Received 31 March 2023; revised 30 May 2023; accepted 22 June 2023. Date of publication 29 June 2023; date of current version 24 July 2023.

Digital Object Identifier 10.1109/OJCOMS.2023.3290625

Secure and Efficient Hierarchical Decentralized Learning for Internet of Vehicles

ZIXUAN LIANG[1], PENGLIN YANG[2], CHENYU ZHANG[1], AND XINCHEN LYU[1,3]

[1]National Engineering Research Center for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China

[2]Business Research Department, China Mobile Research Institute, Beijing 100031, China

[3]Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen 518055, China

CORRESPONDING AUTHOR: X. LYU (e-mail: lvxinchen@bupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1806602; in part by the National Science Foundation of China under Grant 62001048; and in part by the Beijing University of Posts and Telecommunications–China Mobile Research Institute Joint Innovation Center.

(Zixuan Liang and Penglin Yang contributed equally to this work.)

ABSTRACT Decentralized machine learning enables multiple devices to train a global model collaboratively and is a promising paradigm to realize ubiquitous intelligence for the Internet of Vehicles (IoV). Existing work mainly focused on either the data privacy protection techniques or efficient topology orchestration of decentralized machine learning. However, these techniques cannot be directly applied to IoV due to possible accuracy degradations and insufficient topology adaptability, not to mention the joint secure and efficient decentralized learning designs. This paper proposes a secure and efficient hierarchical decentralized learning framework for IoV networks with multiple fog nodes and mobile vehicles. The proposed framework combines federated learning and distributed consensus for vehicle-fog and inter-fog collaborative learning, respectively, and integrates masking with local training to protect data privacy. We propose the network-level masking mechanism and consensus matrix optimization for signaling-efficient implementations in IoV. The network-level masking can eliminate the masking pairing requirements of inter-fog handover of mobile vehicles and is proved to be canceled via distributed consensus. Experimental results on two popular datasets validate the superiority of the proposed framework in terms of learning accuracy, data protection, and signaling efficiency, compared to the existing approaches.

INDEX TERMS Decentralized machine learning, Internet of Vehicles, data privacy, signaling efficiency.

I. INTRODUCTION

DECENTRALIZED machine learning at the network edge has emerged as a promising paradigm that enables multiple edge devices to train a global Artificial Intelligence (AI) model collaboratively without sharing their raw data [1]. Due to its advantage of distributed model training, while preserving data privacy, decentralized machine learning is regarded as the underlying technology to realize ubiquitous intelligence in Internet-of-Vehicles (IoV) [2]. In particular, the vehicles can share and aggregate a comprehensive and accurate AI model for autonomous driving without a central coordinator [3].

Decentralized machine learning for IoV still faces the challenges arising from the data privacy and adaptability for dynamic geo-distributed topology [4]. Although the raw data is only used for local training, the intermediate results (e.g., gradients) are shared to aggregate a global model. The gradients can already disclose the private raw data, e.g., via DeepLeakage-based attacks [5], [6], [7], [8], hampering the data privacy of the participants. Nevertheless, the IoV network is typically dynamic and geo-distributed with multiple roadside units (RSUs) and fast-moving vehicles. Thus, the decentralized machine learning framework needs to operate without centralized aggregation and accommodate dynamic vehicles.

Overview of Related Work: Data privacy protection [9], [10], [11], [12], [13], [14], [15] and topology orchestration [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27] have been widely studied in the topic decentralized machine learning.

1) Data Privacy Protection. In the topic of data privacy of federated learning, several defense techniques (such as differential privacy [9], [10], [11], homomorphic encryption [12], [13], and masking [14], [15]) have been proposed to protect the original client-specific gradient during global aggregation for improving the security and privacy of federated learning.

2) Topology Orchestration. In terms of topology orchestration, some variants of federated learning (including hierarchical federated learning [16], [17], [18], [19], semi-decentralized learning [20], [21], and fully distributed learning [22], [23], [24], [25], [26], [27]) was proposed for efficient decentralized learning in generally distributed graphs or hierarchical tree topologies.

Currently, many advanced resource management techniques (e.g., user scheduling, task offloading, gradient sparsification, and power control [28], [29], [30]) can be applied to enable communication-efficient federated learning. Resource management is an important issue in the IoV scenario with typically computation-restricted vehicles and increasingly computation-intensive model training. We leave task/data offloading mechanism design [29], [31] in decentralized learning IoV networks for our future work. In this paper, we mainly focus on the data privacy and topology orchestration of decentralized learning for IoV networks.

However, the existing work cannot satisfy the requirements of IoV networks, due to possible model accuracy degradations and lack of joint design of data privacy and topology adaptability. On the one hand, model accuracy is of paramount importance for IoV networks (e.g., autonomous driving applications). The existing defense techniques may degrade the learning accuracy (e.g., differential privacy [9], [10], [11]) or incur excessive computation/signaling overhead (e.g., homomorphic encryption and masking [12], [13], [14], [15]). On the other hand, the existing decentralized frameworks [22], [23], [24], [25], [26], [27] are not dedicated designed for hierarchical IoV networks with typically decentralized RSUs and moving vehicles.

To this end, the key research questions of this paper are *how to design secure and efficient (in terms of computation/signaling overhead) hierarchical learning framework for IoV network without compromising learning accuracy*.

Main Contributions: This paper proposes a secure and efficient hierarchical decentralized learning for IoV networks with multiple fog nodes (e.g., RSU and edge server) and dynamic vehicles. The basic idea is to exploit the hierarchical structure to design a hybrid decentralized learning framework that integrates federated learning (between the vehicles and fog nodes) and distributed consensus (among the fog nodes).

For data privacy, we adopt the masking technique to prevent the individual local gradient being accessed by the adversary (e.g., fog node). The masking technique is computation-efficient but may incur additional signaling overhead for masking seed negotiation. For the signaling/learning efficiency in the dynamic IoV topology, we propose the network-level masking mechanism to reduce the signaling overhead and optimize the consensus matrix between the fog nodes to speed up the consensus process.

Different from the existing approaches that separately considered the data privacy and learning efficiency (topology orchestration), this paper combined the privacy-preserving mechanism (i.e., masking) and communication-efficient designs (i.e., network-level masking pairing and consensus optimization) to achieve secure and efficient hierarchical learning for IoV. The key technical contributions are summarized as follows.

- We design the hierarchical decentralized learning framework that combines vehicle-fog federated learning and inter-fog distributed consensus. Masking is adopted to disturb the local gradients before sending them to the fog node to protect individual gradients from attacks.
- We propose the network-level masking mechanism to prevent frequent masking seed negotiation (e.g., at each time of handover), thereby reducing signaling overhead. We prove that the network-level masks can be canceled via distributed consensus for learning efficiency.
- We prove the convergence guarantee of the proposed hierarchical learning framework for general non-convex loss functions. By reaching consensus gradients at each round, the proposed framework follows the performance of the FedAvg algorithm in federated learning.
- We optimize the consensus matrix to improve training efficiency and reduce the consensus signaling overhead. In particular, the matrix optimization problem is reformulated as semi-definite programming and efficiently solved via convex optimization.

The experiments are conducted on two popular datasets, i.e., MNIST and Fashion-MNIST, under both IID and Non-IID data distributions. Experimental results validate the effectiveness of the proposed framework in terms of data privacy protection, learning accuracy, and signaling efficiency, compared to the state-of-the-art.

Paper Organization: The rest of the paper is organized as follows. Section II provides a brief overview of the existing work. Sections III and IV present the system model and detailed operations of the proposed hierarchical decentralized learning for IoV networks, respectively. Section V demonstrates the dedicated designs for the signaling/learning efficiency for IoV. The experiment results are analyzed in Section VI, followed by the conclusion in Section VII.

II. RELATED WORK

This section briefly reviews the existing work on data privacy and topology orchestration in decentralized machine learning. In the following, we analyze the existing work on these two topics and highlight the difference of our work.

A. DATA PRIVACY FOR FEDERATED LEARNING

Gradient leakage attack is one of the main threat in federated learning. In [5], DeepLeakage was proposed to efficiently reconstruct the private raw data and label from the transmitted gradient in federated learning. The reconstruction was based on setting up virtual pairs of raw data and labels, whose gradient is compared to and optimized to approach the targeted/transmitted actual gradient via stochastic gradient descent. Following the basic idea in [5], some variations were proposed to increase the attack efficiency [6], improve the label leakage accuracy [7], and attack against well-trained neural networks [8].

There are several defense methods against gradient leakage attacks, such as differential privacy [9], [10], [11], homomorphic encryption [12], [13], and masking [14], [15]. In particular, differential privacy adds the Laplace/Gaussian noises into the local gradient to achieve a theoretically provable tradeoff between data protection and learning accuracy [9], [10], [11] Homomorphic encryption enables the secure aggregation based on encrypted (e.g., via additive homomorphic encryption) local gradients to obtain the accurate weighted-averaged global gradient at the aggregation server [12], [13]. However, the accuracy loss in differential privacy is not favorable for autonomous driving services in IoV, and the significant computational overhead of homomorphic encryption would introduce an excessive delay in data training.

The defense technique adopted in this paper is masking from multi-party computations. In particular, multiple parties can achieve collaborative computation of the global gradient without any knowledge of the individual gradients [14]. For example, in [15], the double-masking approach was proposed the users' local gradients and the verifiable mechanism was implemented at the server to prove the correctness of gradient aggregation. Masking would not compromise learning accuracy and is lightweight in computational overhead, but also requires high signaling overhead on sharing and pairing the secret keys among the devices.

B. TOPOLOGY ORCHESTRATION FOR DECENTRALIZED LEARNING

The existing decentralized machine learning frameworks can be categorized into four types in terms of topology orchestration, including traditional federated learning (with a star structure), hierarchical federated learning (in a tree structure) [16], [17], [18], [19], semi-decentralized learning (with a two-layer topology) [20], [21], and fully decentralized learning (in a general graph structure) [22], [23], [24], [25], [26], [27]. Traditional federated learning is the typical case with a centralized parameter server; please see the survey [32] for details. In the following, we summarize the other three types of emerging frameworks.

1) Hierarchical Federated Learning: This is typically referred to as the client-edge-cloud hierarchical federated learning [16], where the clients/devices locally train the local learning model and multiple serves are organized in a hierarchical structure to aggregate the local gradients in the edge-cloud order. The hierarchical federated learning fits the tree structure, where the cloud is the root, each node of the middle layers is the edge servers, and the leaves are the clients. Edge association and resource allocation were optimized in [17], [18], [19] to further improve the effectiveness of the hierarchical learning framework.

2) Semi-decentralized Learning: This framework (also known as hierarchical decentralized learning) [20], [21] is typically studied in the network of collaborative device-to-device (D2D) mobile devices and a base station (acting as the parameter server). In [20], each device locally computed a weighted-average gradient via distributed consensus and sent the gradient to the parameter server for aggregation. In [21], the devices were grouped into different clusters and the devices within local clusters were designed to perform a distributed consensus procedure.

3) Fully Decentralized Learning: There are two popular mechanisms to realize fully decentralized learning, i.e., via blockchain [22], [23], [24] or distributed stochastic gradient descent (DSGD) [25], [26], [27]. The blockchain-enabled decentralized learning exploits the distributed ledger features of blockchain to record the local gradients and perform global aggregation [22]. The recently emerging swarm learning [23] architecture belongs to this category and has also been applied to autonomous vehicles [24]. However, given the limited transaction rate of blockchain, this architecture may not fit the case of massive devices.

DSGD is another underlying technique for enabling fully decentralized learning in a generalized graph, which is proved to exhibit an asymptotic convergence rate even in the case of unreliable communication links [25]. The devices can share their local gradients with neighbors and take weighted averages at each iteration to obtain the global learning model. In [26], [27], consensus optimization and distributed consensus techniques (the foundations of DSGD) were exploited to propose decentralized collaborative learning for massive devices in the absence of a centralized controller.

Difference of This Work: We can see that the existing defense techniques and decentralized learning frameworks cannot satisfy the requirements of IoV networks in this paper.

- The defense techniques may degrade the learning accuracy (e.g., differential privacy) or incur excessive computation/signaling overhead (e.g., homomorphic encryption and masking), especially in highly dynamic IoV networks. However, accuracy and computation/signaling efficiency are critical for autonomous driving applications in IoV.
- The existing decentralized learning frameworks are not dedicated and designed for the hierarchical IoV network with typically decentralized RSUs and moving vehicles. Hierarchical federated learning cannot scale to the network of interest, and fully decentralized learning fails to exploit the possibility of partial centralized control by RSUs (hence suffering performance degradation). The semi-decentralized

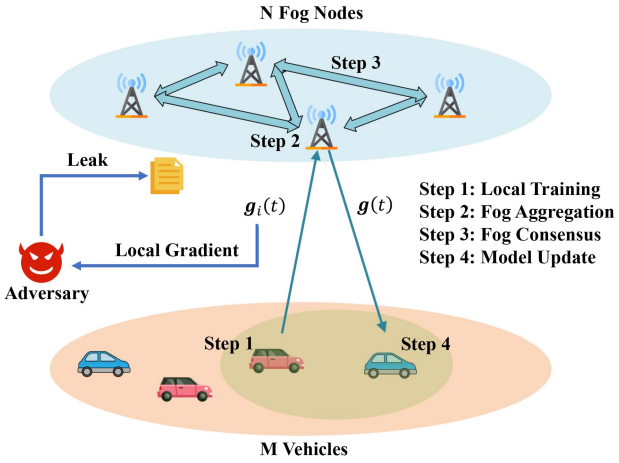

FIGURE 1. The hierarchical decentralized learning framework of *N* fog nodes and *M* vehicles.

learning (i.e., hierarchical decentralized learning) is most similar to this work in terms of topology, but may not scale to the case of multiple RSUs for distributed consensus.

Distinctively different from the existing work that considers security (defense) and efficiency (learning frameworks) separately, this paper focuses on the joint design and proposes secure and efficient hierarchical decentralized learning for IoV networks.

III. SYSTEM MODEL

The hierarchical network consists of N fog nodes (e.g., RSUs and edge servers) and M vehicles, as shown in Fig. 1. The distributed learning system operates in an iterative manner to train a global model for autonomous driving distributively. At each iteration, each vehicle calculates its local gradient by training based on its local sensing data and uploads the gradient to the fog node it belongs to. Then, N fog nodes cooperate to distributively calculate the average gradient and send it back to the vehicles for global convergence. The notations used in this paper are summarized in Table 1.

A. NETWORK MODEL

In the network model, we define the set of fog nodes as $\mathbf{N} = 1, 2, 3, \ldots, N$ and the set of intelligent vehicles as $\mathbf{M} = 1, 2, 3, \ldots, M$. The vehicle uploads the local gradient through the V2X links, and the average gradient is calculated among the fog nodes in a distributed manner through the inter-server link (e.g., via the X2 interface). Here, $\mathbf{E} = \{(i, j)\}$ collects all the inter-server links between the fog nodes.

Given the fast mobility of driving vehicles, the topology of the network, especially the relationship between one intelligent vehicle to its connected fog node, can change dramatically over time. Let $\mathbf{S}_i(t)$ collect the set of vehicles that are in the coverage of fog node i at iteration t. Each vehicle is only connected to one fog node, i.e., $\mathbf{S}_i(t) \cap \mathbf{S}_j(t) = \emptyset$ for $i \neq j$. Let $\mathbf{S}(t) = \{\mathbf{S}_i(t) | i \in \mathbf{N}\}$ collect the coverage relationship of all the vehicles. In summary, the overall hierarchical network topology at iteration t can be denoted by $\mathbf{G}(t) = \{\mathbf{N}, \mathbf{M}, \mathbf{E}, \mathbf{S}(t)\}$.

TABLE 1. Summary of notations.

Notation	Definition
$\mathbf{N}$	Set of fog nodes
$\mathbf{M}$	Set of intelligent vehicles
$\mathbf{E}$	Set of links between fog nodes
$\mathbf{D}$	Global dataset of all the devices
$\mathbf{D}_i$	Local training dataset of vehicle i
$\mathbf{B}_i(t)$	Batch of training data of vehicle i at iteration t
$\mathbf{G}(t)$	Network topology at iteration t
$\mathbf{w}$	Global learning model
$\mathbf{w}_i$	Local learning model of vehicle i
$\mathbf{S}_i(t)$	Set of vehicles served by fog nodes i at iteration t
F	Global loss function
f_i	Local loss function of vehicle i
$g_{(\cdot)}$	Global/local gradient depending on the index
η	Learning rate (i.e., stepsize)
$\mathbf{W}$	Consensus weight matrix
$\mathbf{P}$	Masking pairing matrix
r_i	Random mask generated by vehicle i
s_u^{SK}	Public key hold by vehicle u
s_v^{PK}	Private key hold by vehicle v
$s_{u,v}$	Private shared number between u and v
PRG	Pseudo-random generator
ρ	Spectral radius of a matrix

B. HIERARCHICAL DECENTRALIZED LEARNING MODEL

This section illustrates the basics and preliminaries of the distributed learning model in the hierarchical network. The details of the proposed hierarchical decentralized learning framework will be provided in Section IV. Each vehicle i has its local training dataset $\mathbf{D}_i$, and the global dataset of all the devices can be given by $\mathbf{D} = \cup_{i \in \mathbf{M}} \mathbf{D}_i$. Let $\mathbf{w}(t)$ and $\mathbf{w}_i(t)$ denote the global model and local model of vehicle i at iteration t, respectively.

The objective of the distributed learning system is to find a global model $\mathbf{w}(t)$ that minimizes the global loss function $F(\mathbf{w}(t))$ in terms of the global dataset $\mathbf{D}$. Each vehicle i also maintains a local model $\mathbf{w}_i(t)$ and corresponding local loss function $f_i(\mathbf{w}_i(t))$. The relationship between the local and global loss functions can be given by [33]

$$F(\mathbf{w}(t)) = \frac{\sum_{i=1 \in \mathbf{M}} |\mathbf{D}_i| f_i(\mathbf{w}(t))}{|\mathbf{D}|}, \tag{1}$$

where $|\cdot|$ is the size (number of elements) of a set $(\cdot)$.

Recall that the hierarchical network $\mathbf{G}(t)$ consists of two layers for the vehicles and fog nodes. The vehicles $i \in \mathbf{S}_j(t)$ in the coverage of the same fog node j can perform the steps of federated learning. However, the fog nodes are organized in a distributed manner and need to conduct additional distributed consensus procedures for global convergence. Without diving into the details of distributed consensus, the steps of the hierarchical learning at iteration t are as follows.

Step 1 (Local Training): Consider the mini-batch gradient descent training mechanism [34]. Each vehicle i shuffles its local dataset and randomly select a subset (i.e., batch) of data, denoted by $\mathbf{B}_i(t)$, from $\mathbf{D}_i$ to be trained. The vehicle calculates the batch gradient $g_i(t) = \nabla_{\mathbf{B}_i(t)} f_i(\mathbf{w}_i(t))$ and sends the local gradient to its corresponding fog node.

Step 2 (Fog Aggregation): Each fog node j can collect the received gradient $\mathbf{g}_i(t)$ from the vehicles $i \in \mathbf{S}_j(t)$ in its coverage. Different from the gradient aggregation in federated learning, to maintain the information on sizes of vehicles (i.e., data volume), the fog-level gradient $g_j(t)$ can be calculated by summarizing the local gradients, as given by $g_j(t) = \sum_{i \in \mathbf{S}_j(t)} g_i(t)$.

Step 3 (Fog Consensus): Each fog node j shares the fog-level gradient $g_j(t)$ with its neighboring node m with $(j, m) \in \mathbf{E}$. The details of the consensus steps will be illustrated in Section III-B. The final result of the fog consensus is the global gradient $g(t)$, satisfying

$$g(t) = \frac{1}{M} \sum_{j \in \mathbf{N}} g_j(t). \tag{2}$$

Without loss of generality, the denominator takes the total number of vehicles M, since the batch sizes $|\mathbf{B}_i(t)|$ are typically the same (e.g., 32/64/128 data samples). It can be easily extended to the cases of different batch sizes or even heterogeneous local training epochs, e.g., by letting the denominator be $\sum_{i \in \mathbf{M}} |\mathbf{B}_i(t)|$.

Step 4 (Model Update): The fog node multicasts the global gradient $g(t)$ to all the vehicles. Upon receiving the global gradient, each vehicle i updates its local model according to [35]

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) - \eta g(t), \tag{3}$$

where η is the adjustable learning rate that relates to the convergence speed and model accuracy.

C. THREAT MODEL

We consider the typical curious-but-honest threat model for the fog nodes. In other words, the adversary (e.g., the fog nodes) would follow the standard operations of the hierarchical decentralized learning procedure, but attempt to recover the original private data from the received gradients of the vehicles. This is the popular gradient-based passive attacks in federated learning [36].

Note that the participants can also be malicious, i.e., may deviate from the standard operations of the learning procedure and conduct the adversary attacks (e.g., poisoning attacks) to manipulate the victim's training model. This is the case of active attacks, which can be detected by the victims (e.g., by comparing the model gradients of the participants where the adversary is more likely to exhibit different gradient directions). In this paper, we focus on the passive attack setting [36] (following the standard procedure), which can hardly be perceived and are of practice importance.

For example, DeepLeakage [5] can reconstruct the original input data and private label, denoted by x and y, respectively,

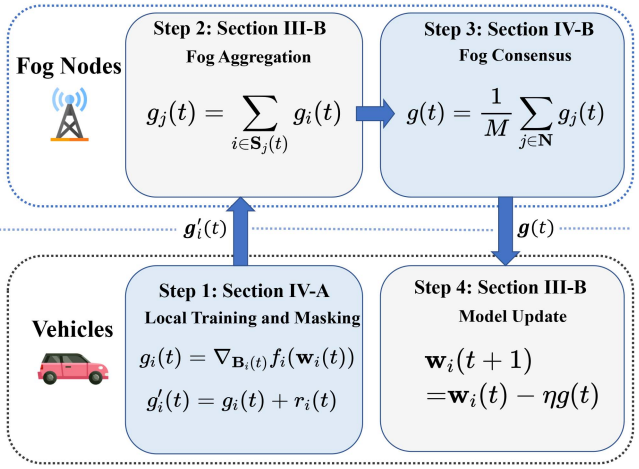

FIGURE 2. The operational steps of the proposed hierarchical decentralized learning framework.

from the received gradient g. In particular, the adversary can generate random dummy input and label pair, namely (x', y'), and tries to optimize the distance between the gradient of the virtual sample pair (x', y') (denoted by $g(x', y')$) and the targeted/received gradient g, i.e.,

$$(x'^*, y'^*) = \arg\min_{(x', y')} ||g(x', y') - g||^2. \tag{4}$$

The above optimization problem can be solved by the gradient descent mechanism iteratively. When the optimization finishes, the optimal dummy pair (x'^*, y'^*) is mostly similar to the targeted gradient and can reveal the original private data and label, thereby degrading the security performance of the learning process.

IV. PROPOSED MASKING-ENABLED HIERARCHICAL DECENTRALIZED LEARNING FRAMEWORK

This section presents the proposed masking-enabled hierarchical decentralized federated learning framework. Fig. 2 shows the operations of the proposed framework that also consists of four basic steps of hierarchical decentralized learning as stated in Section III-B. In particular, the proposed framework adds the masking procedure in the first step of local training to protect the data privacy from the adversary (in Section IV-A), and also details the fog consensus steps to obtain the global gradient distributively (in Section IV-B). Finally, we summarize the remaining challenges of signaling and communication efficiency of the system (in Section IV-C) and propose to optimize the consensus and masking signaling process (as will be stated in Section V).

A. LOCAL TRAINING WITH MASKING PROTECTION

In the following, we illustrate the details of the first step on how to add masking protection in the local training process. Recall that the malicious curious-but-honest participants can reconstruct the private information from the received/targeted gradient. The basic idea of the masking

protection is to add a random mask to the original local gradient such that the adversary cannot retrieve the original gradient for data attacks.

1) Operation of Masking Protection: Let $r_i(t)$ denote the random mask generated by vehicle i at iteration t. The masked gradient $g_i'(t)$ and the generated masks should satisfy

$$g_i'(t) = g_i(t) + r_i(t),$$
$$\sum_{i \in \mathbf{S}_j(t)} r_i(t) = 0, \ \forall j \in \mathbf{N}, \tag{5}$$

where the second equation ensures that all the random masks of the vehicles served by the same fog node can be canceled. As a result, the fog node can also obtain the correct fog-level aggregated gradient from the masked gradient $g_i'(t)$, i.e.,

$$g_j(t) = \sum_{i \in \mathbf{S}_j(t)} g_i'(t) = \sum_{i \in \mathbf{S}_j(t)} g_i(t) + \sum_{i \in \mathbf{S}_j(t)} r_i(t) = \sum_{i \in \mathbf{S}_j(t)} g_i(t). \tag{6}$$

2) Generation of Masks with Key Negotiation: Masking generation is a well-established procedure based on cryptography and key negotiation [14], [15]. In the following, we demonstrate the details of generating the masks $r_i(t)$ satisfying Eq. (5) via the key negotiation among the vehicles. We start by introducing the basics of the key agreement process. A typical key agreement protocol consists of two key algorithms, including KA.gen(pp) and KA.agree(s_u^{SK}, s_v^{PK}) [37]:

1) KA.gen(pp) $\rightarrow (s_u^{SK}, s_v^{PK})$ allows any vehicle u to generate a private-public key pair, where pp is a random number to input.
2) KA.agree(s_u^{SK}, s_v^{PK}) $\rightarrow s_{u,v}$ allows any user u to combine its private key with the public key s_v^{PK} to obtain a private shared number $s_{u,v}$ between u and v.

Based on the key agreement protocol, we can summarize the process of the key initialization and masking generation in Fig. 3. The basic idea is to first initialize the secret keys (i.e., random seeds) at the vehicles and generate the pseudo-randoms at each iteration. The detailed steps of masking seed generation are as follows.

1) *Key Initialization:* Vehicles generate a key pair (s_u^{SK}, s_v^{PK}) through key generation algorithm KA.gen(pp). The private key s_u^{SK} will be kept confidentially by the vehicle, and the public key s_u^{PK} will be sent to the fog node.
2) *Public Key Dissemination:* Each fog node j, acting as the coordinating server, collects all messages from the vehicles and broadcasts the vehicle identification information and corresponding public key (v, s_v^{PK}) to all vehicles $i \in \mathbf{S}_j(t)$ in its coverage.
3) *Seed Generation:* When a vehicle receives other vehicles' public key s_v^{PK}, it can generate a random mask $s_{u,v}$ via KA.agree(s_u^{SK}, s_v^{PK}) $\rightarrow s_{u,v}$.

After finishing the above process, each vehicle u can securely disseminate the random seed $s_{u,v}$ and also receive the random seed $s_{v,u}$ to/from all the vehicles v in the coverage of the same fog node. Given the agreement of the random vector $s_{u,v}$, we can generate the mask of vehicle u served by fog node j as follows:

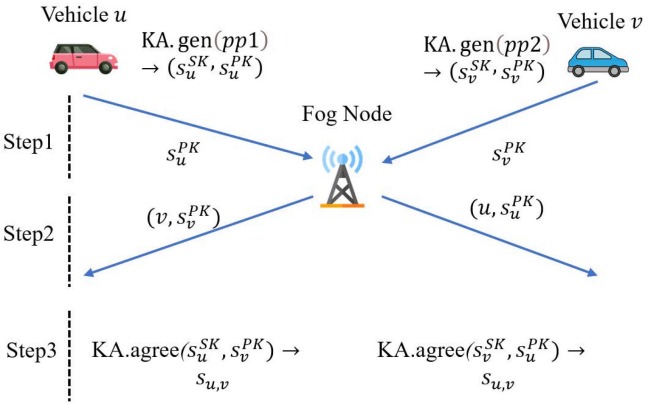

FIGURE 3. The process of key initialization and masking generation.

$$r_u(t) = \sum_{v \in \mathbf{S}_j(t): u<v} \mathrm{PRG}(s_{u,v}) - \sum_{v \in \mathbf{S}_j(t): u>v} \mathrm{PRG}(s_{v,u}), \tag{7}$$

where PRG(s) is a pseudo-random generator based on random seed s [38]. In particular, a pseudo-random generator is a deterministic algorithm that generates a sequence of numbers that appear to be random. With the same sequence index (i.e., iteration t) across the vehicles, the random number PRG($s_{u,v}$) are identical at different vehicles.

We can show that the generated masks $r_u(t)$ in Eq. (7) satisfy the requirement in Eq. (6) to ensure the correctness of fog-level aggregated gradient. In particular, for each fog node j, we have

$$\begin{aligned}\sum_{u \in \mathbf{S}_j(t)} r_u(t) &= \sum_{u \in \mathbf{S}_j(t)} \sum_{v \in \mathbf{S}_j(t): u<v} \mathrm{PRG}(s_{u,v}) \\ &\quad - \sum_{u \in \mathbf{S}_j(t)} \sum_{v \in \mathbf{S}_j(t): u>v} \mathrm{PRG}(s_{v,u}) \\ &= \sum_{u \in \mathbf{S}_j(t)} \sum_{v \in \mathbf{S}_j(t): u<v} \left[\mathrm{PRG}(s_{u,v}) - \mathrm{PRG}(s_{u,v})\right] = 0,\end{aligned} \tag{8}$$

where the second equality can be achieved by rearranging the terms with PRG($s_{v,u}$).

We can see that, by applying the masking protection approach in local training, the adversary (e.g., fog node) cannot have the original local gradient of individual vehicles, hence protecting data privacy. Nevertheless, the fog node can still retrieve the accurate fog-level gradient $g_j(t)$ according to Eq. (6), since the random masks can be canceled with each other as shown in Eq. (8). The masking-enabled protection would not compromise the accuracy of the learning process.

B. DISTRIBUTED FOG CONSENSUS

In the following, we detail the process of distributed fog consensus in Step 3 to obtain the averaged global gradient $g(t)$ without a centralized aggregator according to the

distributed consensus technique [39]. In particular, each fog node would iteratively exchange its fog-level gradient until reaching a consensus. Let $g_i(\tau)$ be the consensus gradient of fog node i at iteration τ, which is initialized as the fog-level gradient at the iteration, i.e., $g_i(0) = g_i(t)$. Each fog node i communicates with its neighboring node j to update the consensus gradient, i.e.,

$$g_i(\tau+1) = W_{ii}g_i(\tau) + \sum_{j\in\mathbf{N},(i,j)\in\mathbf{E}} W_{ij}g_j(\tau),\ \forall i \in \mathbf{N}, \quad (9)$$

where $\mathbf{W} = \{W_{ij}|(i,j) \in \mathbf{E}\}$ is the consensus weight matrix for reaching the averaged global gradient. The consensus matrix $\mathbf{W}$ should satisfy

$$\mathbf{W} = \mathbf{W}^{\mathrm{T}}, \mathbf{W}\mathbf{1} = \mathbf{1}, \quad (10)$$

$$\rho\left(\mathbf{W} - \mathbf{1}\mathbf{1}^T/N\right) < 1, \quad (11)$$

$$\mathbf{W} \in \mathcal{W}. \quad (12)$$

Here, Eq. (10) indicates that W is symmetric and satisfies the doubly stochastic condition. Eq. (11) (together with Eq. (10)) shows that $\mathbf{1}$ is one of the eigenvalues of $\mathbf{W}$ and the other eigenvalues are less than 1 in magnitude. Eq. (12) guarantees that the matrix cannot violate the topology in $\mathbf{E}$.

According to [39], following the above consensus update process, the consensus gradient at each fog node would converge to the average value of the initialized fog-level gradients, i.e., $g_i(\tau) \leftarrow \frac{1}{N}\sum_{j\in\mathbf{N}} g_j(t)$, in a finite number of update rounds. Then, the fog node can adjust the global gradient according to the training data volume, e.g., letting $g(t) = \frac{N}{M}g_i(\tau)$ according to Eq. (2), and multicast $g(t)$ to all the vehicles in its coverage for model update.

C. IMPLEMENTATION DESIGN AND SIGNALING CHALLENGES FOR INTERNET OF VEHICLES

The implementation of the proposed framework can be summarized in Fig. 2. In particular, the vehicles are responsible for *Step 1* of local training with masking protection and *Step 4* of model update, and the fog nodes conduct *Steps 2–3* for fog aggregation and distributed consensus. The operations of *Steps 1 and 3* are presented in Sections IV-A and IV-B and the operations of *Steps 2 and 4* remain the same as in the learning model in Section III-B.

The proposed framework incurs lightweight additional computations of generating pseudo randoms in Eq. (7), and can achieve secure hierarchical decentralized learning in the network of M vehicles and N fog nodes without a centralized aggregation server. However, the proposed framework for IoV may still face the challenges of significant signaling overhead due to the high mobility features of vehicles and iterative distributed consensus among fog nodes. We summarize the challenges as follows.

1) Signaling Overhead due to Frequent Masking Pairing: Recall that the generated mask $r_u(t)$ needs to be fully eliminated at its associated fog node j, as stated in Eq. (8). This requires that all the vehicles in the coverage of the fog node can have the mask seed $s_{u,v}$ via the key negotiation protocol. The key negotiation process requires considerable signaling overhead In the deterministic network, $s_{u,v}$ can be initialized at once to produce the masks via PRG at each iteration to reduce such signaling overhead.

However, in the dynamic IoV network, the set of vehicles served by the same fog node, i.e., $\mathbf{S}_j(t)$, frequently changes over time. As a result, at each time of fog-level handover, the vehicles entering the coverage of a new fog node need to perform a new key negotiation process to obtain the mask seed pairs. Given the high mobility of vehicles, the handover would occur increasingly frequently, resulting in excessive signaling overhead for masking generation.

2) Signaling Overhead due to Distributed Fog Consensus: In the distributed fog consensus step, the fog nodes need to share the consensus gradient with their neighbors for multiple iterations for reaching the convergence. These iterations are necessary for reaching a global consensus to obtain the global gradient, but would also incur considerable signaling overhead for transmitting the gradients among fog nodes. It is critical to reduce the fog-level signaling overhead to further increase the efficiency of the proposed framework.

The existing techniques, such as periodical synchronization, gradient quantization, gradient pruning and sparsification, can be applied to reduce such signaling overhead. Nevertheless, the number of iterations in the distributed consensus process depends on the feature of the consensus matrix $\mathbf{W}$. We need to properly design the matrix to speed up the consensus rate and reduce the inter-fog overhead.

V. PROPOSED SIGNALING-EFFICIENT DESIGNS FOR INTERNET OF VEHICLES

This section demonstrates the proposed signaling-efficient designs to reduce the excessive signaling due to frequent masking pairing and distributed fog consensus in IoV networks. In particular, to reduce the masking signaling overhead at the vehicles, we propose the network-level masking pairing (instead of the fog-level) mechanism and prove the network-wide masking cancellation (to be shown in Section V-A). To reduce the fog-level signaling, we propose to optimize the consensus matrix to speed up the consensus rate (to be shown in Section V-B).

A. NETWORK-LEVEL MASKING PAIRING AND CANCELLATION FOR VEHICLE-LEVEL SIGNALING EFFICIENCY

To address the challenge of signaling overhead due to frequent masking pairing (i.e., vehicle handover), we propose to design the network-level masking pairing (instead of the fog level), i.e., masking pairing across the whole network. Then, the masking pairing happens only when the vehicle runs out of the network (e.g., the city-wide area), thereby significantly reducing the pairing frequency (compared to the handover in fog-level masking pairing). In the following, we introduce the process of network-level masking pairing and prove the added masks can be canceled during the fog-distributed consensus step.

1) Network-Level Masking Pairing: Let $\mathbf{P} = \{p_{ij} | \forall i, j \in \mathbf{M}\}$ be the masking paring matrix among M vehicles. In particular, $p_{ij} \in \{0, 1\}$ denotes the masking pairing conditions, where $p_{ij} = 1$ indicates that vehicles i and j have mutually mask seeds $s_{i,j}$ and $s_{j,i}$; and otherwise, not. $\mathbf{P}$ is symmetric, i.e., $p_{ij} = p_{ji}$.

The paring matrix $\mathbf{P}$ can be sparse (at least ensuring that each vehicle has its masking pairs). Let $d(\mathbf{P})$ denote the minimum degree of all the vehicles in the matrix $\mathbf{P}$. The masking protection performance would increase with $d(\mathbf{P})$ in case all the other paired vehicles collaborate with the adversary to retrieve the original gradient. We do not specify the detailed process of generating $\mathbf{P}$, and only require that $d(\mathbf{P}) \geq 2$ to ensure the effectiveness of masking protection. For example, the pairing matrix $\mathbf{P}$ can be induced based on the relationship of the vehicles (e.g., the social relations of its owners), such that the paired vehicle would not expose the random seed.

Based on the paring matrix $\mathbf{P}$, the mask of vehicle i at iteration t, i.e., $r_i(t)$, can be given by

$$r_i(t) = \sum_{j \in \mathbf{P}_i : i < j} \mathrm{PRG}\big(s_{i,j}\big) - \sum_{j \in \mathbf{P}_i : i > j} \mathrm{PRG}\big(s_{j,i}\big), \tag{13}$$

where $\mathbf{P}_i = \{j | p_{ij} = 1\}$ is the set of vehicles that are paired with vehicle i.

The other steps remain the same except that Eq. (7) is replaced with Eq. (13) to achieve the network-level masking. In the network-level masking case, the masks may not be canceled within each fog node, i.e., $\sum_{i \in \mathbf{S}_j(t)} r_i(t) \neq 0$, since the paired vehicles are not necessarily in the coverage of the same fog node. In other words, the fog-level aggregated gradient may be incorrect, i.e.,

$$g_j(t) = \sum_{i \in \mathbf{S}_j(t)} g_i(t) + \sum_{i \in \mathbf{S}_j(t)} r_i(t) \neq \sum_{i \in \mathbf{S}_j(t)} g_i(t). \tag{14}$$

2) Proof of Network-Level Masking Cancellation: In the following, we prove that the added masks can be successfully canceled during the distributed consensus process.

Theorem 1: In the case that the consensus matrix $\mathbf{W}$ satisfies Eqs. (10)–(12), the added masks $r_i(t)$ can be canceled during the distributed consensus, i.e.,

$$g(t) = \frac{1}{M} \sum_{i \in \mathbf{M}} g_i(t). \tag{15}$$

In other words, the global gradient can converge to the average of the local gradients of all the vehicles via the distributed consensus.

Proof: 1) Proof of Average Consensus: We first prove that the average consensus iterations in Eq. (9) would converge to the average initial value, i.e.,

$$\lim_{T \to \infty} g_i(T) = \lim_{T \to \infty} \mathbf{W}^{\mathrm{T}} g_i(t) = (1/N) \mathbf{1}\mathbf{1}^{\mathrm{T}} g_i(t). \tag{16}$$

Based on Eqs. (10)–(12), we have

$$\begin{aligned} \mathbf{W}^{\mathrm{T}} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N &= \mathbf{W}^{\mathrm{T}}\big(I - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N\big) \\ &= \mathbf{W}^{\mathrm{T}}\big(I - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N\big)^{\mathrm{T}} \\ &= \big(\mathbf{W}\big(I - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N\big)\big)^{\mathrm{T}} \\ &= \big(\mathbf{W} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N\big)^{\mathrm{T}}, \end{aligned} \tag{17}$$

where the first equality is due to Eq. (10), and the second equality is because $I - \mathbf{1}\mathbf{1}^{T}/N$ is a projection matrix. By further exploiting Eq. (11) (which indicates the spectral radius of $\mathbf{W}$ is less than 1), we can obtain

$$\lim_{T \to \infty} \mathbf{W}^{\mathrm{T}} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N = 0. \tag{18}$$

This concludes the proof of Eq. (16). In other words, by taking the denominator of $\frac{N}{M}$ after reaching the distributed consensus, the global gradient $g(t)$ satisfies Eq. (2).

2) Proof of Masking Cancellation: We proceed to prove that the added masks $r_i(t)$ can be canceled in the global gradient. According to Eq. (2), we have

$$\begin{aligned} g(t) &= \frac{1}{M} \sum_{j \in \mathbf{N}} g_j(t) \\ &= \frac{1}{M} \sum_{j \in \mathbf{N}} \left\{ \sum_{i \in \mathbf{S}_j(t)} g_i(t) + \sum_{i \in \mathbf{S}_j(t)} r_i(t) \right\} \\ &= \frac{1}{M} \sum_{i \in \mathbf{M}} \{g_i(t) + r_i(t)\} = \frac{1}{M} \sum_{i \in \mathbf{M}} g_i(t), \end{aligned} \tag{19}$$

where the last equality is due to

$$\begin{aligned} \sum_{i \in \mathbf{M}} r_i(t) &= \sum_{i \in \mathbf{M}} \sum_{j \in \mathbf{P}_i : i < j} \mathrm{PRG}\big(s_{i,j}\big) \\ &\quad - \sum_{i \in \mathbf{M}} \sum_{j \in \mathbf{P}_i \ : \ i > j} \mathrm{PRG}\big(s_{j,i}\big) \\ &= \sum_{i \in \mathbf{M}} \sum_{j \in \mathbf{P}_i \ : \ i < j} \big[\mathrm{PRG}\big(s_{i,j}\big) - \mathrm{PRG}\big(s_{i,j}\big)\big] = 0, \end{aligned} \tag{20}$$

This concludes the proof. ■

3) Learning Performance Analysis: We proceed to prove the convergence guarantee the proposed hierarchical decentralized learning framework. To facilitate the proof, we consider the following typical assumptions for general non-convex loss functions.

Assumption 1: There exist constants $G \geq 0$ and $B \geq 1$, such that

$$\frac{1}{M} \sum\nolimits_{i=1}^{M} ||\nabla f_i(\mathbf{x})||^2 \leq G^2 + B^2 \|\nabla f(\mathbf{x})\|^2, \forall \mathbf{x},$$

Assumption 2: The gradient $g_i = \nabla_{\mathbf{B}_i} f_i(\mathbf{x})$ based on the data samples in batch B_i is an unbiased estimation with bounded variance, i.e.,

$$\mathbb{E}_{\mathbf{B}_i}\Big[||g_i - \nabla f_i(\mathbf{x})||^2\Big] \leq \sigma^2, \text{ for any } i, \mathbf{x}.$$

Assumption 3: The loss $\{f_i\}$ is β-smooth, satisfying

$$||\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{y})|| \leq \beta \|\mathbf{x} - \mathbf{y}\|, \forall i, \mathbf{x}, \mathbf{y}.$$

The assumptions above are widely adopted in the literature for the general non-convex loss functions. By resembling the training process of the proposed framework to the standard FedAvg algorithm in federated learning, we can derive the

theoretical convergence guarantee of the proposed framework in the following theorem.

Theorem 2: Suppose that assumptions 1–3 hold for general non-convex loss f_i. There exist weights $\{\mathbf{w}(T)\}$, and for any step-size $\eta \leq \frac{1}{(1+B^2)8\beta K}$, the output of proposed framework $\mathbf{w}(T)$ satisfies

$$\mathbb{E}\Big[||\nabla f(\mathbf{w}(T))||^2\Big] \leq \mathcal{O}\left(\frac{\beta\sigma\sqrt{MF}}{\sqrt{TKM}} + \frac{F^{2/3}\big(\beta G^2\big)^{1/3}}{(T+1)^{2/3}} + \frac{B^2\beta F}{T}\right)$$

where T and K are the total rounds of training iterations and local batch updates, respectively. For brevity, $F = f(\mathbf{w}(0)) - f^\star$.

Proof: According to Theorem 1, the network-level masks can be canceled and the global gradient can converge to the average of the local gradients via the distributed consensus process, i.e., satisfying Eq. (15). With the same batch sizes, this is the typical setting of the popular FedAvg algorithm in federated learning. As a result, the convergence guarantee analysis of the proposed framework follows the convergence guarantee of FedAvg. Please refer to [40, Th. 5] for the details. ■

B. CONSENSUS MATRIX OPTIMIZATION FOR FOG-LEVEL SIGNALING EFFICIENCY

To reduce the signaling overhead due to distributed fog consensus, we propose to optimize the consensus matrix $\mathbf{W}$ to speed up the consensus rate. We note that the matrix optimization aims to reduce the number of iterations required to reach the fog-level consensus, and can be readily applied in conjunction with the existing communication-efficient techniques (including periodical synchronization, gradient quantization, gradient pruning and sparsification) to further reduce the signaling overhead.

In the following, we optimize the consensus matrix $\mathbf{W}$. The consensus rate (the number of iterations before consensus) is related to the spectral radius of $\mathbf{W}$ [39], i.e., $\rho(\mathbf{W} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N)$. Then, the optimization problem of minimizing the spectral radius can be formulated as

$$\begin{aligned} \min_{\mathbf{W}} \quad & \rho\big(\mathbf{W} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N\big) \\ \text{s.t.} \quad & (10),\ (11),\ (12). \end{aligned} \tag{21}$$

Problem (21) can be transformed into a semi-definite programming (SDP) problem to improve the solving efficiency. Note that the consensus matrix $\mathbf{W}$ is symmetric. Then, we can introduce a scalar variable s as the upper bound of the spectral radius of $\mathbf{W}$ and reformulate the problem as

$$\begin{aligned} \min_{\mathbf{W}} \quad & s \\ \text{s.t.} \quad & \begin{bmatrix} sI & \mathbf{W} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N \\ \mathbf{W} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N & sI \end{bmatrix} \succcurlyeq 0, \\ & (10),\ (11),\ (12), \end{aligned} \tag{22}$$

Algorithm 1 Proposed Secure and Efficient Hierarchical Decentralized Learning Framework

Step 1: Local Training and Masking
1: *Network-level Mask Pairing:* The vehicles follow the network-level masking pairing matrix $\mathbf{P}$ to generate the masks according to Eq. (13).
2: *Local Training with Masks:* Each vehicle conducts the local training process, and generates the masked gradients based on Eq. (5).

Step 2: Fog Aggregation
3: The fog nodes aggregate the masked gradients of its connected vehicles according to Eq. (14).

Step 3: Fog Consensus
4: *Consensus Matrix Optimization:* Optimize the consensus matrix $\mathbf{W}$ by solving the SDP problem (22).
5: *Consensus Update:* Each fog node updates the global gradients according to Eq. (9) iteratively.

Step 4: Model Update
6: After receiving the multicast global gradients, the vehicles update the local model based on Eq. (3).

where the first matrix inequality is the linear matrix inequality to bound the spectral radius of $\mathbf{W}$. Problem (22) is SDP and can be efficiently solved by a convex optimization solver. In this paper, we adopt the CVXPY library [41] (a Python-embedded modeling language for convex optimization) to solve the SDP problem and obtain the optimized consensus matrix to speed up distributed consensus (and hence, reduce the inter-fog communication overhead).

Algorithm 1 summarizes the detailed operational process of the proposed secure and efficient hierarchical decentralized framework. The framework still follows the standard steps in Fig. 2 (as specified in Section IV). However, in lines 1 and 4, we propose the network-level masking pairing (instead of the fog-level) mechanism and optimize the consensus matrix to speed up the consensus rate. The excessive signaling due to frequent masking pairing and distributed fog consensus can be significantly reduced, as will be shown in Figs. 6 and 7.

VI. EXPERIMENTAL RESULTS

This section evaluates the effectiveness of the proposed hierarchical decentralized learning framework. In the following, we will first introduce the experimental settings and then analyze the experimental results in terms of defense effectiveness, model accuracy and communication overhead.

A. EXPERIMENTAL SETTING

We adopt the Pytorch project to implement the proposed hierarchical decentralized learning framework in a simulated network of 5 fog nodes and 20 vehicles. For simulating the dynamic topology, we adopt the SUMO (Simulation of Urban MObility) platform [42] to generate the positions and mobility features of the vehicles in the four-way four-lane crossroads with the size of 100 × 100m. SUMO is an open-source, highly portable, microscopic traffic simulation package, used in different projects to simulate automatic

driving or traffic management strategies. The fog nodes are uniformly located in the simulated network, and the vehicles are served by the fog node with the minimum distance. The simulation duration is 20 minutes.

1) Datasets and Models: We conduct the experiments on the MNIST and Fashion-MNIST datasets for handwritten digits and basic images, respectively. The datasets include 60,000 samples and 10,000 testing samples. The convolutional-neural-network (CNN)-based LeNET learning model [43] is adopted for the classifications of both the MNIST and Fashion-MNIST datasets. The learning rate and batch size are 0.001 and 64, respectively, in the model training experiments.

2) Data Distribution: Both IID and Non-IID data distributions are considered in our experiments. In the IID setting, the data samples are uniformly distributed among the vehicles. In the non-IID setting, we consider the extremely diverse data distributions, where the data samples are sorted by different types according to their labels and assigned to the vehicles. In particular, 60,000 training samples are divided sequentially to 40 pieces (each of 1,500 samples), and each vehicle is randomly assigned with two pieces of data. The maximum label number of vehicles' training data is two, i.e., the resultant data distribution is more diverse than the typical Dirichlet-based Non-IID data.

3) Attack Method: The adversary (e.g., the curious-but-honest fog node) has access to the transmitted local gradient and conducts DeepLeakage [5] to reconstruct the private original data of the vehicles. Some basic designs of DeepLeakage can be found in Section III-C.

4) Comparison Benchmark: For comparison purposes, we also simulate the following benchmark approaches to evaluate the performances of accuracy, defense and communication overhead, as follows.

- *Centralized-Aggregated Federated Learning (FL):* This is the case of basic federated learning where the vehicles upload the local gradient to one centralized server for global synchronization. Here, we assume the availability of a central server in the network and this serves for the evaluation of the learning performance (and provides an upper bound) of the proposed framework.
- *Federated Learning with Differential Privacy (DP-FL):* This is to implement the differential privacy to federated learning for data privacy [9]. The artificial noise is added to the local gradients at the vehicles (instead of adding masks) and the other operations remain the same as the proposed framework.
- *Basic Hierarchical Decentralized Learning (Basic-Hierarchical):* This is the basic design of the proposed framework with fog-level masking and no consensus matrix optimization, as stated in Section IV. The benchmark hierarchical-FL approach serves for evaluating the effectiveness of the dedicated designs in Section V in terms of communication overhead.

The proposed framework is denoted by "Proposed" in the following experimental results. We also note that homomorphic encryption can protect data privacy but would incur 10^4 times of additional computation times (typically about 100s per-round additional training time for each client) than the random generators. Thus, homomorphic encryption is not simulated due to its excessive runtime overhead.

B. RESULT ANALYSIS

In the following, we evaluate the effectiveness of the proposed approach in terms of accuracy performance, data protection performance, and signaling overhead.

1) Accuracy Performance: Fig. 4 plots the accuracy performances of centralized FL, the proposed framework, and DP-FL on the MNIST and Fashion-MNIST datasets under both the IID and non-IID data distributions. Here, we consider two different levels of artificial noises in differential privacy, where the noise control parameter ϵ is set to 1.5 (mild noise) and 0.5 (moderate noise).

We can see in Fig. 4-(a) that, in both the IID and non-IID distributions of the MNIST dataset, the proposed framework can achieve the same accuracy performance (up to 98% accuracy) with the centralized FL benchmark (which assumed a centralized server to aggregate all the gradient of the vehicles and served as the upper bound of accuracy performance). This validates the accuracy performance of the proposed framework. In contrast, the DP-FL would suffer from a lower convergence rate and accuracy losses due to the added artificial noises at the local gradients of the vehicles. The performance degradation also increases with the noise levels of differential privacy.

The results show a similar phenomenon in the Fashion-MNIST dataset in Fig. 4-(b). The differences are the lower accuracy of the proposed and FL approaches (due to the increased complexity of Fashion-MNIST) and the increased performance degradation (up to 15% and 5% accuracy losses during the training and after convergence, respectively). Such performance degradation is already unacceptable for accuracy-critical services. By comparing Figs. 4-(a) and (b), we can see that the performance degradation would become increasingly significant with the complexity of datasets. As a result, the accuracy loss would prevent differential privacy from applications in IoV (especially, autonomous driving).

2) Data Protection Performance: Fig. 5 shows the data attack (via DeepLeakage [5]) results on MNIST and Fashion-MNIST datasets for the centralized FL, the proposed framework, and DP-FL ($\epsilon = \{0.5, 1.5\}$), as the attack iterations increase to 100. We can see that, in both the MNIST and Fashion-MNIST datasets, the attack method can accurately reconstruct the original data for the centralized FL (without protection where the local gradient is available), and cannot retrieve any information for the proposed framework by adding random masks to protect the local gradient. This validates the data protection performance of the proposed masking-enabled framework.

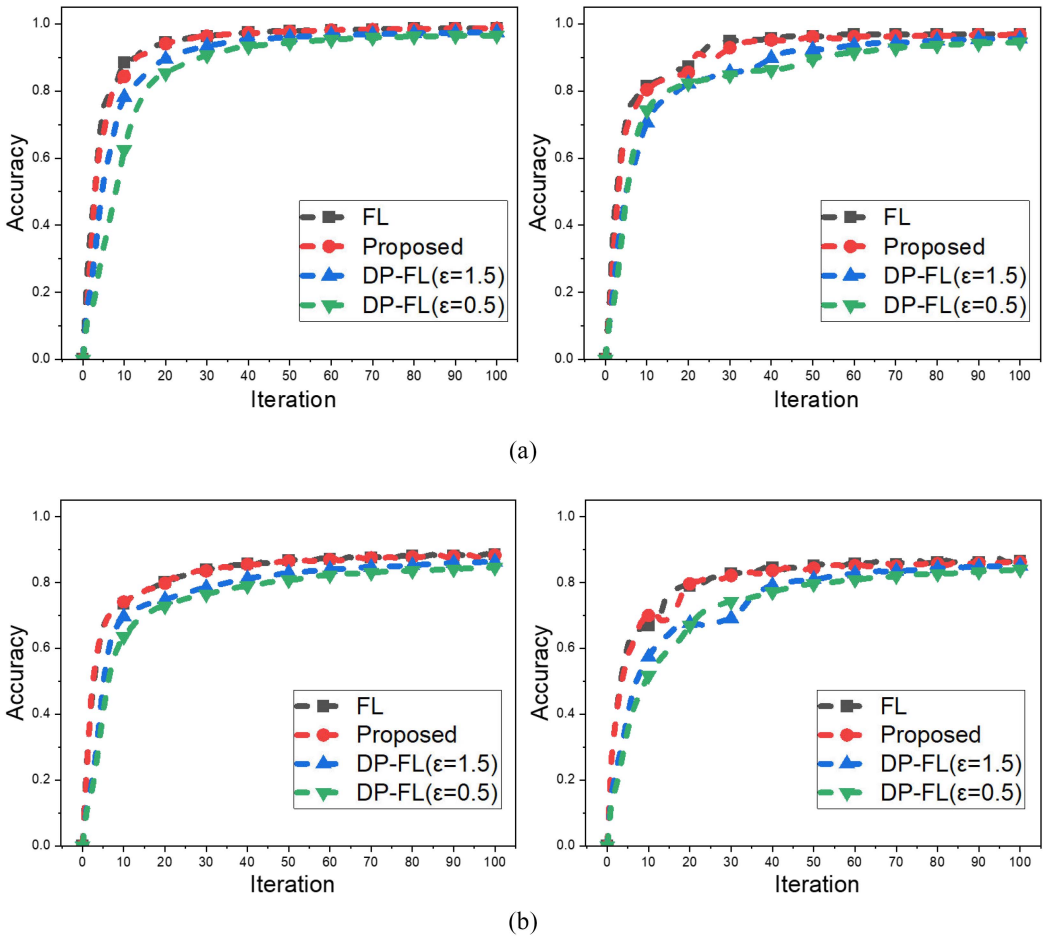

FIGURE 4. The accuracy performances (top-1 accuracy) of centralized FL, the proposed framework, and DP-FL on the MNIST and Fashion-MNIST datasets under IID and non-IID distributions.

The protection performance of DP-FL depends on the level of injected noises into the local gradients. In particular, in the case of mild noises (when $\epsilon = 1.5$), the data can be successfully reconstructed (i.e., one can easily classify the original picture). With the increase of the noise levels, the reconstructed data are increasingly disturbed by many noises and increasingly hard to be classified. In the case of $\epsilon = 0.5$, the reconstructed data for MNIST can hardly be classified. However, the increase in noise levels would also lead to increasing accuracy losses, as already shown in Fig. 4.

3) Signaling Overhead: In the following, we evaluate the fog-level and vehicle-level signaling overhead (due to distributed fog consensus and masking pairing) of the proposed framework and the basic hierarchical decentralized learning (without the signaling reduction designs in Section V). In particular, we evaluate the fog-level signaling overhead via the number of consensus iterations and the vehicle-level signaling overhead via the masking pairing times.

Fig. 6 shows the number of iterations required consensus iterations for reaching a global average among the fog nodes, achieved by the proposed framework and basic hierarchical decentralized learning, as the number of fog nodes N increases. We can see that the number of required iterations decreases with N. This is because, following the stochastic topology generation rule, we generate the fog node topology by setting a constant link establishment probability of 0.3. As a result, with the increase of N, the number of links also increases and the fog nodes are increasingly close to each other, resulting in a decrease of required iterations. Moreover, by optimizing the consensus matrix, the proposed framework can reduce 24.8% average iterations (i.e., signaling overhead) when $N = 10$.

Fig. 7 plots the required masking pairing times of the proposed framework (network-level pairing) and the basic hierarchical decentralized learning (fog-level pairing) as the simulation duration increases. We can see in Fig. 7 that the proposed network-level masking can significantly reduce the signaling overhead, where the pairing times (i.e., corresponding signaling overhead) of the proposed framework is only about 20% of the benchmark (i.e., the fog-level pairing mechanism). This is because the vehicles need to regenerate the pairing seeds at each time of inter-fog handover. In contrast, in the network-level pairing, the pairing seeds are regenerated when the vehicles run out of the

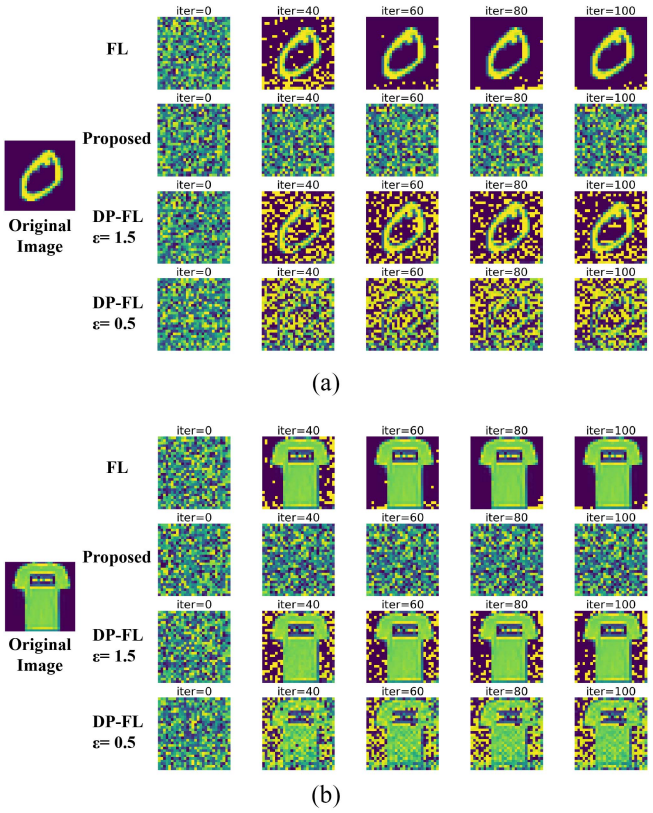

FIGURE 5. The data attack results on MNIST and Fashion-MNIST datasets, as the attack iterations increase to 100.

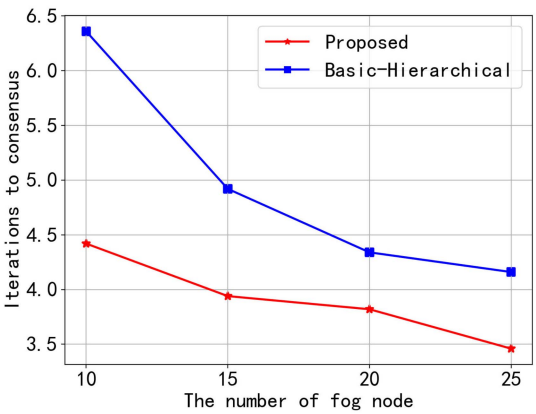

FIGURE 6. The number of iterations required consensus iterations for reaching a global average under different number of fog nodes.

network area, eliminating the unnecessary inter-fog handover overhead. This validates the signaling effectiveness of the proposed network-level masking pairing mechanism.

VII. CONCLUSION

This paper proposed a secure and efficient hierarchical decentralized learning framework for IoV, where federated learning and distributed consensus were integrated for efficient vehicle-fog and inter-fog collaborative learning. We designed the network-level masking mechanism to protect data privacy with reduced signaling overhead, where the

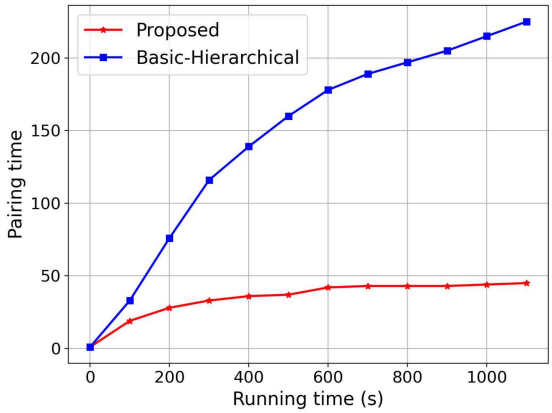

FIGURE 7. The required masking pairing times of the proposed network-level pairing mechanism and the fog-level pairing benchmark.

vehicles can be paired across the coverage of different fog nodes to eliminate the inter-fog handover repairing in the traditional fog-level pairing. The random masks via the network-level pairing were proved to be canceled via distributed consensus, hence preserving learning accuracy. The consensus matrix was optimized via SDP to reduce the signaling due to inter-fog consensus iterations. Experiments were conducted on MNIST and Fashion-MNIST datasets under IID and non-IID distributions. The results validate the effectiveness of the proposed framework in terms of data privacy protection, learning accuracy, and signaling efficiency.

REFERENCES

[1] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1342–1397, 2nd Quart., 2021.

[2] B. Fan, Z. Su, Y. Chen, Y. Wu, C. Xu, and T. Q. S. Quek, "Ubiquitous control over heterogeneous vehicles: A digital twin empowered edge AI approach," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 166–173, Feb. 2023.

[3] M. Chen et al., "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.

[4] M. Dibaei et al., "Investigating the prospect of leveraging blockchain and machine learning to secure vehicular networks: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 683–700, Feb. 2022.

[5] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 14747–14756.

[6] B. Zhao, K. R. Mopuri, and H. Bilen, "IDLG: Improved deep leakage from gradients," 2020, *arXiv:2001.02610*.

[7] A. Wainakh et al., "User-label leakage from gradients in federated learning," 2021, *arXiv:2105.09369*.

[8] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 16937–16947.

[9] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," 2017, *arXiv:1712.07557*.

[10] W. Wei, L. Liu, Y. Wut, G. Su, and A. Iyengar, "Gradient-leakage resilient federated learning," in *Proc. IEEE 41st Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2021, pp. 797–807.

[11] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020.

[12] Y. Aono, T. Hayashi, L. Wang, S. Moriai, and others, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1333–1345, 2017.

[13] J. Ma, S.-A. Naas, S. Sigg, and X. Lyu, "Privacy-preserving federated learning based on multi-key homomorphic encryption," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 5880–5901, 2022.
[14] C. Zhang, S. Ekanut, L. Zhen, and Z. Li, "Augmented multi-party computation against gradient leakage in federated learning," *IEEE Trans. Big Data*, early access, Sep. 22, 2022, doi: 10.1109/TBDATA.2022.3208736.
[15] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, "Verifynet: Secure and verifiable federated learning," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 911–926, 2019.
[16] L. Liu, J. Zhang, S. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.
[17] W. Y. B. Lim, J. S. Ng, Z. Xiong, D. Niyato, C. Miao, and D. I. Kim, "Dynamic edge association and resource allocation in self-organizing hierarchical federated learning networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3640–3653, Dec. 2021.
[18] J. Feng, L. Liu, Q. Pei, and K. Li, "Min-max cost optimization for efficient hierarchical federated learning in wireless edge networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 11, pp. 2687–2700, Nov. 2022.
[19] X. Zhou, W. Liang, J. She, Z. Yan, I. Kevin, and K. Wang, "Two-layer federated learning with heterogeneous model aggregation for 6G supported Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5308–5317, Jun. 2021.
[20] M. Yemini, R. Saha, E. Ozfatura, D. Gündüz, and A. J. Goldsmith, "Semi-decentralized federated learning with collaborative relaying," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2022, pp. 1471–1476.
[21] F. P.-C. Lin, S. Hosseinalipour, S. S. Azam, C. G. Brinton, and N. Michelusi, "Semi-decentralized federated learning with cooperative D2D local model aggregations," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3851–3869, Dec. 2021.
[22] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.
[23] O. L. Saldanha et al., "Swarm learning for decentralized artificial intelligence in cancer histopathology," *Nature Med.*, vol. 28, no. 6, pp. 1232–1239, 2022.
[24] Y. Liu, L. Huo, J. Wu, and A. K. Bashir, "Swarm learning-based dynamic optimal management for traffic congestion in 6G-Driven intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 1, 2020, doi: 10.1109/TITS.2023.3234444.
[25] H. Ye, L. Liang, and G. Y. Li, "Decentralized federated learning with unreliable communications," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 487–500, Apr. 2022.
[26] S. Savazzi, M. Nicoli, and V. Rampa, "Federated learning with cooperating devices: A consensus approach for massive IoT networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4641–4654, May 2020.
[27] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Wireless communications for collaborative federated learning," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 48–54, Dec. 2020.
[28] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, "Communication-efficient federated learning," *Proc. Nat. Acad. Sci.*, vol. 118, no. 17, 2021, Art. no. e2024789118.
[29] J. Li et al., "Budget-aware user satisfaction maximization on service provisioning in mobile edge computing," *IEEE Trans. Mobile Comput.*, early access, Sep. 9, 2022, doi: 10.1109/TMC.2022.3205427.
[30] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.
[31] X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu, and E. Dutkiewicz, "Optimal Online data partitioning for geo-distributed machine learning in edge of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2393–2406, Oct. 2019.
[32] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Engineering*, vol. 8, pp. 33–41, Jan. 2022.
[33] A. Koloskova, N. Loizou, S. Boreiri, M. Jaggi, and S. Stich, "A unified theory of decentralized SGD with changing topology and local updates," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5381–5393.
[34] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist. (COMPSTAT)*, Paris france, 2010, pp. 177–186.
[35] S. Ruder. "An overview of gradient descent optimization algorithms." 2016. [Online]. Available: http://arxiv.org/abs/1609.04747.
[36] H. Yang, M. Ge, D. Xue, K. Xiang, H. Li, and R. Lu, "Gradient leakage attacks in federated learning: Research frontiers, taxonomy and future directions," *IEEE Netw.*, early access, Apr. 24, 2023, doi: 10.1109/MNET.001.2300140.
[37] E. Klaoudatou, E. Konstantinou, G. Kambourakis, and S. Gritzalis, "A survey on cluster-based group key agreement protocols for WSNs," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 429–442, 3rd Quart., 2011.
[38] K. Bhattacharjee and S. Das, "A search for good pseudo-random number generators: Survey and empirical studies," *Comput. Sci. Rev.*, vol. 45, Aug. 2022, Art. no. 100471.
[39] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.
[40] S. P. Karimireddy, S. Kale, M. Mohri, S. J. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," 2019, *arXiv:1910.06378*.
[41] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, 2016.
[42] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO-simulation of urban mobility," *Int. J. Adv. Syst. Meas.*, vol. 5, nos. 3–4, pp. 128–138, 2012.
[43] Y. LeCun. "LeNet-5, convolutional neural networks." 2015. [Online]. Available: URL: http://yann. lecun. com/exdb/lenet

ZIXUAN LIANG received the B.E. degree from the Minzu University of China in 2021. He is currently pursuing the master's degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications. His research interests include Internet of Vehicles, edge intelligence, and distributed learning.

PENGLIN YANG received the Doctoral degree from the China Academy of Space Technology in 2019. He is currently works with China Mobile Research Institute as a Security Engineer. His main research interests include network security, system security, and security issues in federated machine-learning aspect.

CHENYU ZHANG received the B.E. degree from the Shandong University of Science and Technology in 2020, and the master's degree from the Beijing University of Posts and Telecommunications in 2023. His research interests include Internet of Vehicles, edge intelligence, and federated learning.

XINCHEN LYU received the B.E. degree from the Beijing University of Posts and Telecommunications (BUPT) in 2014, and the dual Ph.D. degrees from BUPT and the University of Technology Sydney in 2019. He is currently an Associate Professor with the National Engineering Research Center for Mobile Network Technologies, BUPT, and an Associate Researcher with the Department of Broadband Communication, Peng Cheng Laboratory. His research interests include the resource management and security of edge intelligence and its applications in future wireless networks.