Federated Learning With Local Differential Privacy for Autonomous Electronic Vehicles: Enhancing Security and Performance

Xin Wang, Byung-Gyu Kim, *Senior Member, IEEE*, Mohammed Amoon, Sachin Kumar, and Zhonghua Liu

***Abstract*—The rapid advancement of autonomous electronic vehicles (AEVs) has revolutionized the transportation industry, offering enhanced safety, efficiency, and convenience. However, the collaborative nature of AEVs, which involves exchanging data and models among vehicles and central servers, raises significant concerns regarding privacy and security. Federated learning (FL) has emerged as a promising approach to enable collaborative model training while preserving data confidentiality. Nevertheless, the vulnerability of FL to inference attacks and the potential presence of malicious participants pose substantial challenges. To address these issues, we propose a novel FL framework with local differential privacy (LDP-FL) tailored specifically for AEVs. Our framework incorporates LDP mechanism to protect individual AEV data by perturbing the model updates before sharing them with the central server. Moreover, we introduce a performance loss constraint mechanism to minimize the impact of the privacy-preserving noise on the model's performance. The LDP-FL algorithm ensures secure and efficient collaborative learning among AEVs while preserving data privacy. Extensive experiments using the Udacity dataset demonstrate the superiority of LDP-FL compared to state-of-the-art baseline methods for global accuracy, performance loss, running time, energy consumption, communication latency, and robustness against malicious attacks. The LDP-FL framework achieves a remarkable balance between privacy, security, and performance.**

***Index Terms*—Autonomous electronic vehicles, federated learning, local differential privacy, privacy-preserving.**

Received 21 July 2024; revised 9 October 2024, 30 November 2024, and 5 January 2025; accepted 17 February 2025. Date of publication 14 March 2025; date of current version 14 August 2025. This work was supported by the Researchers Supporting Project, King Saud University Riyadh, Saudi Arabia, under Grant RSPD2025R968. *(Corresponding author: Zhonghua Liu.)*

Xin Wang is with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: dnsy_heinrich@neueet.com).

Byung-Gyu Kim is with the Department of AI Engineering, Sookmyung Women's University, Seoul 04310, Republic of Korea (e-mail: bg.kim@sookmyung.ac.kr).

Mohammed Amoon is with the Department of Computer Science, Community College, King Saud University, Riyadh 11437, Saudi Arabia (e-mail: mamoon@ksu.edu.sa).

Sachin Kumar is with the Department of Data Science, Galgotias College of Engineering and Technology, Greater Noida 201310, India (e-mail: imsachingupta@rediffmail.com).

Zhonghua Liu is with the Computer Center, Shengjing Hospital of China Medical University, Shenyang 110004, China (e-mail: liuzh@sj-hospital.org).

Digital Object Identifier 10.1109/TCE.2025.3551511

I. Introduction

THE ADVENT of autonomous electronic vehicles (AEVs) represents a shift in transportation, offering a future where roads are safer, traffic flows more efficiently, and mobility is enhanced for all [1], [2]. AEVs are self-driving vehicles with advanced sensors, artificial intelligence, and control systems that enable them to navigate and operate without human intervention. As these self-driving marvels become increasingly common, they serve as mobile data hubs, constantly collecting and processing vast amounts of information from sophisticated sensors [3], [4]. These include high-resolution cameras that capture the vehicle's surroundings, LiDAR systems that create detailed 3D maps of the environment, radar for detecting obstacles and measuring distances, and GPS for precise location tracking [5]. This wealth of data is the lifeblood of the intricate machine learning models forming autonomous driving systems' brains, enabling AEVs to navigate complex road conditions, make split-second decisions, and continuously improve their performance. However, this data-driven approach comes with significant privacy implications [6]. The information gathered by AEVs is not just technical but deeply personal, potentially offering a window into the lives of passengers. From daily commutes and favorite destinations to driving behaviors and personal habits, the data collected by AEVs could, if mishandled, reveal intimate details about individuals' lives, raising critical questions about privacy and data protection in the age of autonomous transportation.

The conventional approach to machine learning, which involves centralizing vast amounts of data in a single location for model training, presents a significant privacy conundrum in the context of AEVs. While computationally efficient, this method creates a potential single point of failure for data breaches and raises concerns about the concentration of sensitive information [7]. Enter federated learning (FL), a groundbreaking paradigm reshaping the landscape of collaborative AI development [8], [9], [10]. FL allows AEVs to contribute to improving global models without compromising the privacy of their locally stored data. Each vehicle or fleet becomes a node in a vast network in this decentralized learning ecosystem, training models on their unique datasets derived from real-world driving experiences. Instead of raw data, only the distilled wisdom in the form of model updates is shared with a central server. This server then acts as an orchestrator, aggregating these insights to refine a master model, which is disseminated back to the participating vehicles.

While FL offers significant privacy advantages by decentralized raw data, it is not a complete panacea for all security concerns in AEVs. The shared model parameters exchanged during the FL process can still be susceptible to inference attacks, where malicious actors employ sophisticated techniques to extract sensitive information about the underlying data or the individual AEVs [11], [12]. Furthermore, the distributed nature of FL introduces additional security risks, as compromised or malicious AEVs participating in the collaborative learning process may deliberately manipulate the model updates they send to the central server. Such attacks can have severe consequences, ranging from degrading the overall performance of the learned models to launching coordinated assaults that jeopardize the safety and reliability of the autonomous driving system. To address these critical challenges, it is imperative to develop efficient and lightweight cryptographic primitives specifically tailored for FL in AEVs. These secure computation techniques should provide robust privacy and security guarantees, protecting the confidentiality and integrity of the shared model parameters while being computationally feasible for resource-constrained AEVs. By integrating advanced cryptographic mechanisms into the FL framework, we can establish a strong foundation for building trustworthy and resilient AEVs that can withstand the ever-evolving landscape of cyber threats.

The main contributions of this paper are as follows:

1) We propose LDP-FL, a novel FL framework with local differential privacy (LDP) for AEVs, to enhance the privacy and security of collaborative model training.
2) We design an LDP mechanism that adds noise to the model parameters during the transmission process, preventing inference attacks and protecting sensitive information.
3) We introduce a performance loss constraint mechanism to minimize the impact of LDP on the model's performance, ensuring the effectiveness of the trained models.

The remainder of this paper is organized as follows. Section II provides an overview of related work on FL and privacy-preserving techniques in AEVs. Section III describes the problem formulation and the proposed LDP-FL framework. Section IV discusses the experimental setup and results. Finally, Section V concludes the paper and outlines future research directions.

II. Related Work

The integration of FL in AEVs has gained the considerably significant attention in recent years due to its ability to enable collaborative learning while preserving data privacy. Rani et al. [13] studied FL-based misbehavior detection for the 5G-enabled AEVs. Several studies have explored the application of FL in specific aspects of autonomous driving. For instance, Song et al. [14] introduced a federated transformer learning approach for Bird's eye view perception. Moreover, integrating FL with other emerging technologies, such as edge computing and 5G networks, has been investigated to enhance further the efficiency and reliability of FL in AEVs. Wu et al. [15] proposed a social-aware decentralized cooperative caching for IoV, using the FL framework to train the collaborative caching algorithm based on deep reinforcement learning.

As the adoption of AEVs continues to grow, ensuring the robustness and security of FL algorithms becomes increasingly critical. Ensuring the privacy of sensitive information is a fundamental requirement in FL, especially when applied to AEVs. Various privacy-preserving techniques have been proposed to address this challenge in recent years. Differential privacy (DP) has emerged as a popular approach for protecting individual data points in FL [16], [17], [18], [19]. The combination of multiple privacy-preserving techniques has also been investigated to provide comprehensive privacy guarantees. Parekh et al. [20] presented GeFL: gradient encryption-aided privacy preserved FL for AVs. Chen et al. [21] proposed a novel Byzantine-fault-tolerant (BFT) decentralized FL method with privacy preservation for AVs called BDFL. Wang et al. [22] proposed FL-assisted connected AV (FLCAV). Wang et al. [23] proposed an attack against FL-based AV framework (ATT_FLAV) to evaluate and enhance the robustness of the FL-based autonomous driving models. Li et al. [24] proposed a traceable identity-based privacy-preserving scheme to protect vehicular message privacy with the improved Dijk-Gentry-Halevi-Vaikutanathan (DGHV) algorithm (tiDGHV). He et al. [25] proposed Bift, a blockchain-based FL system for CAVs.

While these works address important aspects of AEV privacy, they do not specifically tackle the privacy challenges in collaborative model training using FL. Existing methods do not adequately address balancing privacy preservation with model performance in resource-constrained AEVs. Additionally, they lack mechanisms to handle the dynamic nature of AEV networks and the potential for adversarial attacks in a federated learning setting. Additionally, our focus is on LDP; other privacy-preserving techniques, such as homomorphic encryption and secure multi-party computation, offer alternative approaches to protecting data in federated learning systems. However, these methods often incur higher computational overhead compared to LDP. Our work fills this gap by providing a comprehensive framework for privacy-preserving FL for AEVs.

III. Proposed Method

A. Problem Description

Consider a scenario where multiple AEVs, each equipped with various sensors and data collection capabilities, participate in a FL process to train a shared machine learning model for tasks such as object detection, trajectory prediction, or decision making. Let $\mathrm{V} = V_1, V_2, \ldots, V_N$ denote the set of participating AEVs, where N is the total number of vehicles. Each AEV V_i has its local dataset D_i, which consists of data collected from its sensors, such as camera images, LiDAR point clouds, and GPS coordinates [26], [27].

TABLE I
NOTATION TABLE

Symbol	Description
V	Set of participating AEVs
N	Total number of AEVs
V_i	The i -th AEV
D_i	Local dataset of the i -th AEV
M	Global model for autonomous driving tasks
ε	Privacy budget for LDP
δ	Probability of privacy failure
C	Clipping threshold for gradients
σ	Noise scale for LDP
T	Total number of communication rounds
η	Learning rate for model updates
p	Perturbation probability

Drivers' psychological comfort and trust are crucial factors in autonomous vehicle adoption. Our framework incorporates user experience feedback mechanisms that help maintain driver confidence while preserving privacy. Regular transparent updates about data protection measures are provided through the vehicle interface, helping drivers understand how their information is being protected without compromising system performance.

The objective of the FL process is to collaboratively train a global model M that can be used by all participating AEVs for autonomous driving tasks. The global model is trained by aggregating the locally trained models from each AEV, without requiring the direct sharing of raw data. However, even though the raw data remains on the individual AEVs, there is still a risk of information leakage through the shared model parameters.

We propose a FL framework with LDP for AEVs to mitigate this risk [28]. We aim to prevent honest-but-curious participants from inferring sensitive information about individual AEVs or their data by analyzing the shared model parameters during the training process.

Each AEV V_i performs local training on its dataset D_i to update the global model parameters in our proposed framework. Before sharing the updated model parameters with the central server or other AEVs, an LDP mechanism is applied to add noise to the parameters, obscuring the contributions of individual data points. The central server then aggregates the noisy updates from all participating AEVs to obtain the updated global model.

The LDP-FL framework incorporates a dynamic participation mechanism to handle AEVs joining or leaving the federation. New vehicles undergo an initialization phase, receiving the current global model and gradually increasing their contribution weight. Departing vehicles' contributions are phased out over several rounds to maintain model stability.

For clarity, the terms used in this study are described in Table I.

For a newly joining AEV j at time t, its contribution weight $w_j(t)$ is calculated as:

$$w_j(t) = \min\left(1, \frac{t - t_{j,\text{join}}}{T}\right). \tag{1}$$

where $t_{j,\text{join}}$ is the joining time, T is the warmup period, and t is the current time.

For a departing AEV k, its contribution weight $w_k(t)$ phases out as:

$$w_k(t) = \max\left(0, 1 - \frac{t - t_{k,\text{leave}}}{\tau}\right). \tag{2}$$

where $t_{k,\text{leave}}$ is the departure time, τ is the phase-out period, and t is the current time.

The global model update is then weighted accordingly:

$$\theta_{t+1} = \theta_t + \eta \sum_i w_i(t) \Delta\theta_i. \tag{3}$$

where θ_t is the global model at time t, η is the learning rate, and $\Delta\theta_i$ is the update from AEV i.

The LDP mechanism ensures that the shared model parameters do not reveal sensitive information about the individual AEVs or their local datasets. By providing a strong privacy guarantee, our framework allows AEVs to collaboratively train a robust and accurate model for autonomous driving tasks while preserving the confidentiality of their sensitive data.

The LDP-FL framework incorporates adaptive computation offloading to account for heterogeneity in AEV hardware capabilities. AEVs with limited computational resources can offload more intensive tasks to nearby edge servers or more capable vehicles, ensuring that all participants can contribute effectively to the federated learning process.

The framework implements multi-user privacy partitioning for shared vehicle scenarios. When multiple users access the same vehicle, the system creates isolated privacy zones for each user's data, ensuring individual privacy protection while maintaining system efficiency. This partitioning enables secure data handling in ride-sharing applications without compromising individual user privacy.

B. LDP Mechanism

Our framework accommodates varying privacy expectations across different geographical regions and cultures. The system implements adjustable privacy thresholds that can be configured according to local regulations and cultural norms while maintaining core security features. This flexibility allows for broader international deployment while respecting regional privacy sensitivities.

In the LDP-FL framework, each AEV V_i performs local training on its dataset D_i to update the global model parameters. The local objective function for AEV V_i is defined as:

$$\text{mi}n_{\mathbf{w}_i}\mathrm{L}_i(\mathbf{w}_i) = \frac{1}{|D_i|}\sum_{j=1}^{|D_i|} \mathrm{l}\left(\mathbf{w}_i; d_{i,j}\right). \tag{4}$$

where $\mathbf{w}_i$ represents the local model parameters of AEV V_i, L_i is the local objective function, $\mathrm{l}(\cdot)$ is the loss function, and $d_{i,j}$ denotes the j-th data sample in the local dataset D_i.

The LDP-FL framework implements an asynchronous update mechanism to mitigate the impact of communication failures or network disruptions. AEVs can continue local training and cache updates during disconnected periods, submitting them when connectivity is restored. The central server employs a staleness-aware aggregation technique to weight these delayed updates appropriately.

To ensure LDP, we apply the Gaussian mechanism to perturb the local model updates before sharing them with the central server [29]. The sensitivity of the local model update is defined as:

$$\Delta_i = \max_{\mathbf{w}_i,\mathbf{w}_{i'}} |\nabla \mathrm{L}_i(\mathbf{w}_i) - \nabla \mathrm{L}_i(\mathbf{w}_{i'})|_2 \le \frac{2C}{|D_i|}. \tag{5}$$

where C is the clipping threshold for the gradients, and $|D_i|$ is the size of the local dataset.

To achieve (ε, δ)-LDP, the Gaussian noise added to the local model update should have a standard deviation of:

$$\sigma_i = \frac{\Delta_i \sqrt{2\ln(1.25/\delta)}}{\varepsilon}. \tag{6}$$

where ε is the privacy budget, δ is the probability of privacy failure, and the term 1.25 is a constant that arises from the mathematical derivation of the Gaussian mechanism for achieving (ε, δ)-differential privacy. It is related to the tail bounds of the Gaussian distribution and the desired probability of privacy failure δ.

The perturbed local model update $\widetilde{\mathbf{w}}_i$ is then computed as:

$$\widetilde{\mathbf{w}}_i = \mathbf{w}_i + \mathrm{N}\left(0, \sigma_i^2 \mathbf{I}\right). \tag{7}$$

where $\mathrm{N}(0, \sigma_i^2 \mathbf{I})$ represents a Gaussian distribution with zero mean and standard deviation σ_i.

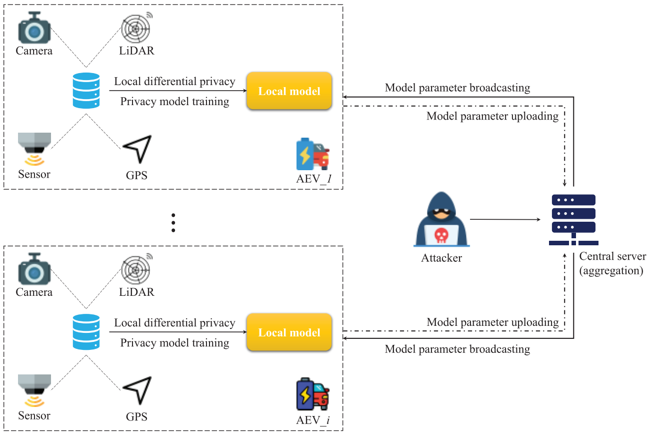
Camera
LiDAR
Local differential privacy
Privacy model training
Local model
Sensor
GPS
AEV_I
Model parameter broadcasting
Model parameter uploading
Attacker
Central server (aggregation)
Camera
LiDAR
Local differential privacy
Privacy model training
Local model
Model parameter uploading
Model parameter broadcasting
Sensor
GPS
AEV_i

Fig. 1. LDP-FL framework for AEVs.

The LDP-FL framework, illustrated in Fig. 1, showcases a distributed approach to collaborative learning for AEVs. At the heart of the framework lies a central server, acting as a coordinator and aggregator, surrounded by multiple participating AVs. Each AV, equipped with its local dataset, independently trains a model using this data. The AVs employ the LDP mechanism to ensure privacy, introducing carefully calibrated noise into the model updates. These perturbed updates are then securely transmitted to the central server. Upon receiving the noisy updates from all participating AVs, the central server performs an aggregation process, typically by averaging the updates, to construct an updated global model. This refined model is then disseminated back to the AVs, enabling them to initiate the next round of local training. Through this iterative process, the LDP-FL framework facilitates collaborative learning among AVs while prioritizing data privacy and security.

The LDP-FL framework employs a hierarchical aggregation structure to enhance scalability for larger AEV fleets. AEVs are grouped into clusters, with local aggregators handling intra-cluster communication and aggregation. This approach reduces the load on the central server and improves overall system scalability. The LDP-FL algorithm integrates the LDP mechanism and the performance loss constraint mechanism into the FL process for AEVs. The algorithm consists of two main components: the server-side aggregation and the client-side local training with privacy-preserving noise addition.

The server-side aggregation is performed by a central server, which coordinates the FL process among the participating AEVs. While on the client side, each participating AEV performs local training on its dataset, which consists of data collected from various sensors such as cameras, LiDAR, and GPS. The local training process involves computing the gradients of the local objective function concerning the current global model parameters.

Environmental conditions significantly influence data collection quality in autonomous vehicles. During adverse weather conditions such as heavy rain or snow, our framework adjusts data collection frequencies and implements adaptive filtering mechanisms to maintain data quality. The system automatically compensates for reduced sensor reliability by enhancing collaboration between nearby vehicles while maintaining privacy standards.

The LDP-FL algorithm proceeds iteratively, with multiple communication rounds between the server and the participating AEVs. In each round, the selected AEVs perform local training with privacy-preserving noise addition, and the server aggregates the received updates to improve the global model. This process continues until a predefined number of rounds is reached or a certain convergence criterion is met.

The framework includes an intuitive interface layer that mediates between human inputs and AI decisions. This interface provides real-time feedback about privacy-protected decisions while allowing drivers to understand and override system choices when necessary, creating a balanced approach to human-AI collaboration without compromising data security.

Algorithm 1 presents the pseudocode for the central server in the LDP-FL framework. The server initializes the global model $\mathbf{w}_0$ and iteratively performs the following steps for each communication round t:

1) Select a subset of AEVs S_t to participate in the current round.
2) Distribute the current global model $\mathbf{w}_t$ to the selected AEVs.
3) Receive the noisy local model updates $\widetilde{\mathbf{w}}_{i,t}$ from the participating AEVs.
4) Compute the updated global model $\mathbf{w}_{t+1}$ by averaging the received noisy updates.

Algorithm 2 describes the local training process for each AEV in the LDP-FL framework. Each AEV V_i performs the following steps:

1) Receive the current global model $\mathbf{w}_t$ from the central server.

Algorithm 1 LDP-FL Server

```
Input: Initial global model parameters $\mathbf{w}_0$, number of rounds $T$, set of clusters $\mathcal{C}$, set of AEVs $\mathcal{V}$
Output: Final global model parameters $\mathbf{w}_T$
initialize $\mathbf{w}_0$
for $t = 1$ to $T$ do
    initialize cluster_updates = { }
    for each cluster $c$ in $\mathcal{C}$ do
        initialize local_aggregato$r\_update = 0$
        select active AEVs $\mathcal{V}_c \subseteq \mathcal{V}$ in cluster $c$
        distribute $\mathbf{w}_{t-1}$ to AEVs in $\mathcal{V}_c$
        for each AEV $i$ in $\mathcal{V}_c$ do
            receive update $\Delta\mathbf{w}_i$ from AEV $i$
            local_aggregator_update += $\Delta\mathbf{w}_i/|\mathcal{V}_c|$
        end for
        cluster_updates [$c$] = local_aggregator_update
    end for
    cluster_updates [$c$]/$|\mathcal{C}|$
end for
```

Algorithm 2 LDP-FL AEV

```
Input: Current global model parameters received from the server $\mathbf{w}_t$, local dataset $D_i$ of AEV $V_i$, learning rate $\eta$ for local model updates, clipping threshold for gradients $C$, privacy budget $\varepsilon_i$ for AEV $V_i$, $p = 1$ for maximum privacy protection, and probability $\delta_i$ of privacy failure for AEV $V_i$.
Output: $\widetilde{\mathbf{w}}_{i,t}$: The noisy local model update to be sent to the server
initialize $\mathbf{w}_t$, $D_i$
$\mathbf{w}_{i,t} \leftarrow \mathbf{w}_{t-\eta}\nabla \mathrm{L}_i(\mathbf{w}_t)$
$\mathbf{w}_{i,t} \leftarrow \mathbf{w}_{i,t}/\max\left(1, \frac{|\mathbf{w}_{i,t}|2}{C}\right)$
// check performance loss constraint:
if $\alpha(\mathrm{P}_{i,t}) < \mathbb{E}v_0 \sim p_0[v_1]$ then
    adjust $\sigma_i$ to satisfy constraint
end-if
$\sigma_i \leftarrow \frac{2C\sqrt{2\ln(1.25/\delta)}}{|D_i|\varepsilon}$
$\widetilde{\mathbf{w}}_{i,t} \leftarrow \mathbf{w}_{i,t} + \mathrm{N}(0, \sigma_i^2\mathbf{I})$
send $\widetilde{\mathbf{w}}_{i,t}$ to the central server
apply robust aggregation to filter out potential malicious updates
aggregate filtered updates to obtain global model update
```

2) Perform local training on its dataset D_i to compute the local model update $\mathbf{w}_{i,t}$.
3) Clip the local model update to ensure bounded sensitivity.
4) Apply the LDP mechanism to perturb the local model update and obtain $\widetilde{\mathbf{w}}_{i,t}$.
5) Send the noisy local model update $\widetilde{\mathbf{w}}_{i,t}$ to the central server.

C. Performance Loss Constraint Mechanism

While the LDP mechanism enhances the privacy of FL in AEVs, it can also lead to performance degradation due to the added noise. To mitigate the impact of privacy-preserving noise on the model's performance, we introduce a performance loss constraint mechanism to minimize the performance loss while ensuring a desired level of privacy.

Let P_i denote the performance loss of AEV V_i resulting from the LDP mechanism. The performance loss is defined as the difference between the loss of the model trained with the perturbed local updates and the loss of the model trained with the original local updates:

$$\mathrm{P}_i = \mathrm{L}_i(\widetilde{\mathbf{w}}_i) - \mathrm{L}_i(\mathbf{w}_i). \tag{8}$$

where $\widetilde{\mathbf{w}}_i$ represents the perturbed local model parameters, and $\mathbf{w}_i$ represents the original local model parameters.

The overall performance loss P across all participating AEVs after T communication rounds is given by:

$$\mathrm{P} = \sum_{i=1}^{N}\sum_{t=1}^{T} \exp\big(\alpha(\mathrm{P}_{i,t})\big). \tag{9}$$

where $\mathrm{P}_{i,t}$ is the performance loss of AEV V_i in round t. α is a monotonic function that maps the performance loss to a non-negative value. Specifically, we define α as the KL divergence between the perturbed and original model distributions, which naturally satisfies these properties while providing a mathematically rigorous measure of information loss due to perturbation.

To constrain the performance loss while ensuring differential privacy, we formulate an optimization problem that minimizes the overall performance loss subject to the LDP constraint:

$$\begin{aligned} &\min\nolimits_{\sigma_1,\ldots,\sigma_N} \quad \mathrm{P} \\ &\qquad\qquad \text{s.t.} \quad \sigma_i \geq \frac{\Delta_i\sqrt{2\ln(1.25/\delta)}}{\grave{o}}, \forall i \in 1, \ldots, N. \end{aligned} \tag{10}$$

where σ_i is the noise scale for AEV V_i, Δ_i is the sensitivity of the local model update, ε is the privacy budget, and δ is the probability of privacy failure.

The performance loss constraint is integrated into the neural network training process through a modified loss function. We use a Lagrangian multiplier approach to incorporate the constraint, allowing the network to balance, minimize the original loss, and satisfy the performance loss constraint during backpropagation [30]. The $\mathrm{L}(\sigma_1, \ldots, \sigma_N, \lambda_1, \ldots, \lambda_N)$ is used to optimize the noise scales σ_i while satisfying the privacy constraints. The noise scales are optimized independently of the model weights to ensure we achieve the desired level of privacy while minimizing the performance loss. The model weights are updated through standard gradient descent on the task-specific loss function, with the optimized noise scales applied to the resulting gradients.

$$\begin{aligned} &\mathrm{L}(\sigma_1, \ldots, \sigma_N, \lambda_1, \ldots, \lambda_N) \\ &= \mathrm{P} + \sum_{i=1}^{N} \lambda_i\left(\frac{\Delta_i\sqrt{2\ln(1.25/\delta)}}{\varepsilon} - \sigma_i\right). \end{aligned} \tag{11}$$

where λ_i is the Lagrange multiplier associated with the constraint for AEV V_i.

Taking the partial derivatives of the Lagrangian function with respect to σ_i and λ_i, and setting them to zero, we obtain:

$$\frac{\partial \mathrm{L}}{\partial \sigma_i} = \frac{\partial \mathrm{P}}{\partial \sigma_i} - \lambda_i = 0, \ \forall i \in 1, \ldots, N. \tag{12}$$

$$\frac{\partial \mathrm{L}}{\partial \lambda_i} = \frac{\Delta_i\sqrt{2\ln(1.25/\delta)}}{\varepsilon} - \sigma_i = 0, \quad \forall i \in 1, \ldots, N. \tag{13}$$

To compute $\frac{\partial \mathrm{P}}{\partial \sigma_i}$, we need to evaluate the performance loss function $\alpha(\mathrm{P}_{i,t})$. To properly $\alpha(\mathrm{P}_{i,t})$, we use the log-likelihood ratio between the perturbed and original model distributions:

$$\alpha\left(\mathrm{P}_{i,t}\right) = \mathrm{KL}(p_1 || p_0) = \int p_1(z) \log\left(\frac{p_1(z)}{p_0(z)}\right) \mathrm{dz}. \tag{14}$$

where p_1 represents the probability density function of the perturbed model update, p_0 represents the probability density function of the original model update, and KL denotes the Kullback-Leibler divergence.

Let v_0 denote the probability density function of the original model update, which follows a Gaussian distribution with zero mean and standard deviation σ_i:

$$v_0(z) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{z^2}{2\sigma_i^2}\right). \tag{15}$$

Let v_1 denote the probability density function of the perturbed model update, which follows a mixture of Gaussian distributions with means 0 and Δ_i, and standard deviation σ_i:

$$\begin{aligned} v_1(z) = {} & q\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(z-\Delta_i)^2}{2\sigma_i^2}\right) \\ & +(1-q)\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{z^2}{2\sigma_i^2}\right). \end{aligned} \tag{16}$$

where q is the probability of the model update being perturbed.

To derive Eq. (14), we start with the definition of Renyi divergence and apply it to our specific case of Gaussian distributions [31]. We then simplify the resulting expression using the properties of logarithms and exponentials. Similarly, for Eq. (15), we use the definition of mutual information and express it in terms of the Renyi divergence we just derived.

$$E_{v_1 \sim p_1}[v_1] = \int_{-\infty}^{\infty} v_1(z)\left(\frac{v_1(z)}{v_0(z)}\right)^{\beta-1} \mathrm{dz}. \tag{17}$$

$$E_{v_0 \sim p_0}[v_0] = \int_{-\infty}^{\infty} v_0(z)\left(\frac{v_0(z)}{v_1(z)}\right)^{\beta-1} \mathrm{dz}. \tag{18}$$

where $\beta > 1$ is a constant.

To determine the monotonicity of $\alpha(\mathrm{P}_{i,t})$, we define $\gamma = \exp(\frac{(z-\Delta_i)^2}{2\sigma_i^2})$ and evaluate the derivative of $\alpha(\mathrm{P}_{i,t})$ with respect to γ:

$$\begin{aligned} \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{d\gamma} &= \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{\mathrm{dz}} \cdot \frac{\mathrm{dz}}{d\gamma} \\ &= \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{\mathrm{dz}} \cdot \frac{1}{\gamma} \cdot \left(\frac{q}{1-q}\right)^{\frac{1}{\beta-1}} \cdot (\gamma-1)^{\frac{1}{\beta-1}-1} \\ &= \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{\mathrm{dz}} \cdot \frac{1}{\gamma} \cdot \left(\frac{q}{1-q}\right)^{\frac{1}{\beta-1}} \cdot \left(\gamma^{\frac{1}{\beta-1}} - \gamma^{\frac{1}{\beta-1}-1}\right). \end{aligned} \tag{19}$$

Considering $\beta > 1$, we have:

$$\frac{q}{1-q} \cdot \gamma^{\frac{1}{\beta-1}} - 1 \geq \left(\frac{q}{1-q}\right)^{\frac{1}{\beta-1}} - 1. \tag{20}$$

Through algebraic manipulations, we obtain:

$$\left(\frac{q}{1-q}\right)^{\frac{1}{\beta-1}} \geq 1 + \frac{1}{q} \cdot \left(1 - \frac{1}{q}\right)^{\frac{1}{\beta-1}}. \tag{21}$$

Substituting this result into Eq. (16), we can rewrite $\frac{\mathrm{d}\alpha(\mathrm{P}_{i,t})}{\mathrm{d}\gamma}$ as:

$$\begin{aligned} \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{d\gamma} = {} & \frac{d\alpha\left(\mathrm{P}_{i,t}\right)}{\mathrm{dz}} \cdot \frac{1}{\gamma} \cdot \left(\frac{q}{1-q}\right)^{\frac{1}{\beta-1}} \\ & \cdot \left(\gamma^{\frac{1}{\beta-1}} - 1 - \frac{1}{q} \cdot \left(1 - \frac{1}{q}\right)^{\frac{1}{\beta-1}}\right). \end{aligned} \tag{22}$$

Let $\psi(\gamma) = \gamma^{\frac{1}{\beta-1}} - 1 - \frac{1}{q} \cdot (1 - \frac{1}{q})^{\frac{1}{\beta-1}}$. Taking the first and second derivatives of $\psi(\gamma)$ with respect to γ, we obtain:

$$\begin{aligned} \frac{d\psi(\gamma)}{d\gamma} &= \frac{1}{\beta-1} \cdot \gamma^{\frac{1}{\beta-1}-1} \\ \frac{d^2\psi(\gamma)}{d\gamma^2} &= \frac{1}{(\beta-1)^2} \cdot \left(\frac{1}{\beta-1} - 1\right) \cdot \gamma^{\frac{1}{\beta-1}-2}. \end{aligned} \tag{23}$$

Let $h(\sigma)$ denote the privacy-utility trade-off function defined as:

$$h(\sigma) = \exp\left(-\frac{\sigma^2}{2}\right). \tag{24}$$

which characterizes how increasing the noise scale σ affects the utility of the model while providing privacy guarantees.

The derivation from Eqs. (14) to (23) establishes a more computationally tractable form of the optimization problem. The key motivation is to transform the original constrained optimization in Eq. (11) into a simplified inequality constraint based on the log-likelihood ratio. First, by expressing the performance loss in terms of the log-likelihood ratio between perturbed and original model updates (Eqs. (14) to (16)), we obtain a more direct measure of information loss due to privacy-preserving perturbations. The subsequent derivation through Eqs. (17) to (23) leads to the key inequality. This inequality, derived from analyzing the monotonicity of $h(\sigma)$, provides a more efficient optimization criterion than the original Lagrangian formulation.

While the derived inequality appears to depart from the original Lagrangian formulation, it serves as a constraint in our modified optimization problem:

$$\begin{aligned} & \min \ \mathrm{L}(\mathbf{w}, \sigma) \\ & \text{s.t.} \ \ \alpha(\mathrm{P}i, t) \geq \mathbb{E}v_0 \sim p_0[v_1]. \end{aligned} \tag{25}$$

where $\min \mathrm{L}(\mathbf{w}, \sigma)$ is our original loss function. This reformulation provides computational advantages as it reduces the optimization to a constrained problem with a simpler constraint structure while maintaining the privacy guarantees of the original formulation.

D. Privacy and Security Analysis

Next, we analyze the privacy and security guarantees provided by the LDP-FL framework. We first prove that the LDP mechanism satisfies (ε, δ)-LDP. Then, we discuss the robustness of LDP-FL against inference attacks and malicious participants.

Let M denote the LDP mechanism applied to the local model update $\mathbf{w}_i$ of AEV V_i. For any two neighboring datasets D_i and $D_{i'}$ that differ in a single data point, and any subset of outputs $S \subseteq Range(\mathrm{M})$, we have:

$$\begin{aligned}
&Pr[\mathrm{M}(\mathbf{w}_i(D_i)) \in S] \\
&= \int S \frac{1}{\left(2\pi\sigma_i^2\right)^{d/2}} \exp\left(-\frac{|\mathbf{z} - \mathbf{w}_i(D_i)|2^2}{2\sigma_i^2}\right) d\mathbf{z} \\
&\leq \int S \frac{1}{\left(2\pi\sigma_i^2\right)^{d/2}} \exp\left(-\frac{|\mathbf{z} - \mathbf{w}_i(D_{i'})|_2^2 - 2\Delta_i|\mathbf{z} - \mathbf{w}_i(D_{i'})|2}{2\sigma_i^2}\right) d\mathbf{z} \\
&\leq \exp\left(\frac{\varepsilon_i}{2}\right) \cdot \int S \frac{1}{\left(2\pi\sigma_i^2\right)^{d/2}} \exp\left(-\frac{|\mathbf{z} - \mathbf{w}_i(D_{i'})|_2^2}{2\sigma_i^2}\right) \\
&\exp\left(\frac{\Delta_i\sqrt{2\mathrm{qTln}(1/\delta_i)}}{\epsilon_i}\right) d\mathbf{z} \\
&\leq exp(\varepsilon_i) \cdot \Pr[\mathrm{M}(\mathbf{w}_i(D_{i'})) \in S] + \delta_i. \qquad (26)
\end{aligned}$$

where d is the dimension of the model update vector.

To satisfy $(\varepsilon_i, \delta_i)$-LDP, the following condition should hold:

$$\exp\left(\frac{\Delta_i\sqrt{2\mathrm{qTln}(1/\delta_i)}}{\varepsilon_i}\right) \leq exp(\varepsilon_i). \qquad (27)$$

Solving for σ_i, we obtain:

$$\sigma_i \geq \frac{\Delta_i\sqrt{2\mathrm{qTln}(1/\delta_i)}}{\varepsilon_i}. \qquad (28)$$

It is guaranteed that the local model updates shared by the AEVs during the FL process do not reveal sensitive information about individual data points in their local datasets. The privacy budget ε_i and the parameter δ_i control the level of privacy provided by the LDP mechanism for each AEV.

To further analyze the privacy guarantees of the LDP-FL algorithm, we consider the composition of the LDP mechanism over multiple communication rounds [32], [33]. According to the sequential composition theorem, the overall privacy guarantee for AEV V_i after T rounds is $(\varepsilon_{i'}, \delta_{i'})$-differential privacy, where:

$$\varepsilon_{i'} = \sum_{t=1}^{T} \varepsilon_{i,t}. \qquad (29)$$

$$\delta_{i'} = \sum_{t=1}^{T} \delta_{i,t}. \qquad (30)$$

where $\varepsilon_{i,t}$ and $\delta_{i,t}$ denote the privacy parameters for AEV V_i in round t.

We can allocate the privacy budget across the communication rounds using the advanced composition theorem to achieve a desired overall privacy level $(\varepsilon_{i'}, \delta_{i'})$. By setting $\varepsilon_{i,t} = \frac{\varepsilon_{i'}}{\sqrt{2Tln(1/\delta_{i'})}}$ and $\delta_{i,t} = \frac{\delta_{i'}}{T}$, the LDP-FL algorithm guarantees $(\varepsilon_{i'}, \delta_{i'})$-differential privacy for AEV V_i after T rounds.

In addition to privacy concerns, the FL process in AEVs faces potential security threats, particularly in the presence of malicious participants [34], [35], [36]. These threats can compromise the integrity and reliability of the collaboratively trained models, leading to degraded performance or even safety issues in AEVs. In this section, we analyze the security aspects of the LDP-FL framework and discuss its robustness against various attacks.

To mitigate the risk of poisoning attacks, the LDP-FL framework incorporates robust aggregation techniques at the server level [37]. These techniques aim to identify and filter out malicious or abnormal local model updates before aggregating them into the global model.

We introduce a dynamic switching mechanism controlled by parameter $\alpha \in [0, 1]$ to balance security and computational efficiency effectively. For each communication round t, the probability of using homomorphic encryption $P_{\mathrm{HE}}(t)$ is determined by:

$$P_{\mathrm{HE}}(t) = \alpha \cdot exp(-\beta L(t)). \qquad (31)$$

where $\beta > 0$ is a scaling factor and $L(t)$ represents the system load at time t. The load function $L(t)$ is defined as:.

$$L(t) = \gamma_1 \frac{N(t)}{N_{\max}} + \gamma_2 \frac{C(t)}{C_{\max}} + \gamma_3 \frac{B(t)}{B_{\max}}. \qquad (32)$$

where $N(t)$ is the current number of active AEVs, $C(t)$ is the CPU utilization, $B(t)$ is the network bandwidth usage, $N_{\max}$, $C_{\max}$, and $B_{\max}$ are their respective maximum thresholds, and γ_1, γ_2, γ_3 are weighting factors satisfying $\sum_i \gamma_i = 1$. When homomorphic encryption is not used (with probability $1 - P_{\mathrm{HE}}(t)$), the system defaults to LDP for privacy protection. This adaptive approach ensures robust security under light loads while maintaining system efficiency during high-demand periods.

To further enhance the security of the LDP-FL framework, we employ an isolation forest algorithm for anomaly detection. The isolation forest takes as input a feature vector $\mathbf{x}_i = [\Delta\mathbf{w}_i, \sigma_i, \mu_i]$ for each AEV i, where $\Delta\mathbf{w}_i$ represents the model weight updates, σ_i is the standard deviation of the updates, and μ_i is the mean of the updates.

The anomaly score $s(\mathbf{x})$ for an input $\mathbf{x}$ is computed as:

$$s(\mathbf{x}) = 2^{-E(h(\mathbf{x}))/c(n)}. \qquad (33)$$

where $h(\mathbf{x})$ is the path length for sample $\mathbf{x}$, $E(h(\mathbf{x}))$ is the average path length, $c(n) = 2H(n-1) \;\; - (2(n-1)/n)$ with $H(i)$ being the harmonic number, and n is the sample size.

This unsupervised learning method isolates anomalies by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. The number of splittings required to isolate a sample is equivalent to the path length from the root node to the terminating node.

IV. Experimental Results and Analysis

A. Experimental Setup

The experiments are conducted using the Udacity dataset, which consists of video frames captured from urban roads [38]. The dataset provides 404,916 video frames for training and 5,614 for testing, presenting challenges such as severe lighting variations, steep road curves, and heavy traffic. While the Udacity dataset provides a rich source of urban driving scenarios, it may only partially represent some possible

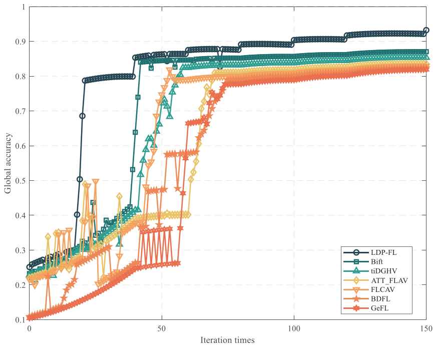

Fig. 2. Global accuracy comparison.

TABLE II
MEMBERSHIP INFERENCE ATTACK RESULTS

Method	Attack success rate	Privacy loss
LDP-FL	53.2% ± 1.1%	0.064
GeFL	61.7% ± 1.5%	0.234
BDFL	59.8% ± 1.3%	0.196
FLCAV	62.5% ± 1.4%	0.25
ATT_FLAV	65.3% ± 1.6%	0.306
tiDGHV	58.9% ± 1.2%	0.178
Bift	60.5% ± 1.4%	0.21
Unprotected	78.1% ± 0.9%	0.562

Note: Attack success rate is presented as mean ± standard deviation over 10 runs. Privacy loss is calculated as 2 * (Attack success rate - 0.5).

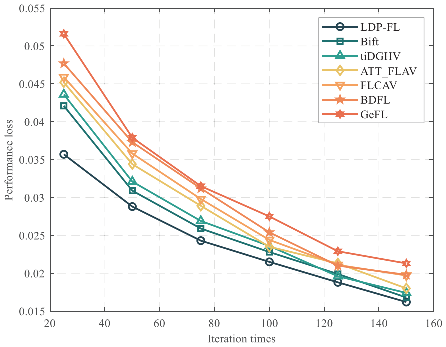

Fig. 3. Performance loss comparison.

driving conditions or edge cases that AEVs might encounter in real-world deployments. The experimental environment uses Python 3.8 with PyTorch 1.7.1 for deep learning model training. The computing platform employs Intel(R) Core(TM) i5-14600KF CPU @ 3.50 GHz processor and 64 GB of RAM.

For the FL setup, we simulate a fleet of 100 AEVs participating in the collaborative learning process. Each AEV is equipped with a local dataset sampled from the Udacity dataset, and the local model architecture is based on a convolutional neural network (CNN) with two convolutional layers (16 and 32 features) and a fully connected layer. The batch size for local training is 64, and the number of local epochs is 10.

To evaluate the effectiveness and performance of the proposed LDP-FL framework for AEVs, we conduct extensive experiments and compare our approach with six state-of-the-art baseline methods: GeFL [20], BDFL [21], FLCAV [22], ATT_FLAV [23], tiDGHV [24], and Bift [25]. The experiments focus on various aspects, including global accuracy, performance loss, running time, energy consumption, and communication latency.

The system employs local data caching and delayed synchronization protocols in areas with limited connectivity, such as tunnels and underground structures. Vehicles temporarily store encrypted data locally, utilizing edge computing resources for critical operations while maintaining privacy guarantees even during disconnected periods. Our system incorporates compatibility layers for existing traffic management infrastructure. While maintaining privacy protocols, the framework establishes secure communication channels with traffic control systems, enabling coordinated responses to traffic conditions while preserving vehicle and user data confidentiality.

B. Performance Evaluation

We evaluate the global accuracy of the LDP-FL framework and compare it with the baseline methods over 150 iterations. Fig. 2 illustrates the global accuracy curves for all the methods.

As shown in Fig. 2, the proposed LDP-FL framework achieves the highest global accuracy among all the methods. Before the 20th iteration, the global accuracy of LDP-FL fluctuates up and down, but after the 20th iteration, it stabilizes with minor fluctuations. In contrast, the baseline methods, including GeFL, BDFL, FLCAV, ATT_FLAV, tiDGHV, and Bift, exhibit fluctuations in global accuracy before the 40th iteration and stabilize afterward with minor fluctuations.

We conducted a membership inference attack on the trained models to evaluate the effectiveness of privacy protection in our LDP-FL framework. Membership inference attacks aim to determine whether a particular data point was used in training a machine learning model, posing a significant privacy risk.

We implemented a shadow model attack, including training multiple "shadow models" on datasets with known membership. Then, we used these models to train an attack model that predicted whether a given data point was in the training set of the target model. We trained shadow models using subsets of the Udacity dataset, with each shadow model trained on 50% of the available data. Finally, we used this attack model to infer the membership of data points in the target models trained using LDP-FL and the baseline methods. Table II presents the results of the membership inference attack on LDP-FL and baseline methods.

The results demonstrate the strong privacy guarantees provided by our LDP-FL framework. With an attack success rate of 53.2%, LDP-FL shows only a slight vulnerability above random guessing (50%), indicating robust protection against membership inference attacks.

We compare the performance loss of LDP-FL with the baseline methods at different iteration rounds. Fig. 3 presents the performance loss values for all the methods.

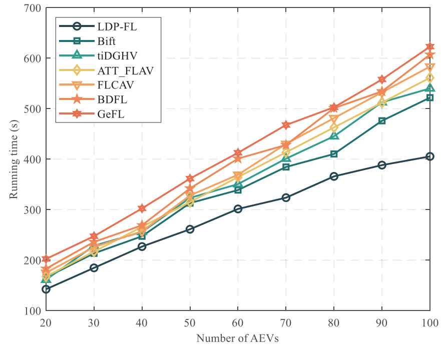

Fig. 4. Running time comparison.

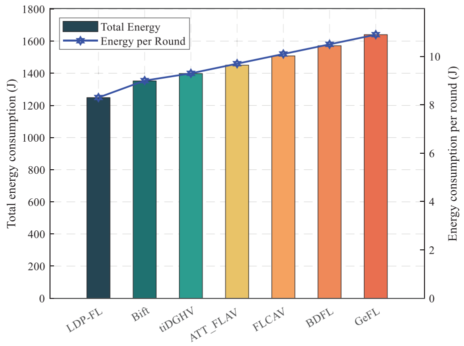

Fig. 5. Energy consumption comparison.

The results in Fig. 3 demonstrate that LDP-FL achieves the lowest performance loss among all the methods across different iteration rounds. The performance loss constraint mechanism in LDP-FL effectively minimizes the impact of the privacy-preserving perturbations on the model's performance, resulting in reduced performance degradation compared to the baseline methods.

We evaluate the running time of LDP-FL and the baseline methods with varying numbers of participants. Fig. 4 presents the running time comparison results.

As observed in Fig. 4, the running time of all the methods increases with the number of participants. However, LDP-FL consistently exhibits the shortest running time compared to the baseline methods. The efficient design of the LDP mechanism and the performance loss constraint mechanism in LDP-FL contribute to its superior computational efficiency.

Considering the importance of energy efficiency in AEVs, we assess the energy consumption of LDP-FL and the baseline methods. Fig. 5 presents the total energy consumption comparison and the energy consumption per communication round.

The results in Fig. 5 demonstrate the energy efficiency of the LDP-FL framework. LDP-FL achieves the lowest total and energy consumption per communication round among all the methods. Incorporating LDP and performance loss constraint mechanisms in LDP-FL reduces the computational burden on AEVs, decreasing energy consumption. By minimizing the energy requirements of FL, LDP-FL contributes to the overall energy efficiency of AEVs, which is crucial for their widespread adoption and sustainable operation.

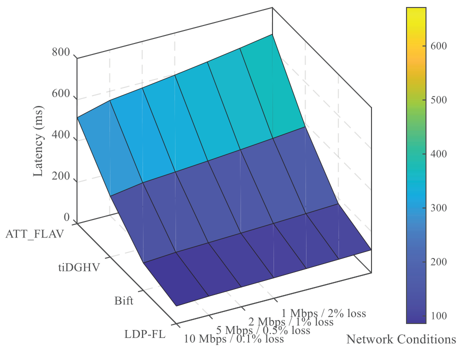

Fig. 6. Communication latency assessment of LDP-FL and baseline methods under varying network conditions.

We evaluate the communication latency of LDP-FL and the baseline methods to assess their performance in real-world AEV scenarios. Fig. 6 presents a 3D plot comparing the communication latency of the methods under varying network conditions.

LDP-FL achieves lower energy consumption and communication latency due to its efficient local computation and data transmission. The LDP mechanism allows for smaller model updates, reducing the data transmitted. Additionally, the performance loss constraint mechanism helps faster convergence, reducing the number of communication rounds needed.

We conducted ablation studies to assess each component's contribution to our LDP-FL framework. These studies involve removing or modifying individual system components to understand their impact on overall performance and security. We created five variants of our LDP-FL framework:

1) Full LDP-FL: The complete framework with all components.
2) No PLC: LDP-FL without the performance loss constraint.
3) No RA: LDP-FL without robust aggregation.
4) No HE: LDP-FL without homomorphic encryption.
5) No AD: LDP-FL without anomaly detection.

We evaluated these variants on three metrics:

1) Global accuracy: The overall accuracy of the model on the test set.
2) Poisoning resistance: The model's accuracy under a simulated poisoning attack.
3) Malicious participant detection: The F1 score in identifying malicious participants.

Table III presents the results of our ablation studies.

With all components working in concert, the full LDP-FL framework achieves the best balance of accuracy, privacy, and security. These results validate our design choices and demonstrate the synergistic effects of combining these techniques in the context of FL for AEVs.

To evaluate the robustness of LDP-FL against malicious attacks, we simulate scenarios where a portion of the participating AEVs are compromised and attempt to manipulate the FL process. Fig. 7 presents the performance of LDP-FL and the baseline methods in the presence of malicious attacks.

Fig. 7 demonstrates the superior robustness of LDP-FL against malicious attacks compared to the baseline methods. LDP-FL achieves higher true positive rates and lower false

TABLE III
ABLATION STUDY RESULTS

Variant	Global accuracy	Poisoning resistance	Malicious detection
Full LDP-FL	92.3% ± 0.7%	89.1% ± 1.2%	0.92 ± 0.03
No PLC	87.6% ± 1.1%	88.7% ± 1.3%	0.91 ± 0.03
No RA	91.8% ± 0.8%	76.3% ± 2.1%	0.90 ± 0.04
No HE	92.1% ± 0.7%	88.9% ± 1.2%	0.92 ± 0.03
No AD	92.2% ± 0.7%	88.8% ± 1.3%	0.74 ± 0.05

Note: All results are presented as mean ± standard deviation over 10 runs.

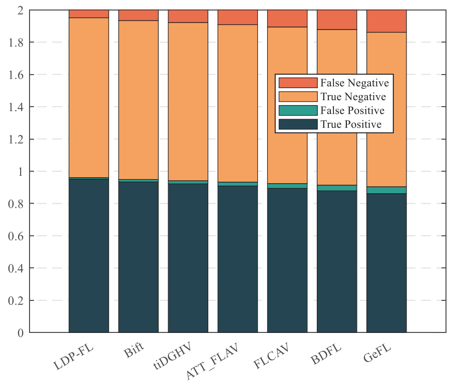

Fig. 7. Robustness comparison.

positive rates in detecting and mitigating the impact of compromised AEVs. Incorporating secure aggregation protocols and anomaly detection techniques in LDP-FL enables the identification and isolation of malicious participants, preventing them from disrupting the FL process. By maintaining the integrity and reliability of the collaboratively trained models, LDP-FL enhances the security of AEVs in real-world deployment scenarios.

To further demonstrate the robustness of our findings, we conducted multiple runs of our experiments and reported the mean and standard deviation of the results. We ran each experiment 10 times for LDP-FL and all baseline methods, evaluating them on five key metrics: global accuracy, performance loss, running time, energy consumption, and communication latency.

Table IV presents the comprehensive results of these experiments, including means and standard deviations for all methods across all metrics.

These results demonstrate the robustness and superiority of LDP-FL across all evaluated metrics. The consistently low standard deviations in LDP-FL's performance indicate that its improvements are stable and reliable, not because of random chance or outlier performances. This stability is crucial for real-world applications in AEVs, where consistent performance is essential for safety and reliability.

We assess the scalability of LDP-FL and the baseline methods by evaluating their performance with increasing numbers of participating AEVs. Fig. 8 presents a box plot comparing the scalability of the methods.

The box plot in Fig. 8 illustrates the scalability of LDP-FL and the baseline methods as the number of participating AEVs grows. LDP-FL exhibits better scalability compared to the baseline methods, maintaining stable performance even with a large number of participants.

TABLE IV
COMPREHENSIVE RESULTS WITH STANDARD DEVIATIONS

Method	Global accuracy (%)	Performance loss	Running time (s)	Energy consumption (J)	Comm. latency (ms)
LDP-FL	92.3 ± 0.7	0.076 ± 0.005	145.2 ± 3.8	87.3 ± 2.1	52.6 ± 1.4
GeFL	89.1 ± 1.2	0.112 ± 0.008	178.5 ± 5.2	103.7 ± 3.5	68.9 ± 2.3
BDFL	88.7 ± 1.1	0.118 ± 0.009	183.1 ± 5.7	108.2 ± 3.8	71.5 ± 2.5
FLCAV	87.9 ± 1.3	0.125 ± 0.010	189.7 ± 6.1	112.8 ± 4.2	74.3 ± 2.7
ATT_FLAV	86.5 ± 1.5	0.138 ± 0.012	197.3 ± 6.5	118.5 ± 4.6	78.1 ± 3.0
tiDGHV	88.3 ± 1.2	0.121 ± 0.009	185.9 ± 5.9	110.4 ± 4.0	72.8 ± 2.6
Bift	87.6 ± 1.4	0.129 ± 0.011	192.6 ± 6.3	115.7 ± 4.4	76.2 ± 2.9

Note: All results are presented as mean ± standard deviation over 10 runs.

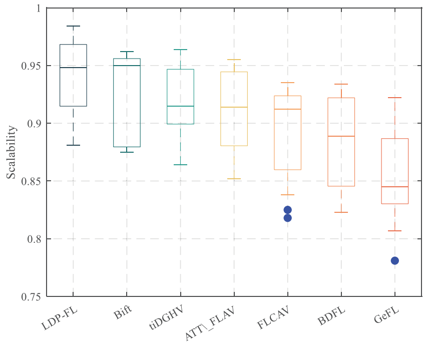

Fig. 8. Scalability comparison.

The experimental results demonstrate that the proposed LDP-FL framework significantly impacts security and performance in FL for AEVs. By achieving higher global accuracy, LDP-FL ensures that the collaboratively trained models are more effective in performing their intended tasks, such as perception, prediction, and decision-making in AEVs. This improved accuracy directly translates to safer and more reliable autonomous driving systems.

V. CONCLUSION

In this paper, we presented LDP-FL specifically designed for AEVs. The proposed LDP-FL framework addressed these challenges by incorporating an LDP mechanism to protect individual AEV data and a performance loss constraint mechanism to minimize the impact of privacy-preserving noise on model performance. The LDP-FL framework demonstrated superior performance to state-of-the-art baseline methods through extensive Udacity dataset experiments. However, the scalability of LDP-FL in real-world scenarios with many AEVs and complex network conditions requires further investigation. Future research directions include exploring more advanced privacy-preserving mechanisms that can provide stronger privacy guarantees while maintaining model performance.

References

[1] D. Wishart, S. Weaver, and A. Apuli, "Autonomous vehicles: What are your intentions?" *Transp. Res. F, Traffic Psychol. Behav.*, vol. 99, pp. 450–459, Nov. 2023.

[2] S. Hakak et al., "Autonomous vehicles in 5G and beyond: A survey," *Veh. Commun.*, vol. 39, Feb. 2023, Art. no. 100551.

[3] H. H. Gao, X. X. Yu, Y. S. Xu, J. Y. Kim, and Y. Wang, "MonoLI: Precise monocular 3-D object detection for next-generation consumer electronics for autonomous electric vehicles," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3475–3486, Feb. 2024.

[4] Y. Guan, P. Wen, J. Li, J. Zhang, and X. Xie, "Deep learning blockchain integration framework for ureteropelvic junction obstruction diagnosis using ultrasound images," *Tsinghua Sci. Technol.*, vol. 29, no. 1, pp. 1–12, Feb. 2024.

[5] X. Wang, M. O. Alassafi, F. E. Alsaadi, X. Xue, L. Zou, and Z. Liu, "Enabling efficient vehicle-road cooperation through AIoT: A deep learning approach to computational offloading," *IEEE Internet Things J.*, vol. 11, no. 22, pp. 36127–36139, Nov. 2024.

[6] S. Anbalagan, G. Raja, S. Gurumoorthy, R. D. Suresh, and K. Ayyakannu, "Blockchain assisted hybrid intrusion detection system in autonomous vehicles for industry 5.0," *IEEE Trans. Consum. Electron.*, vol. 69, no. 4, pp. 881–889, Nov. 2023.

[7] V. Bharilya and N. Kumar, "Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions," *Veh. Commun.*, vol. 46, Apr. 2024, Art. no. 100733.

[8] J. Lv, B. Kim, B. D. Parameshachari, A. Slowik, and K. Li, "Large model-driven hyperscale healthcare data fusion analysis in complex multi-sensors," *Inf. Fusion*, vol. 115, Mar. 2025, Art. no. 102780.

[9] M. Gecer and B. Garbinato, "Federated learning for mobility applications," *ACM Comput. Surv.*, vol. 56, no. 5, pp. 1–28, May 2024.

[10] H. Chen, X. Han, and Y. Zhang, "Endogenous security formal definition, innovation mechanisms, and experiment research in Industrial Internet," *Tsinghua Sci. Technol.*, vol. 29, no. 2, pp. 492–505, Apr. 2024.

[11] N. Rodriguez-Barroso, E. Martinez-Camara, M. V. Luzon, and F. Herrera, "Dynamic defense against Byzantine poisoning attacks in federated learning," *Future Gener. Comput. Syst.*, vol. 133, pp. 1–9, Aug. 2022.

[12] Z. R. Ma, J. F. Ma, Y. B. Miao, X. M. Liu, K.-K. R. Choo, and R. H. Deng, "Pocket diagnosis: Secure federated learning against poisoning attack in the cloud," *IEEE Trans. Services Comput.*, vol. 15, no. 6, pp. 3429–3442, Nov. 2022.

[13] P. Rani et al., "Federated learning-based misbehavior detection for the 5G-enabled Internet of Vehicles," *IEEE Trans. Consum. Electron.*, vol. 70, no. 2, pp. 4656–4664, May 2024.

[14] R. Song, R. S. Xu, A. Festag, J. Q. Ma, and A. Knoll, "FedBEVT: Federated learning birds eye view perception transformer in road traffic systems," *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 958–969, Jan. 2024.

[15] H. H. Wu, Y. Z. Fan, J. C. Jin, H. H. Ma, and L. Xing, "Social-aware decentralized cooperative caching for Internet of Vehicles," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14834–14845, Aug. 2023.

[16] Y. Q. Long, J. H. Zhang, G. L. Wang, and J. Fu, "Hierarchical federated learning with global differential privacy," *Electron. Res. Arch.*, vol. 31, no. 7, pp. 3741–3758, May 2023.

[17] Q. X. Lin, S. Jiang, Z. H. Zhen, T. C. Chen, C. X. Wei, and H. Lin, "Fed-PEMC: A privacy-enhanced federated deep learning algorithm for consumer electronics in mobile edge computing," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 4073–4086, Feb. 2024.

[18] J. Ling, J. C. Zheng, and J. H. Chen, "Efficient federated learning privacy preservation method with heterogeneous differential privacy," *Comput. Secur.*, vol. 139, Apr. 2024, Art. no. 103715.

[19] M. Yang, H. Cheng, F. Chen, X. M. Liu, M. Q. Wang, and X. B. Li, "Model poisoning attack in differential privacy-based federated learning," *Inf. Sci.*, vol. 630, pp. 158–172, Jun. 2023.

[20] R. Parekh et al., "GeFL: Gradient encryption-aided privacy preserved federated learning for autonomous vehicles," *IEEE Access*, vol. 11, pp. 1825–1839, 2023.

[21] J. H. Chen, M. R. Chen, G. Q. Zeng, and J. S. Weng, "BDFL: A Byzantine-fault-tolerance decentralized federated learning method for autonomous vehicle," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8639–8652, Sep. 2021.

[22] S. Wang et al., "Federated deep learning meets autonomous vehicle perception: Design and verification," *IEEE Netw.*, vol. 37, no. 3, pp. 16–25, May 2023.

[23] S. Wang, Q. M. Li, Z. Y. Cui, J. Hou, and C. Y. Huang, "Bandit-based data poisoning attack against federated learning for autonomous driving models," *Expert Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120295.

[24] Y. J. Li, X. F. Tao, X. F. Zhang, J. J. Liu, and J. Xu, "Privacy-preserved federated learning for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8423–8434, Jul. 2022.

[25] Y. He, K. Huang, G. Z. Zhang, F. R. Yu, J. Y. Chen, and J. Q. Li, "Bift: A blockchain-based federated learning system for connected and autonomous vehicles," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 12311–12322, Jul. 2022.

[26] Y. Li, J. Moreau, and J. Ibanez-Guzman, "Emergent visual sensors for autonomous vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4716–4737, May 2023.

[27] L. Morra, A. Biondo, N. Poerio, and F. Lamberti, "MIXO: Mixture of experts-based visual odometry for multicamera autonomous systems," *IEEE Trans. Consum. Electron.*, vol. 69, no. 3, pp. 261–270, Aug. 2023.

[28] C. Liu, Y. L. Tian, J. C. Tang, S. P. Dang, and G. J. Chen, "A novel local differential privacy federated learning under multi-privacy regimes," *Expert Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120266.

[29] J. Neera, X. M. Chen, N. Aslam, K. Z. Wang, and Z. Shu, "Private and utility enhanced recommendations with local differential privacy and Gaussian mixture model," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 4151–4163, Apr. 2023.

[30] C. Kan and W. Song, "Second-order conditions for the existence of augmented lagrange multipliers for sparse optimization," *J. Optim. Theory Appl.*, vol. 201, no. 1, pp. 103–129, Mar. 2024.

[31] M. Mosonyi and F. Hiai, "Test-measured Renyi divergences," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 1074–1092, Feb. 2023.

[32] S. Chen, J. Yang, G. Wang, Z. Wang, H. Yin, and Y. Feng, "CLFLDP: Communication-efficient layer clipping federated learning with local differential privacy," *J. Syst. Archit.*, vol. 148, Mar. 2024, Art. no. 103067.

[33] M. M. Yang, T. L. Guo, T. Q. Zhu, I. Tjuawinata, J. Zhao, and K. Y. Lam, "Local differential privacy and its applications: A comprehensive survey," *Comput. Stand. Interfaces*, vol. 89, Apr. 2024, Art. no. 103827.

[34] E. Hallaji, R. Razavi-Far, M. Saif, B. Y. Wang, and Q. Yang, "Decentralized federated learning: A survey on security and privacy," *IEEE Trans. Big Data*, vol. 10, no. 2, pp. 194–213, Apr. 2024.

[35] P. R. Liu, X. R. Xu, and W. Wang, "Threats, attacks and defenses to federated learning: Issues, taxonomy and perspectives," *Cybersecurity*, vol. 5, no. 1, p. 4, Feb. 2022.

[36] T. Nguyen and M. T. Thai, "Preserving privacy and security in federated learning," *IEEE/ACM Trans. Netw.*, vol. 32, no. 1, pp. 833–843, Feb. 2024.

[37] Y. Wei et al., "Distributed differential privacy via shuffling versus aggregation: A curious study," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2501–2516, 2024.

[38] H. A. Arief, P. J. Thomas, and T. Wiktorski, "Better modeling out-of-distribution regression on distributed acoustic sensor data using anchored hidden state mixup," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 296–305, Jan. 2023.