International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 4, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n4.02
Journal homepage: https://wepub.org/index.php/IJCSIT/index

A Federated Learning Scheme for Privacy Preservation in the Internet of Vehicles

Zihao Shen, Nuo Zhang

School of Computer Science and Technology, Henan Polytechnic University, Henan Jiaozuo 454000, China

ABSTRACT

With the rapid development of Internet of Vehicles (IoV) technology, the exponential growth of vehicle data has provided core support for intelligent services such as traffic flow prediction and autonomous driving decision-making. However, as vehicle data contains sensitive information including owner identity, driving trajectories, and data collected by onboard sensors, traditional centralized training models face severe challenges regarding data privacy leakage. To address the aforementioned issues, this paper proposes a Federated Learning Scheme for Privacy Preservation in the Internet of Vehicles (FLSPP). Experimental results demonstrate that the proposed scheme can effectively resist privacy security threats, such as inference attacks, while ensuring high-precision model training. It provides a solution that balances privacy and efficiency for data sharing and collaborative learning in IoV environments.

KEYWORDS

Internet of Vehicles; Privacy Preservation; Federated Learning

1. INTRODUCTION

With the explosive growth of perceived data in the IoV, how to achieve efficient model training while ensuring user privacy has become a focal point in the industry. As an emerging distributed machine learning paradigm, federated learning [1] exhibits remarkable advantages in privacy protection: it allows vehicles to train models locally using private data and only upload gradients or model parameters to the central server, thereby physically realizing "data not leaving the local site" and effectively reducing the risk of unauthorized access to sensitive data such as location information and driving habits [2, 3]. Nevertheless, traditional federated learning still faces severe challenges in the highly dynamic scenarios of the IoV [4]. Due to its high reliance on a centralized aggregation server, the entire network will be paralyzed once the server encounters a single point of failure or malicious attack [5]; in addition, although the central server does not directly access raw data, vehicle privacy can still be inversely inferred from frequently updated model parameters through inference attacks [6]. Therefore, how to ensure the privacy protection of vehicle data in the IoV is a problem worthy of research.

To address the shortcomings of traditional federated learning in terms of decentralization, trustworthiness, and security, integrating blockchain technology with federated learning has become a prominent research focus [7, 8]. For instance, Wang et al. [9] proposed a secure federated learning architecture based on consortium blockchain, whose core logic lies in the coordinated use of homomorphic encryption and an improved Multi-Krum aggregator to eliminate anomalous models without revealing parameter details. Although this approach enhances robustness, the computational overhead and communication latency caused by ciphertext operations remain to be optimized. Lv et

al. [10] constructed a framework that integrates blockchain with federated learning, enhancing the security of model parameters by introducing differential privacy and Gaussian perturbation mechanisms. However, this improvement in privacy protection often comes at the cost of data utility, as excessive noise interference inhibits the convergence accuracy of the global model to a certain extent.

Leveraging its characteristics of decentralization, immutability, and traceability, blockchain can replace traditional centralized servers to construct a transparent and trustworthy model aggregation environment [11]. By automatically executing aggregation logic through smart contracts, blockchain not only effectively resists single points of failure but also utilizes consensus mechanisms to verify parameters uploaded by vehicles. It filters out "poisoning" data provided by malicious attackers and incentivizes high-quality nodes to participate in training through token mechanisms [12]. Based on this background, this paper proposes a Federated Learning Scheme for Privacy Preservation in the Internet of Vehicles (FLSPP). The scheme integrates the decentralized architecture of blockchain with the privacy-preserving features of federated learning. By designing a hierarchical aggregation federated learning mechanism, it significantly enhances system robustness while further reinforcing the data privacy defenses within the IoV environment.

2. SYSTEM MODEL

FLSPP constructs a multi-layer distributed collaborative computing architecture designed to ensure model training efficiency and parameter privacy within IoV environments. As shown in Figure 1, the architecture primarily consists of five key entities: the Trusted Authority (TA), the Task Publisher, Mobile Edge Servers (MESs), Roadside Units (RSUs), and vehicles.

TA: It is primarily responsible for generating and distributing system-wide public parameters and initial key pairs to establish a root of trust among entities. Additionally, it performs rigorous identity auditing and registration management for task publishers, MES nodes, RSU infrastructures, and participating vehicles, assigning them unique identifiers. It also provides traceability and revocation support in the event of malicious attacks or node violations.

Task Publisher: It is mainly responsible for constructing the initial deep learning model architecture based on specific traffic scenarios, such as autonomous driving path planning and traffic density prediction. It must also determine parameters such as the number of federated learning training rounds and convergence goals, and publish task announcements to the entire network via the blockchain. Furthermore, it broadcasts the initial model parameters to the blockchain management domain to serve as the starting point for training.

MES: Physically distributed MES nodes are divided into multiple independent logical management domains, which jointly construct a distributed blockchain framework based on domain-divided governance. In response to the perturbed noise actively added by vehicle terminals for privacy protection, MES performs denoising processing on the aggregated parameters to restore model accuracy. As a decentralized ledger, the blockchain records the global model hash values, training progress, etc., of each round, ensuring the process is tamper-proof and traceable.

RSU: As a communication access and preprocessing base station deployed at the physical road edge, it mainly undertakes the responsibility of low-latency interaction between vehicles and the core network in highly dynamic environments. RSU is responsible for receiving in real time the model parameter fragments uploaded by vehicles within its coverage area, which have been split using additive secret sharing technology [13], and conducts preliminary integrity verification and local aggregation on them to reduce communication redundancy uploaded to the MES layer.

Vehicles: As the data holders, their core responsibility is to perform gradient updates on the global model using locally collected private information, such as sensor data and driving trajectories, while ensuring that the data never leaves the vehicle. Upon completion of the local update, random noise is

first applied to the model parameters. Subsequently, additive secret sharing is employed to partition the noise-added parameters into multiple uncorrelated secret shares. These shares are then encrypted and transmitted directionally to designated RSU nodes. This achieves the objective of "raw data never leaving the vehicle, and parameter information never being leaked in its entirety."

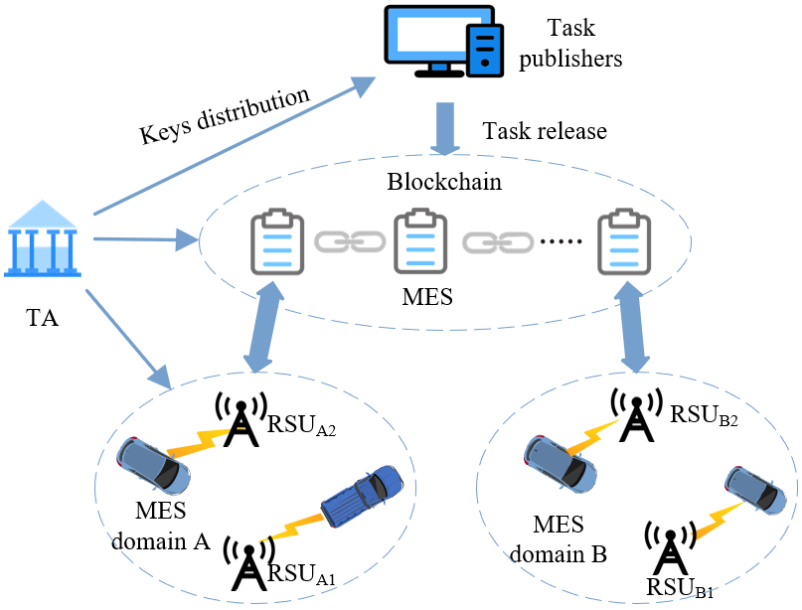

Figure 1. System model

3. IMPLEMENTATION OF THE SCHEME

3.1. System Initialization Phase

1) The task publisher releases a federated learning task to the blockchain and constructs an initial global model W_0 based on business requirements (e.g., road condition prediction).

2) The blockchain distributes the task to the sub-chains of each management domain according to geographic location information.

3) The TA generates a system-level master key and public parameters, and distributes identity certificates and asymmetric encryption key pairs to vehicles and RSUs.

3.2. Local Training

1) The vehicle downloads the global model W_g^t for the current iteration from the blockchain node.

2) The vehicle performs iterative optimization using its local private dataset D_i to derive updated model parameters W_i^{t+1}.

3) To defend against inference attacks, the vehicle injects random noise into the parameters according to Formula (1), where z_i is generated based on the shared key between the vehicle and the MES.

$$W_i^* = W_i + z_i \tag{1}$$

4) Assuming the number of RSUs is n, the vehicle uses additive secret sharing to divide W_i^* into multiple shards $S_{i,1}, S_{i,2}, \ldots, S_{i,n}$.

5) The vehicle encrypts the shards using the RSU's public key and sends them to different RSUs within its coverage area, ensuring that a single RSU cannot obtain the complete local parameters.

3.3. Local Aggregation of Roadside Units

1) Assume the number of vehicles is h. The RSU receives and decrypts the encrypted shards from all vehicles within its service coverage.

2) The RSU performs partial aggregation on the received shards and generates the regional shard summary value according to Formula (2).

$$S_{\mathrm{RSU}}^{k} = \sum_{j=1}^{h} S_{i,j} \tag{2}$$

3) The RSU submits the aggregated shard result S_{RSU}^{k} to the blockchain management domain to which it belongs.

3.4. Global Aggregation of Mobile Edge Servers

1) The MES nodes within the domain collect the aggregated shard values uploaded by various RSUs and compute a noisy global aggregation model S_i based on the additive secret sharing mechanism.

$$S_i = \sum_{k=1}^{n} S_{\mathrm{RSU}}^{k} \tag{3}$$

2) Since random noise is introduced at the vehicle side, the blockchain nodes perform denoising according to Formula (4), thereby restoring model accuracy while maintaining privacy protection.

$$W_g^{t+1} = \frac{S_i - \sum_{i=1}^{h} z_i}{|h|} \tag{4}$$

3) The MES uploads the global model parameters W_g^{t+1} to the blockchain.

The flowchart of the federated learning process is illustrated in Figure 2.

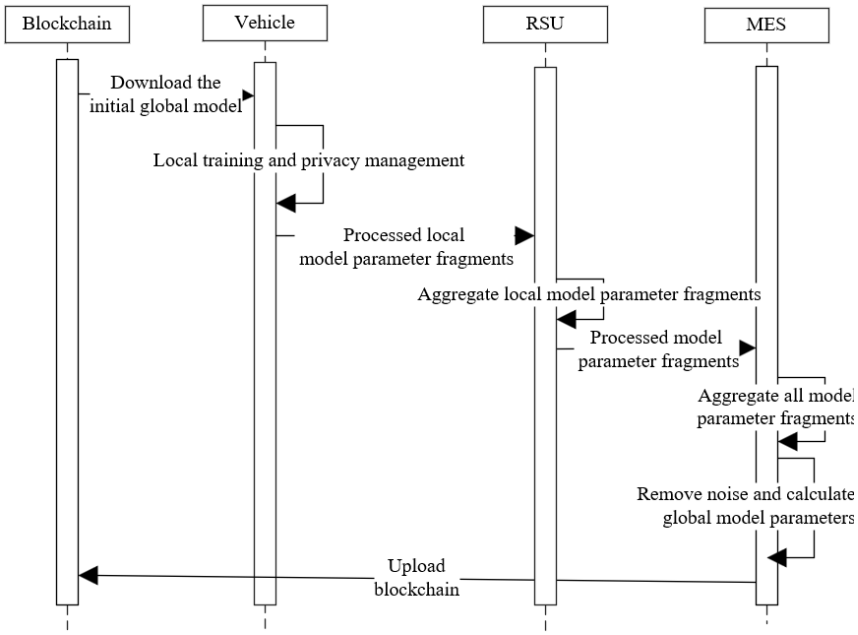

Figure 2. Federated learning process diagram

4. SAFETY ANALYSIS

4.1. Defending Against Inference Attacks

By introducing a noise mechanism during the local training phase, the vehicle injects random noise before uploading model parameters, so that the output gradients no longer correspond precisely to the original data, thereby mathematically breaking the attacker's ability to use reverse engineering to infer specific vehicle trajectories or private sample correlations from parameter deviations. In parallel, secret sharing is employed to further fragment, encrypt, and distribute the noise-added parameters, ensuring that no single intermediate node (such as an RSU or management domain) can obtain the complete gradient information, thus fundamentally blocking the attacker's path to reconstruct original data features.

4.2. Defending Against Collusion Attacks

Through the additive secret sharing mechanism, this scheme splits the noise-added model parameters of vehicles into multiple mathematically independent encrypted shards and distributes them to multiple RSUs controlled by different management domains, ensuring that any single set of nodes only holds meaningless data fragments. Such physically distributed storage and logically multi-domain isolation greatly increase the collusion cost for attackers to bribe cross-regional and multi-level nodes simultaneously, so that when attackers fail to gather a sufficient number of shards, they cannot restore the complete model gradients, nor can they conduct any effective privacy sniffing through the correlation between fragments.

4.3. Defending Against Single Point of Failure and Tampering

The architecture leverages the decentralized consensus mechanism of blockchain to replace the single central server found in traditional federated learning, distributing model aggregation and verification logic across multiple independent administrative domains. This ensures that even if specific RSUs or nodes within a certain region go offline due to attacks or failures, other active domains can take over tasks and maintain the continuity of training. Concurrently, all denoised model updates must undergo consensus verification by multi-party nodes before being recorded on the immutable distributed ledger. This fundamentally eliminates the possibility of a single malicious node illegally tampering with global model parameters, thereby ensuring training data consistency and high service availability across the entire IoV system.

5. ANALYSIS OF SIMULATION EXPERIMENT RESULTS

To verify the superiority of the proposed method, the experiment employs the representative real-world dataset MNIST as a benchmark platform and conducts multiple rounds of horizontal comparative tests between FLSPP and the BPFL scheme [9]. Through in-depth statistical analysis of the experimental data, the average accuracy of each scheme after convergence is calculated and derived. The specific performance indicator comparisons and evolution trends are shown in Figure 3.

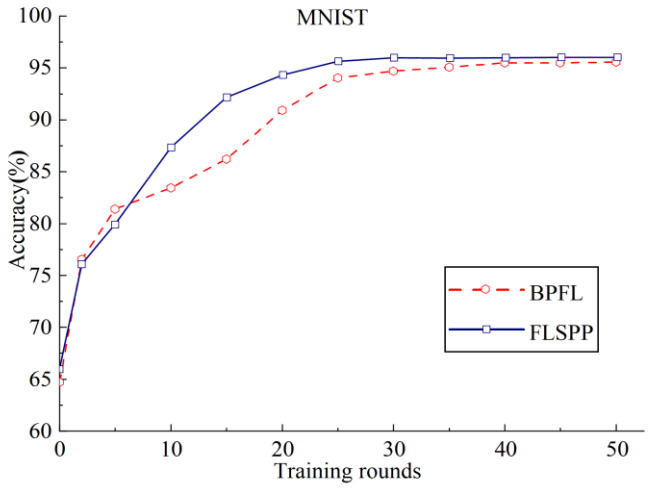

Figure 3. Model accuracy with increasing rounds

Observation of the accuracy curves relative to the number of iterations reveals that the proposed scheme not only achieves superior final precision but also exhibits a more stable convergence profile. This validates the stability of the blockchain-based multi-management domain architecture regarding model updates. By mitigating the stochastic bias introduced by a single aggregation node, the proposed scheme effectively alleviates uncertainty in distributed training, thereby achieving a higher average accuracy during the later stages of the experiment.

To further verify the reliability of the proposed scheme, the real-world datasets MNIST and CIFAR-10 were selected as benchmarking platforms to conduct quantitative comparative experiments between the FLSPP and BPFL schemes regarding average training latency. By recording the time overhead of the entire process from model initialization to convergence and calculating the arithmetic mean through multiple samplings, the experiment aims to evaluate the computational redundancy introduced by different privacy protection mechanisms. The statistical results of the experiment are illustrated in Figure 4.

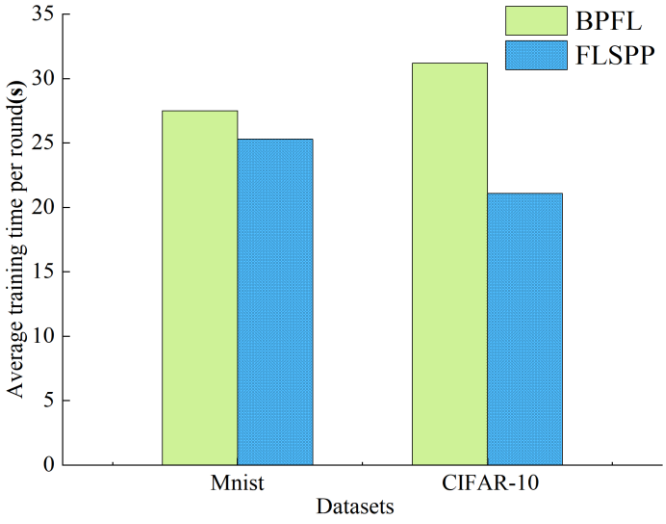

Figure 4. Average training time for different datasets

The experimental results show that FLSPP is significantly superior to the comparison schemes in terms of average training time. This is mainly attributed to the additive secret sharing mechanism adopted in this paper, which replaces the complex large-number encryption process in traditional schemes with efficient operations, thus greatly reducing the computational burden on the vehicle side. This low-complexity design improves the execution efficiency of a single local iteration and ensures efficient operation under the premise of privacy preservation.

6. CONCLUSION

This paper addresses the challenge of balancing data privacy protection with model training efficiency in the IoV environment by proposing a federated learning scheme for privacy preservation in the IoV. By employing a hierarchical aggregation mechanism and incorporating random noise alongside additive secret sharing techniques, this scheme logically achieves fragmented storage and transmission of model gradients, effectively defending against inference attacks and malicious collusion among nodes. Experimental results demonstrate that while ensuring high-strength privacy security, the proposed scheme not only maintains excellent model accuracy but also significantly reduces the average training time of the system, benefiting from the optimization of the decentralized architecture. In summary, this scheme provides a secure, efficient, and highly robust distributed learning paradigm for the complex and dynamic IoV environment.

REFERENCES

[1] Mcmahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data [C]. Artificial Intelligence and Statistics, 2017: 1273-1282.

[2] Zhao K, Hu J, Shao H, et al. Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy [J]. Reliability Engineering \& System Safety, 2023, 236: 109246.

[3] Feng Y, Guo Y, Hou Y, et al. A survey of security threats in federated learning [J]. Complex & Intelligent Systems, 2025, 11(2).

[4] Wang L, Zhao X, Lu Z, et al. Enhancing privacy preservation and trustworthiness for decentralized federated learning [J]. Information Sciences, 2023, 628: 449-468.

[5] Chaudhary R K, Kumar R, Saxena N. A systematic review on federated learning system: a new paradigm to machine learning [J]. Knowledge and Information Systems, 2025, 67(2): 1811-1914.

[6] Cui Y, Zhu J. MChain-SFFL: multi-chain aggregation privacy preserving for server-free federated learning [J]. IEEE Transactions On Network and Service Management, 2024, 21(4): 4861-4870.

[7] Jiang W, Chen M, Tao J. Federated learning with blockchain for privacy-preserving data sharing in internet of vehicles [J]. China Communications, 2023, 20(3): 69-85.

[8] Ning W, Zhu Y, Song C, et al. Blockchain-based federated learning: a survey and new perspectives [J]. Applied Sciences-Basel, 2024, 14(20).

[9] Wang N, Yang W, Wang X, et al. A blockchain based privacy-preserving federated learning scheme for internet of vehicles [J]. Digital Communications and Networks, 2024, 10(1): 126-134.

[10] Lv P, Xie L, Xu J, et al. Misbehavior detection in vehicular ad hoc networks based on privacy-preserving federated learning and blockchain [J]. IEEE Transactions On Network and Service Management, 2022, 19(4): 3936-3948.

[11] Yang Z, Wang R, Wu D, et al. Blockchain-enabled trust management model for the internet of vehicles [J]. IEEE Internet of Things Journal, 2021, 10(14): 12044-12054.

[12] Atlam H F, Ekuri N, Azad M A, et al. Blockchain forensics: a systematic literature review of techniques, applications, challenges, and future directions [J]. Electronics, 2024, 13(17).

[13] Shen M, Wang J, Zhang J, et al. Secure decentralized aggregation to prevent membership privacy leakage in edge-based federated learning [J]. IEEE Transactions On Network Science and Engineering, 2024, 11(3): 3105-3119.