbacks of federated learning, forming a "blockchain-driven federated learning (BlockFed)" paradigm. This paradigm leverages distributed ledgers to record training processes and smart contracts to automate incentives/supervision, building a trusted collaborative environment [2].

Specifically, blockchain features correspond to traditional federated learning (FL) problems as follows:

For FL's reliance on centralized servers (which brings single-point failure risks), blockchain's decentralized architecture uses a peer-to-peer network without core control nodes—even if some nodes fail, overall operation remains unaffected, avoiding single-point risks.

To solve the trust deficit between FL participants (which makes model updates vulnerable to tampering), blockchain's immutability and traceability permanently write model update records into linked blocks; any tampering triggers abnormal hash values, thus ensuring data and model integrity.

For FL's insufficient incentives (which reduce participants'willingness to contribute), blockchain's smart contract automation pre-sets contribution evaluation and reward rules, automatically distributing rewards based on node participation and model quality to boost engagement.

To address FL's opaque training process (which hinders responsibility tracing), blockchain's transparency and traceability synchronously records training, model updates, and aggregation results on the chain for all nodes to view; when issues arise, the responsible entity can be quickly identified.

Against FL's weak node identity authentication (which leaves systems open to malicious node attacks), blockchain combines asymmetric encryption (for node identity verification) and consensus mechanisms (e.g., PoW, PoS) to screen trusted nodes, resisting malicious node injection.

However, the integration process of these technologies faces challenges such as performance loss and mechanism adaptation. Therefore, exploring a reasonable architecture design has become the key to promoting the practicalization of the technology [3].

1.2 Research Status

Current BlockFed research focuses on three main directions: security enhancement, incentive mechanism design, and performance optimization. Some studies have proposed three core coupling structures, which adapt to the needs of different scenarios by adjusting the functional allocation between blockchain and federated learning [4]. Existing studies have initially verified the differentiated characteristics of the structures, but the analysis of their rationality, practical defects, and adaptation boundaries is not systematic, and the optimization ideas need to be further improved.

1.3 Research Content and Significance

Taking the three coupling structures as the research object, this paper clarifies their core connotations and theoretical adaptability, analyzes the inherent defects in practical applications, and proposes targeted optimization strategies. The research can provide support for improving the relevant theoretical system and provide references for architecture selection in practical deployment, reducing the cost of technology integration and implementation.

2. Relevant Basic Theories and Coupling Structures

2.1 Fundamentals of Federated Learning

Federated learning is a distributed machine learning paradigm, characterized by "data available but not visible", which allows multiple participants to construct a global model through collaborative training without sharing raw data [1]. Its typical training process includes as Fig. 2: the central server issues initial models; each participant trains on local data to generate model

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33
updates; the server aggregates the updates to obtain a global model; the global model is fed back to participants for iterative optimization until the model converges [4].

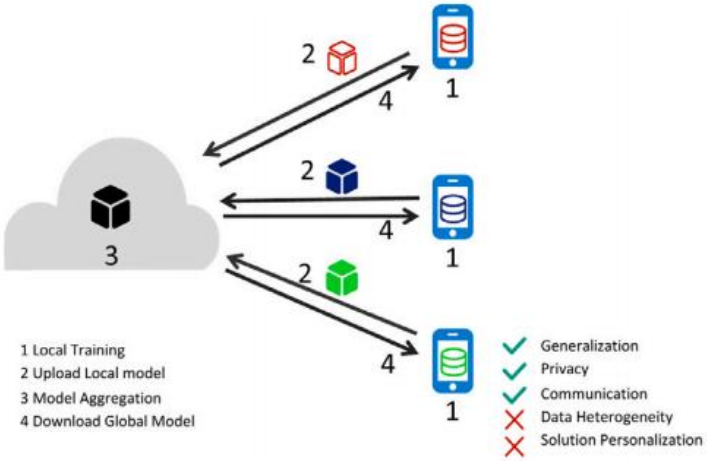


Fig. 2 Typical Training Process of Federated Learning and Its Key Characteristics

According to data distribution characteristics, federated learning (FL) falls into three modes, each tailored to distinct scenario needs. However, traditional FL relies on a centralized server for model aggregation and coordination, carrying inherent limitations—including single-point failure, data privacy leakage risks, lack of trust among participants, and insufficient incentives—that restrict its large-scale deployment [5,6].

The three FL modes are defined as follows:

Horizontal Federated Learning: Participants share consistent data feature dimensions but distinct samples; only model parameter updates are exchanged. For example, multiple hospitals collaborate to train diagnostic models using data (with features like "age/symptoms") for the same disease.

Vertical Federated Learning: Participants have highly overlapping samples but different feature dimensions; intermediate calculation results are shared after encrypted sample alignment. A typical scenario is banks (with user financial data) and e-commerce platforms (with user consumption data) collaborating to build credit evaluation models.

Transfer Federated Learning: Participants have large discrepancies in data distribution (e.g., differences in sample size or features); knowledge sharing is enabled via transfer learning. For instance, tertiary hospitals (with sufficient data) transfer disease diagnosis model capabilities to primary hospitals (with scarce data).

2.2 Fundamentals of Blockchain Technology

Among the core components of blockchain, blocks are the basic units for data storage. As shown in Fig. 3, each block contains information such as block number, hash of the previous block, and transaction data, forming an immutable chain structure through hash association. Consensus mechanisms, smart contracts, and encryption algorithms are responsible for ensuring node consistency, automating rule execution, and protecting data security, respectively [7].

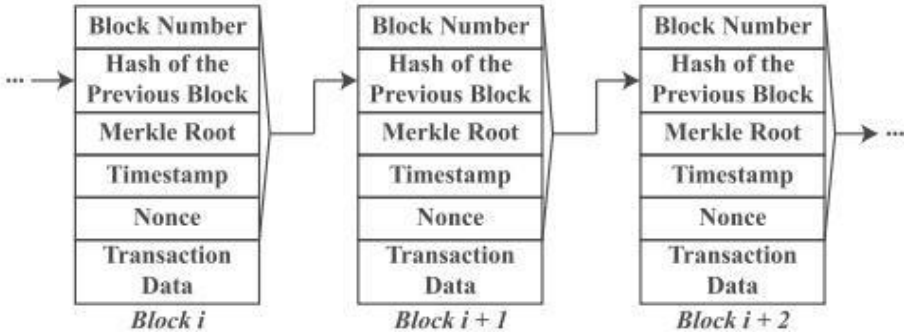


Fig. 3 Schematic Diagram of the Typical Structure of a Blockchain Block.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

Blockchain realizes trust collaboration without a central authority through consensus mechanisms (such as Proof of Work (PoW) and Proof of Stake (PoS)), and ensures data integrity and node identity legitimacy with encryption technology. Its distributed architecture can naturally avoid single-point failure risks, providing technical support for solving the centralization drawbacks of federated learning [6,7].

2.3 Overview of Three Coupling Structures

The integration of blockchain and federated learning forms three core coupling structures through functional allocation and node role design. Essentially, it adapts to the needs of different scenarios by adjusting the integration degree between the two [4]:

Fully Decoupled Type: System nodes are divided into two categories, undertaking only federated learning tasks or blockchain tasks respectively. The two types of nodes interact through pre-set rules. The blockchain subsystem is responsible for recording the training process and aggregating models, while the federated learning subsystem focuses on local training and model update upload.

Overlapping Type: All nodes have both federated learning and blockchain functions, participating in local model training, global model aggregation, as well as block generation and consensus verification, realizing in-depth functional integration.

Overlapping Transformable Type: There are three types of nodes, including clients that only participate in federated learning, nodes that only participate in blockchain, and composite nodes that undertake both functions. Node roles can be dynamically adjusted to balance the degree of decentralization and resource consumption [5].

3. Core Ideas and Design Rationality of the Three Coupling Structures

3.1 Fully Decoupled Type

The architecture of the fully decoupled type is shown in Fig. 4. Its process can be refined into six steps: (1) Client Calculation: Federated learning clients complete local model training based on local data to generate model update parameters; (2) Model Upload: Clients encrypt and upload model update parameters to blockchain nodes; (3) Model Verification: Blockchain nodes verify the validity and legality of model updates through pre-set rules to eliminate malicious parameters; (4) Model Aggregation: Blockchain nodes aggregate all valid model updates using the federated averaging algorithm (FedAvg) to generate a global model; (5) Model Update: Blockchain nodes feed back the global model to each federated learning client; (6) Reward Distribution: Smart contracts automatically calculate rewards based on client contributions and distribute them to corresponding clients through blockchain nodes.

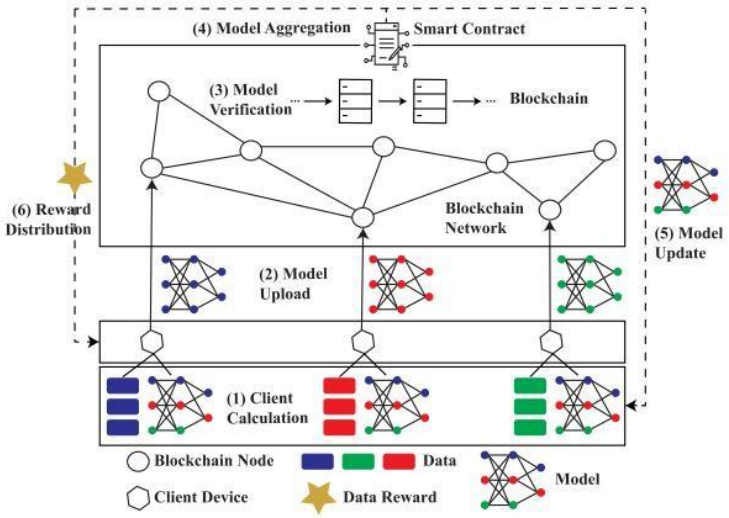


Fig. 4 Architecture Diagram of the Fully Decoupled Type.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

The core idea of this architecture is to adopt a dual-subsystem design with separated functions, assigning the model training of federated learning and the ledger recording/consensus verification of blockchain to independent nodes. Interaction such as model upload, result feedback, and reward distribution is only realized through pre-set interfaces [4]. Each node undertakes a single function: federated learning nodes focus on local training and model updates, while blockchain nodes are responsible for verification, aggregation, and recording.

The rationality and innovation of the design are reflected in two aspects: first, it reduces the resource threshold of nodes, as nodes do not need to bear both high-intensity computing and consensus tasks, adapting to terminal device scenarios with limited resources; second, the architecture has a high degree of modularity, allowing separate optimization of the training algorithms of federated learning and the consensus mechanisms of blockchain, reducing the complexity of system design and maintenance. This separated design solves the resource conflict problem when a large number of nodes participate, providing a lightweight deployment solution for multi-device scenarios such as the industrial Internet of Things [8].

3.2 Overlapping Type

The architecture of the overlapping type is shown in Fig. 5. Its closed-loop process is as follows: (1) Local Calculation: Each node completes model training based on local data to generate model update parameters; (2) Model Broadcast and Verification: Nodes broadcast model update parameters to the entire network and verify the validity of parameters broadcast by other nodes; (3) Model Aggregation: Each node synchronously aggregates all valid model updates in the network to generate a global model (the aggregation result is unified due to consistent node data); (4) Model Update: Nodes optimize local models based on the aggregated global model to complete one iteration; (5) Model Update On-Chain: Nodes write model aggregation results and parameter update trajectories into blocks to complete consensus verification; (6) Reward Distribution: Smart contracts automatically distribute rewards according to nodes' training contributions and consensus participation.

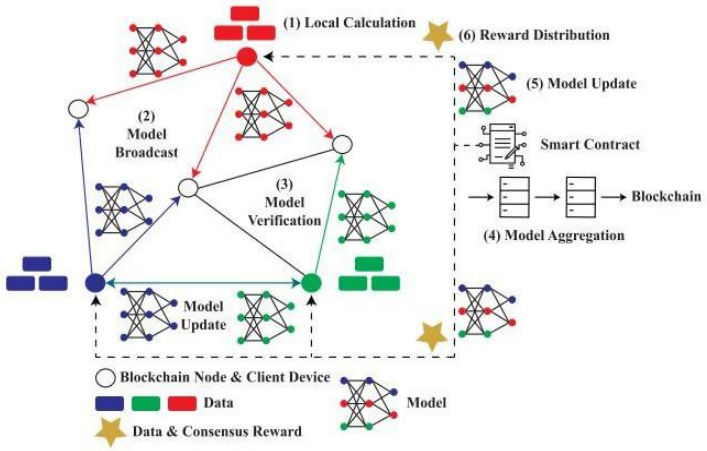


Fig. 5 Architecture Diagram of the Overlapping Type.

The core idea of this architecture is to realize in-depth functional integration. All nodes have complete functions of both federated learning and blockchain, not only training models and broadcasting updates through local data, but also participating in block generation, consensus verification, and global model aggregation [4]. Nodes communicate directly to synchronize model and ledger data without intermediate transfer, forming a closed-loop process of "training-verification-consensus-aggregation".

The rationality and innovation of the design are prominent in two aspects [9]: first, the unified node role simplifies the system topology, avoids cross-subsystem communication delays, and significantly reduces overall communication overhead; second, the full participation of nodes in consensus and training enhances the traceability and credibility of model updates, reducing the risk

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33
of tampering by malicious nodes. This design breaks through the communication bottleneck of traditional separated architectures, and is suitable for distributed collaboration scenarios with sufficient node resources and high real-time requirements, such as collaborative modeling of autonomous vehicles.

3.3 Overlapping Transformable Type

The architecture of the overlapping transformable type is shown in Fig. 6. Its core process is: (1) Local Calculation: Pure clients and composite nodes complete model training based on local data respectively to generate model updates; (2) Model Exchange: Pure clients upload model updates to composite nodes, which synchronously receive all updates and conduct preliminary verification; (3) Model Verification: Composite nodes and pure blockchain nodes collaboratively complete consensus verification of model updates to eliminate invalid parameters; (4) Model Aggregation: Composite nodes aggregate parameters based on verified results to generate a global model; (5) Model Update: Composite nodes feed back the global model to pure clients, and synchronize aggregation results and training trajectories to pure blockchain nodes to complete ledger recording; (6) Reward Distribution: Smart contracts distribute rewards differentially according to the contributions of different types of nodes.

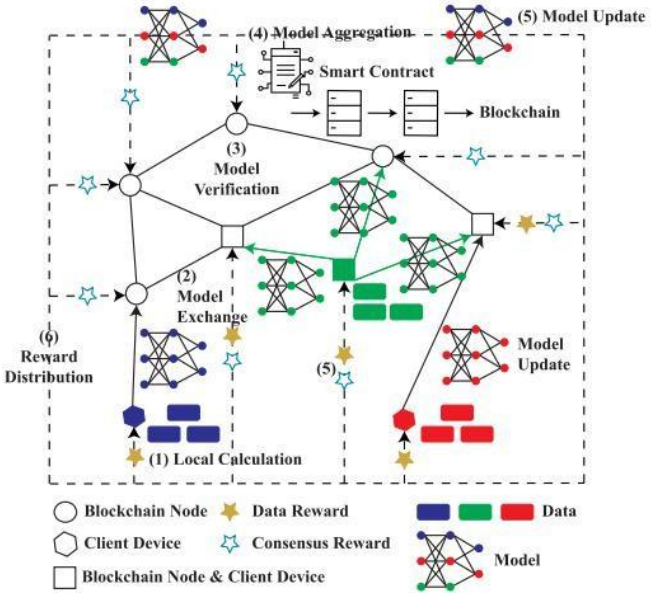


Fig. 6 Architecture Diagram of the Overlapping Transformable Type.

The core idea of this architecture is to introduce composite nodes with both federated learning and blockchain functions, forming a three-level architecture with pure federated learning nodes and pure blockchain nodes. Node roles can be dynamically adjusted according to resource status and task requirements [4]. Composite nodes undertake core coordination functions, while pure nodes are responsible for specific tasks, balancing system performance and resource consumption through flexible role allocation.

The rationality and innovation of the design are reflected in two aspects: on the one hand, it realizes optimal resource allocation through dynamic role adaptation, allowing high-resource nodes to undertake composite functions and low-resource nodes to focus on single tasks, balancing system efficiency and universality; on the other hand, it integrates the lightweight advantage of the fully decoupled type and the low-latency advantage of the overlapping type, avoiding the inherent defects of a single architecture and providing elastic expansion capabilities for complex scenarios. This compromise design solves the collaboration problem of nodes with different resource levels, and is suitable for multi-agent, heterogeneous resource application scenarios such as smart healthcare [10].

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

4. Practical Limitations and Common Defects of Coupling Structures

4.1 Practical Limitations of the Fully Decoupled Type

The fully decoupled type has four main limitations:

First, high cross-subsystem communication overhead — model transmission and verification between nodes are prone to network congestion, hampering training iteration efficiency [3,4].

Second, challenges in incentive mechanism design—quantifying contributions of two node types is difficult, easily causing unfair distribution.

Third, poor collaborative client selection — unreliable nodes are hard to identify, potentially introducing low-quality data. Additionally, separate deployment and maintenance of two independent subsystems raise hardware and operational costs.

Practical evaluation should focus on communication delay, node participation rate, and model convergence speed, requiring a cross-node distributed test environment to simulate performance under different node access scales.

4.2 Practical Limitations of the Overlapping Type

The overlapping type's limitations focus on three key points:

Strict node resource demands: Nodes must handle training, consensus, and storage tasks simultaneously—ordinary terminal devices struggle to adapt, limiting large-scale adoption [4].

Consensus-training collaboration conflicts: Dynamic parameter matching between the two is challenging, often causing aggregation delays or consensus instability.

Concentrated security risks: A single node compromise could trigger both data leakage and ledger tampering. This type only suits high-resource, high-trust closed environments.

Evaluation should prioritize resource occupancy, collaboration efficiency, and security; testing requires high-performance node clusters to verify stability.

4.3 Practical Limitations of the Overlapping Transformable Type

This type's limitations stem from architectural complexity and dynamic mechanism challenges:

Complex design & scheduling: Three-level nodes need refined functional division, protocols, and permissions; dynamic role adjustment relies on complex scheduling algorithms—poor design may cause unreasonable allocation or long delays [4].

Superimposed mechanism difficulties: Differentiated incentive rules for diverse nodes (to align with consensus mechanisms) risk interest conflicts and consensus splits.

State synchronization issues: Role switches require synchronous updates to permissions, storage, and tasks—this may cause data inconsistency, model loss, or ledger delays.

The practical threshold is high; evaluation should focus on scheduling efficiency, synchronization consistency, and fault tolerance, tested via heterogeneous node environments simulating role changes or node failures.

4.4 Common Defects of Coupling Structures

Beyond personalized limitations, the three structures share four core bottlenecks for BlockFed's large-scale deployment:

Performance-security trade-off: Blockchain consensus/encryption increases delay, while function simplification weakens trust [4,6].

Poor scalability: Node count, model dimensions, or task complexity growth causes significant performance drops.

Compatibility gaps: No unified standards for FL algorithms, blockchain consensus, and smart contract platforms—heterogeneous system integration is difficult.

Low operability: Testing requires both distributed training and blockchain deployment, demanding high hardware/technical capabilities, limiting ordinary-scenario use [11].

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33

5. Optimization Schemes and Conclusion

5.1 Overall Improvement and Innovation Directions

In response to the inherent defects of the federated learning and blockchain coupling scheme, against the backdrop of personalized limitations and common issues across the three coupling structures, holistic improvements can be made from three core dimensions: balancing performance and security, improving scalability, and optimizing compatibility, while supporting dynamic mechanism design to enhance scenario adaptability.

Balancing performance and security [12] can be achieved by adopting a combination of lightweight consensus mechanisms and privacy protection optimization: introduce federated learning-specific consensus mechanisms such as Proof of Federated Learning, which reuses nodes' model training computing power for consensus verification to eliminate extra resource overhead, while dynamic noise adjustment for differential privacy (DP) balances privacy and model accuracy by flexibly tuning noise intensity per scenario requirements, and a hierarchical encryption strategy applies differentiated encryption to model parameters and training trajectories to reduce overhead while maintaining security [3].

Improving scalability [13] relies on a hierarchical blockchain architecture and dynamic load balancing mechanism for effectiveness: a two-level "main chain-side chain" architecture offloads local model training, parameter verification, and data storage to side chains, leaving the main chain to handle only global consensus, reward distribution, and core rule storage, which reduces single-chain load pressure, while a dynamic load balancing algorithm flexibly allocates tasks and consensus nodes based on each side chain's task volume and node resource status, preventing local overload and enhancing overall system scalability.

Optimizing compatibility and operability [14] can be realized by building a standardized interface and adaptation framework: unified interface specifications for FL and blockchain integration will align model parameter formats, data interaction protocols, and consensus adaptation standards to support seamless integration of mainstream algorithms and platforms, while a lightweight integration toolkit encapsulates core functions like consensus optimization, encryption processing, and role scheduling, simplifying heterogeneous system integration and lowering deployment and operational thresholds.

Optimizing dynamic role scheduling [15] addresses the overlapping transformable type's challenges through mechanism improvement: a reinforcement learning-driven intelligent scheduling algorithm automatically adjusts node roles and task allocation based on real-time node resources, task priorities, and historical contribution data, balancing resource utilization and system performance, while a multi-dimensional contribution evaluation system quantifies node contributions to enable differentiated incentives and resolve unfair distribution issues.

5.2 Conclusion

The coupling schemes of federated learning and blockchain, through three differentiated structures (fully decoupled, overlapping, and overlapping transformable), provide innovative ideas and technical paths for decentralization, privacy protection, and trust construction in distributed training. Its core value lies in precise complementarity: using the decentralization and immutability of blockchain to make up for the centralization defects of federated learning, solving the problems of lack of trust and single-point failure; at the same time, relying on the privacy-preserving collaborative training mode of federated learning to expand the application boundary of blockchain in data value integration scenarios, realizing the dual goals of "privacy protection" and "trusted collaboration".

However, such coupling schemes generally face common problems such as the inherent contradiction between performance and security, insufficient scalability, poor compatibility, and weak practical operability. In addition, each structure has personalized limitations, restricting its large-scale deployment in practical scenarios. In the future, through optimization directions such as

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P33
lightweight consensus design, hierarchical architecture deployment, standardized adaptation frameworks, and dynamic role scheduling, existing defects can be effectively alleviated, and the practicality and deployment capabilities of the system can be improved.

In general, the coupling of federated learning and blockchain is an effective path to solve the contradiction between data privacy protection and collaborative training. Its improvement and optimization need to focus on the three cores of "efficient resource utilization, flexible function adaptation, and easy system deployment", emphasizing both technological innovation and practical feasibility. Future research can further focus on scenario-specific adaptation optimization, quantify the effect of optimization schemes through simulation experiments and empirical tests, and promote the transformation of BlockFed technology from theoretical research to industrial application.

References

[1] Zheng Z, Xie S, Dai H N, et al. Blockchain challenges and opportunities: A survey[J]. International journal of web and grid services, 2018, 14(4): 352-375.

[2] Issa W, Moustafa N, Turnbull B, et al. Blockchain-based federated learning for securing internet of things: A comprehensive survey[J]. ACM Computing Surveys, 2023, 55(9): 1-43.

[3] Qu Y, Uddin M P, Gan C, et al. Blockchain-enabled federated learning: A survey[J]. ACM Computing Surveys, 2022, 55(4): 1-35.

[4] Zhu J, Cao J, Saxena D, et al. Blockchain-empowered federated learning: Challenges, solutions, and future directions[J]. ACM Computing Surveys, 2023, 55(11): 1-31.

[5] Zhao Y, Zhao J, Jiang L, et al. Privacy-preserving blockchain-based federated learning for IoT devices[J]. IEEE Internet of Things Journal, 2020, 8(3): 1817-1829.

[6] Qammar A, Karim A, Ning H, et al. Securing federated learning with blockchain: a systematic literature review[J]. Artificial Intelligence Review, 2023, 56(5): 3951-3985.

[7] Chai H, Leng S, Chen Y, et al. A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(7): 3975-3986.

[8] Singh S, Rathore S, Alfarraj O, et al. A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology[J]. Future Generation Computer Systems, 2022, 129: 380-388.

[9] Wang X, Zhao Y, Qiu C, et al. InFEDge: A blockchain-based incentive mechanism in hierarchical federated learning for end-edge-cloud communications[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(12): 3325-3342.

[10] Li Y, Chen C, Liu N, et al. A blockchain-based decentralized federated learning framework with committee consensus[J]. IEEE Network, 2020, 35(1): 234-241.

[11] Ahmed A A, Alabi O O. Secure and scalable blockchain-based federated learning for cryptocurrency fraud detection: A systematic review[J]. IEEE Access, 2024, 12: 102219-102241.

[12] Ji S, Zhang J, Zhang Y, et al. LAFED: A lightweight authentication mechanism for blockchain-enabled federated learning system[J]. Future Generation Computer Systems, 2023, 145: 56-67.

[13] Yuan S, Cao B, Sun Y, et al. Secure and efficient federated learning through layering and sharding blockchain[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(3): 3120-3134.

[14] Nguyen D C, Ding M, Pham Q V, et al. Federated learning meets blockchain in edge computing: Opportunities and challenges[J]. IEEE Internet of Things Journal, 2021, 8(16): 12806-12825.

[15] Zhang S, Wang Z, Zhou Z, et al. Blockchain and federated deep reinforcement learning based secure cloud-edge-end collaboration in power IoT[J]. IEEE Wireless Communications, 2022, 29(2): 84-91.