Over-the-Air Federated Learning With Joint Privacy-Accuracy Optimization

Hexin Feng, *Student Member, IEEE*, Rui Wang, *Senior Member, IEEE*, Erwu Liu, *Senior Member, IEEE*, Wei Ni, *Fellow, IEEE*, Dusit Niyato, *Fellow, IEEE*, and Abbas Jamalipour, *Fellow, IEEE*

***Abstract*—Federated learning (FL) contributes to data privacy by not disclosing raw data, but encounters challenges of privacy leakage from local gradient uploading. This paper introduces a novel over-the-air computation (AirComp)-based FL system that balances privacy and accuracy by leveraging the waveform superposition and channel propagation characteristics of AirComp. Specifically, we derive the privacy leakage metric to explicitly account for the effects of waveform aggregation and communication noise. We analyze the convergence upper bound to capture model update errors stemming from artificial and communication noise. We formulate a new joint privacy-accuracy optimization problem by incorporating privacy leakage in the model training objective, guiding the learning process towards enhanced privacy protection. We then employ convex optimization techniques to derive the optimal power scaling and artificial noise intensity. Simulations demonstrate up to 80% reduction in privacy leakage compared to baselines under stringent privacy constraints, while maintaining competitive learning performance. Our method exhibits enhanced robustness under low signal-to-noise ratios, achieving 40% lower privacy leakage under equivalent privacy budgets.**

***Index Terms*—Rényi differential privacy, over-the-air computation, federated learning.**

TABLE I
NOTATION AND DEFINITIONS

Notation	Definition
c_t	Uniform power scaling factor at the learning round t
$f_k(\boldsymbol{w}_t)$	The local loss function of user k
$F(\boldsymbol{w})$	The global loss function
$\boldsymbol{g}_k(\boldsymbol{w}_t)$	The local gradient of user k
$\boldsymbol{h}_{k,t}$	Channel coefficient from user k at the learning round t
M	The number of users
$\boldsymbol{m}_t$	The wireless channel noise
$\boldsymbol{w}_t$	The model parameters
C	The bound of gradients
D_m	Size of local dataset of the m-th user
E	Number of local updates
K	Number of selected users
$\mathcal{K}_t$	A subset of selected users
T	Total number of learning rounds
α	Order of the probability
α_k	The transmit scalar of user k
δ	The fault-tolerant probability
ϵ	The privacy budget
η	Server learning rate
η_t	Local user learning rate
$\theta_{m,t}$	The evaluation parameter of user m

I. INTRODUCTION

FEDERATED learning (FL) enables multiple local users to train machine learning (ML) models collaboratively. It is a functional approach for applications that require low data processing latency and involve many participants to exchange information, e.g., image classification [1], industrial Internet of Things [2], and autonomous vehicles [3]. Despite its advantages, FL faces obstacles, including communication bottlenecks, security, and privacy leakage [4], [5]. To accomplish efficient model aggregation, the authors of [6] proposed over-the-air computation (AirComp) to make users transmit their local updates concurrently in an uncoded manner by taking advantage of the waveform superposition property [7], [8], [9], merging computation and communication and resulting in low transmission latency and excellent spectral efficiency. Moreover, the bandwidth requirements or communication latency of wireless computation do not rise with the number of users, unlike traditional orthogonal multiple access (OMA). This significantly reduces the communication bottleneck of FL.

A. Motivation and Challenges

Recent advances in AirComp-FL have demonstrated significant potential, but the vulnerability of FL systems to gradient inversion attacks remains a critical privacy concern that has not been adequately addressed in AirComp-FL [10]. While FL algorithms aim to protect privacy, current research has revealed that FL may not guarantee privacy and robustness and remains vulnerable to the leakage of private training data, resulting in leakage [11], [12]. The following attacks can be launched against current FL protocol designs: i) attacks from

Received 8 January 2025; revised 4 August 2025; accepted 28 September 2025. Date of publication 10 October 2025; date of current version 28 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62271352 and Grant 42171404 and in part by the Fundamental Research Funds for the Central Universities under Grant 22120250094. The associate editor coordinating the review of this article and approving it for publication was Dr. Zhipeng Cai. *(Corresponding author: Rui Wang.)*

Hexin Feng is with the College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China (e-mail: 2310200@tongji.edu.cn).

Rui Wang is with the College of Electronics and Information Engineering, Shanghai Institute of Intelligent Science and Technology, and the National College of Elite Engineers, Tongji University, Shanghai 201804, China (e-mail: ruiwang@tongji.edu.cn).

Erwu Liu is with the College of Electronics and Information Engineering and the Department of Ophthalmology, Tongji Hospital, School of Medicine, Tongji University, Shanghai 201804, China (e-mail: erwu.liu@ieee.org).

Wei Ni is with the School of Engineering, Edith Cowan University, Perth, WA 6027, Australia, and also with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: wei.ni@ieee.org).

Dusit Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dniyato@ntu.edu.sg).

Abbas Jamalipour is with the School of Electrical and Computer Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: a.jamalipour@ieee.org).

Digital Object Identifier 10.1109/TIFS.2025.3620108

curious servers, in which the server tries to alter participants' perceptions of global parameters, manipulate the collaborative training process, or deduce private information from local updates during training; ii) attacks from adversaries, who try to alter and contaminate the global model parameter, or deduce private information about other participants [13]. The entire training procedure may reveal private information, potentially resulting in profound leakage. For example, adversaries can utilize aggregated gradients to rebuild users' training data [11], [12]. It was shown in [14] that even a small part of gradients could reveal sensitive local information.

Existing defense mechanisms for FL predominantly focus on traditional architectures, relying on explicit gradient perturbation [15], [16] or compression techniques [17]. While addressing privacy concerns, these methods often introduce additional communication overhead and performance trade-offs, particularly in wireless settings. Moreover, current FL privacy studies frequently abstract the communication process as an idealized digital channel, overlooking the privacy implications of physical-layer properties, e.g., the implementation of wireless FL in cell-free massive MIMO systems presents unique privacy challenges ascribed to the distributed nature of access points and users. The multi-point wireless transmission architecture inherently increases privacy vulnerabilities through multiple potential points of data interception. The system's scale and complexity make comprehensive privacy measures particularly challenging [4], [15].

Emerging research has begun to harness the intrinsic privacy advantages of AirComp to enhance system privacy [18]. These approaches leverage channel noise as a natural source of gradient perturbation, thereby integrating physical layer privacy-enhancing properties with privacy-preserving techniques while maintaining transmission efficiency [19]. *A significant research gap lies in the quantification and mitigation of privacy leakage of AirComp-FL systems in the presence of adversaries attempting to decompose the aggregated AirComp signals to restore individual gradients.* This privacy risk associated with AirComp-FL remains largely unexplored in the literature.

B. Related Work

Existing studies have integrated AirComp into FL to address the communication bottleneck of FL. The authors in [9] first presented FL via AirComp and demonstrated that AirComp reduces latency over broadband channels, compared to OMA. This method was later extended to one-bit AirComp in [20]. However, the process alters the direction of the gradient descent and would degrade learning performance. In [18], the learning performance of AirComp-FL was enhanced through simultaneous optimization of receive beamforming and device selection. The authors in [21] used gradient sparsification and compression with AirComp to decrease the dimensionality of exchanged model updates in fading channels and Gaussian channels. The authors in [22] jointly optimized the reconfigurable intelligent surface (RIS) and transmit powers to minimize the convergence upper bound of FL under imperfect aggregation. A recent work [23] pioneered distributed multi-modal foundation models for 6G networks, integrating pipeline parallelism with gradient compression and AirComp to overcome wireless bottlenecks while enabling cross-modal vision-language fusion.

To further improve privacy, encryption and obfuscation of shared gradients [24], [25] are two common approaches. With differential privacy (DP), a rigorous mathematical framework was proposed to quantify privacy disclosure [26]. Its efficiency and simplicity have led to its broader applications. To achieve DP, recent studies added random perturbations, e.g., Gaussian [26], Laplacian [27], and Binomial mechanisms [28], to gradients. To enjoy free privacy protection, the authors of [19] designed an FL method based on both OMA and non-orthogonal multiple access (NOMA) transmissions with AirComp. This was done by utilizing the intrinsic channel noise to enhance privacy. Later, the inherent anonymity of AirComp was demonstrated in [29], which contributes to privacy and minimizes artificial noise added to the local updates. For AirComp-FL, the authors in [30] achieved a superior privacy-accuracy trade-off by using an RIS.

The existing works [19], [29], [30] explored power allocation designs with unified noise scales for all users to balance the privacy-accuracy trade-off, but did not take into account the extra privacy leakage resulting from the regular exchange of the user's model updates. The privacy leakage of model updates was initially modeled statistically and verified experimentally in [31]. Later, the privacy leakage developed in [31] was quantified analytically in [32], adapting to different contexts and data sets. These works [19], [29], [30] overlooked a critical vulnerability: signal decomposition attacks, wherein a multi-antenna adversary employs advanced signal processing, e.g., zero-forcing (ZF) beamforming, to extract individual gradients from aggregated signals.

C. Contribution

This paper presents a new privacy-enhanced AirComp-FL (PEA-FL) framework that harnesses the waveform superposition and channel propagation properties of AirComp to strike a balance between model accuracy and privacy. Through rigorous analysis and system optimization, we demonstrate that PEA-FL can meet stringent privacy requirements and maintain model accuracy. The key contributions of this paper are summarized as follows.

- We are the first to reveal adversaries that are capable of using array signal detection techniques, e.g., zero-forcing, to decompose AirComp signals, restore individual gradients, and expose AirComp-FL to privacy risks.
- We propose PEA-FL, which leverages the waveform superposition property and inherent anonymity of AirComp, along with DP perturbation, to enhance communication efficiency and strengthen privacy.
- We rigorously derive the privacy leakage function that quantifies information exposure under the new decomposition attacks. The impact of the attacks is incorporated into the training objective of PEA-FL, enabling adaptive privacy budget allocation to resist the attacks.
- We derive the convergence upper bounds of PEA-FL by capturing model update errors caused by both artificial and communication noise. We utilize Rényi DP (RDP) to establish tight privacy bounds by analyzing the effective noise perturbations.

- We formulate a joint privacy-accuracy optimization problem to balance model training performance and privacy. Using convex optimization techniques, we derive the optimal uniform power scaling factor and artificial noise intensity, facilitating convergence while ensuring robust resistance against attacks.

Extensive simulations validate our system design, analysis, and algorithm, and indicate that the proposed PEA-FL framework significantly improves the privacy-accuracy trade-off over traditional methods. Under stringent privacy constraints, PEA-FL reduces privacy leakage by 80% compared to baselines while maintaining comparable system performance. Under the same privacy budget settings, PEA-FL demonstrates superior robustness under low signal-to-noise ratios (SNRs), reducing privacy leakage by 40%. Multi-antenna adversaries capable of array signal processing, e.g., ZF, represent a realistic and increasingly relevant threat in future wireless systems. This study provides the analytical tools and defense mechanisms to address this threat class, which has not been sufficiently covered in the existing literature.

The remainder of this paper is structured as follows. Section II outlines the FL model, DP method, and privacy threat. Section III delineates the proposed method for PEA-FL. In Section IV, we analyze the convergence upper bound, derive privacy constraints, present the privacy leakage function, and cast a new joint loss-leakage minimization problem. Section V provides numerical results. Section VI concludes the paper.

Notation. Italic, boldface lower-case, and boldface upper-case letters denote scalars, vectors, and matrices, respectively. $\|\boldsymbol{x}\|$ denotes its ℓ^2-norm of vector $\boldsymbol{x}$. $x[i]$ denotes the i-th entry of $\boldsymbol{x}$. $x[i,j]$ stands for the (i,j)-th entry of $\boldsymbol{X}$. $\mathcal{CN}(\mu,\sigma^2)$ denotes the circularly-symmetric complex Gaussian distribution with mean μ and covariance σ^2. $|\cdot|$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ stand for cardinality, transpose, conjugate transpose, and inverse, respectively. $\mathbb{E}[\cdot]$ takes expectation. $\langle \boldsymbol{a}, \boldsymbol{b}\rangle$ denotes the inner product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex values, respectively. $\mathbb{C}^{x\times y}$ is the space of $x\times y$ complex matrices. $\mathbf{1}$, $\boldsymbol{I}$, and $\mathbf{0}$ denote the all-one column vector, identity matrix, and all-zero matrix, respectively.

II. System Model

This section describes wireless FL and its privacy threat model, and introduces the concepts of DP and Rényi DP, which are adopted to defend against the privacy threat model.

A. Federated Learning (FL) Model

The considered AirComp-FL system involves a single-antenna server, M single-antenna users, and a multiple-antenna adversary, see Fig. 1. This setting reflects lightweight edge computing platforms, e.g., IoT gateways [2], where physical or cost constraints limit the number of antennas installed at a base station with the servers. Notably, the method proposed in this paper is not restricted to single-antenna servers, and can extend to the case of multi-antenna servers.

Let $\mathcal{D}_m$ denote the local dataset of the m-th user with size $|\mathcal{D}_m| = D_m$, $m \in \{1,\ldots,M\}$. There are a total of T training rounds in the system. The channels remain unchanged during a training round. Before each training round, K users are

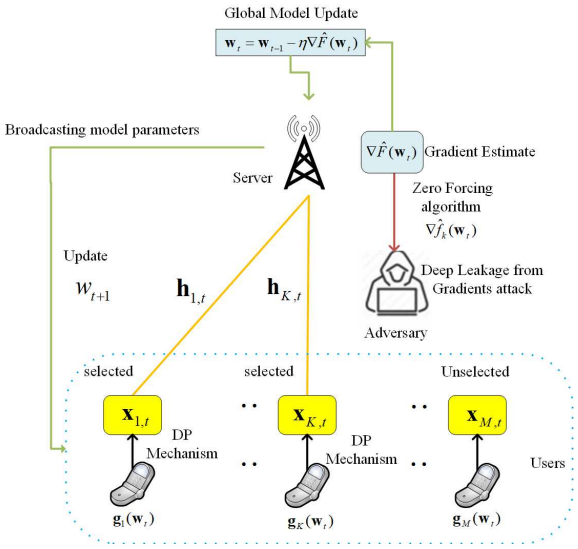


Fig. 1. The wireless FL system under DLG attacks.

selected to participate in training. In the training round t, each selected user k, $k \in \mathcal{K}_t$, where $\mathcal{K}_t$ is the set of selected users for training round t, performs multiple local gradient descent steps to update the local model parameters based on its dataset $\mathcal{D}_k$ with size $|\mathcal{D}_k| = D_k$. The users adopt the stochastic gradient descent to minimize the local loss function $f_k(\boldsymbol{w}_t)$ with respect to the model parameter $\boldsymbol{w}_t \in \mathbb{R}^I$, as given by

$$f_k(\boldsymbol{w}_t) = \frac{1}{D_k}\sum_{(\boldsymbol{x},y)\in\mathcal{D}_k} f_k(\boldsymbol{w}_t;\boldsymbol{x},y), \tag{1}$$

where $f_k(\boldsymbol{w}_t;\boldsymbol{x},y)$ is the local loss function that measures the prediction error of $\boldsymbol{w}_t$ on the input feature $\boldsymbol{x}$ concerning its label y in $\mathcal{D}_k$. For instance, $f_k(\boldsymbol{w}_t)$ is often the cross-entropy loss in image classification tasks [33]. Suppose that the users have equal data sizes [9], i.e., $D_k = D$, $k \in \mathcal{K}_t$. The goal of FL can be written as

$$F(\boldsymbol{w}_t) = \sum_{k\in\mathcal{K}_t} \frac{D_k f_k(\boldsymbol{w}_t)}{\sum_{k\in\mathcal{K}_t} D_k} = \frac{1}{K}\sum_{k\in\mathcal{K}_t} f_k(\boldsymbol{w}_t). \tag{2}$$

By minimizing $f_k(\boldsymbol{w}_t)$, each user k updates and uploads its local model, $\boldsymbol{g}_k(\boldsymbol{w}_t) \in \mathbb{R}^I$, as given by

$$\boldsymbol{g}_k(\boldsymbol{w}_t) = \nabla f_k(\boldsymbol{w}_t) = -\eta_t \sum_{e=1}^{E} \nabla f_k(\boldsymbol{w}_{t,e}^k), \tag{3}$$

where η_t is the local learning rate of user k.

Each user's model update includes the gradients of a batch of images. The server receives the gradients aggregated over-the-air from the selected users to solve the following problem:

$$\boldsymbol{w}_t^* = \arg\min_{\boldsymbol{w}_t} F(\boldsymbol{w}_t). \tag{4}$$

The server uses the gradients aggregated over-the-air to update the global model with the global learning rate.

B. Privacy Threat Model

Assume that the server is "honest-but-curious," i.e., honestly implementing the proposed method while remaining curious

about the training data of the local models [30]. Also, assume that an adversary can launch the Deep Leakage from Gradients (DLG) attack to recover the users' training data from their uploaded gradients. This models external, potentially compromised entities with advanced signal processing capabilities, e.g., rogue base stations or surveillance devices [32].

As described in [11], the local gradients can be recovered pixel-wise accurately for images at the adversary by minimizing the gradient match loss, i.e., the distance between the real and fake loss function gradients, i.e., $\nabla f(\boldsymbol{w})$ and $\nabla f(\boldsymbol{w}')$. Specifically, the adversary generates dummy data for the inputs $\boldsymbol{x}'$ and labels y', and uses the dummy data to generate training gradients. The gradient match loss is given by

$$\left\{\boldsymbol{x}'^*, y'^*\right\} = \min_{\boldsymbol{x}', y'} \left\|\nabla f(\boldsymbol{w}') - \nabla f(\boldsymbol{w})\right\|^2, \tag{5}$$

where $\boldsymbol{x}'^*$ and y'^* stand for training images and labels, respectively. The adversary minimizes the gradient match loss using the standard gradient-based methods to produce the dummy inputs $\boldsymbol{x}'$ and labels y'. The adversary obtains the training images $\boldsymbol{x}'^*$, nearly identical to the real images $\boldsymbol{x}$, as the gradient match loss approaches zero.

C. Rényi DP

To protect users' data privacy, we introduce the DP method, where the users utilize perturbation mechanisms on their gradients when transmitting the gradients. Hence, the server cannot precisely recover the gradients from the received signals. The traditional DP, involving two parameters $\epsilon > 0$ and $\delta \in (0, 1)$, is defined as follows.

Definition 1: ((ϵ, δ)-DP): A randomized mechanism $\mathcal{M} : S \to \mathbb{R}$ with domain S and range $\mathbb{R}$ satisfies (ϵ, δ)-DP, if any neighboring dataset $\mathcal{A}, \mathcal{A}' \in S$, and $O \in \mathbb{R}$

$$\Pr(\mathcal{M}(\mathcal{A}) \in O) \leq e^{\epsilon} \Pr(\mathcal{M}(\mathcal{A}') \in O) + \delta. \tag{6}$$

RDP is an expansion of DP with a parameter α. As $\alpha \to \infty$, RDP is equivalent to DP [34]. RDP facilitates analyzing composite heterogeneous mechanisms closely and conveying precise guarantees on the tails of privacy loss. A relaxation of DP based on the Rényi divergence was suggested in [34].

Definition 2: (Rényi Divergence): For two probability distributions $\mathcal{P}$ and $\mathcal{Q}$ defined over $\mathbb{R}$, the Rényi divergence of orders $\alpha > 1$ is given by

$$D_\alpha(\mathcal{P}\|\mathcal{Q}) = \frac{1}{\alpha - 1} \log \mathrm{E}_{x\sim\mathcal{Q}} \left(\frac{\mathcal{P}(x)}{\mathcal{Q}(x)}\right)^{\alpha}, \tag{7}$$

where all logarithms are taken to the base of the natural number, and $\mathcal{P}(x)$ denotes the probability density of $\mathcal{P}$ at x.

The sensitivity Δs of a function reflects the difference in the corresponding outputs for any neighboring inputs $\mathcal{A}, \mathcal{A}' \in S$, which is defined as follows.

Definition 3: (Sensitivity Δs): For the function $f : S \to \mathbb{R}$, its sensitivity is defined as

$$\Delta s = \max_{\mathcal{A}, \mathcal{A}' \in S} \left\|f(\mathcal{A}) - f(\mathcal{A}')\right\|. \tag{8}$$

The relationship between the Rényi divergence with α and the privacy budget ϵ is defined as follows.

Definition 4: (Gaussian Mechanism-Based RDP): Given $\alpha > 1$ and $\epsilon \geq 0$, the Gaussian mechanism for approximating $\mathcal{M}$ is defined as $\mathcal{M}(\mathcal{A}) + N_0$, where $N_0 \sim \mathcal{N}(0, \sigma^2)$ is the normal distribution with zero mean and deviation σ^2. The Rényi divergence between the Gaussian mechanism and its offset for any neighboring inputs $\mathcal{A}, \mathcal{A}' \in S$ is defined as

$$D_\alpha(\mathcal{M}(\mathcal{A})\|\mathcal{M}(\mathcal{A}')) \leq \alpha(\Delta s)^2/(2\sigma^2) \leq \epsilon. \tag{9}$$

RDP can be interpreted as $(\epsilon(\alpha), \delta)$-DP, with $\epsilon(\alpha) = \epsilon + \frac{\log(1/\delta)}{\alpha - 1}$. The parameter α directly influences the required noise variance as $\sigma^2 \geq \frac{\alpha(\Delta s)^2}{2\left(\epsilon(\alpha) - \frac{\log(1/\delta)}{\alpha-1}\right)}$, as shown in [34].

III. Proposed Privacy-Enhanced AirComp-FL

This section outlines the proposed PEA-FL. We consider a widely used FL configuration [35], in which the size of the local dataset $\mathcal{D}_k$ specifies the weight of the local gradient of the k-th user in over-the-air global aggregation.

A. General Process

Any training round of the considered FL system, i.e., the t-th training round, $1 \leq t \leq T$, comprises the following steps.

1) User Selection: Users with better channel conditions must reduce their transmit power to align the transmit scalar at the server to conduct AirComp. By considering channel heterogeneity (i.e., variations among users in channel fading) and data heterogeneity (i.e., variations in data quantity and model loss), the server selects an adequate subset of the users.

2) Broadcast: The server broadcasts the current global model parameter $\boldsymbol{w}_t$ to the selected users.

3) Noisy Gradient Computation: After receiving the global model parameter, each selected user updates its local gradient $\boldsymbol{g}_k(\boldsymbol{w}_t)$ regarding the local dataset $\mathcal{D}_k$ using (3). Then, each user clips its local gradient and perturbs the gradient using the Gaussian noise generated with the optimum variance (as will be optimized in Section IV-C). Each user uploads the perturbed local gradients to the server using the same time-frequency resource of the wireless channels.

4) Model Aggregation: Upon the receipt of the aggregated signals from the selected users, the server obtains a weighted sum of the users' local gradient $\boldsymbol{r}_t = \sum_{k\in\mathcal{K}_t} D_k \boldsymbol{g}_k(\boldsymbol{w}_t)$. Let $\hat{\boldsymbol{r}}_t$ be the estimate of $\boldsymbol{r}_t$, which is caused by channel fading, communication and artificial noise. The server updates the global gradient $\boldsymbol{w}_{t+1}$ based on $\hat{\boldsymbol{r}}_t$ through gradient descent, i.e.,

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \frac{\eta}{\sum_{k\in\mathcal{K}_t} D_k} \hat{\boldsymbol{r}}_t, \tag{10}$$

where $\frac{1}{\sum_{k\in\mathcal{K}_t} D_k}$ is the scaling factor.

B. User Selection

Based on the local model loss and channel characteristics, the server selects the users to optimize the learning performance. We define a new evaluation parameter, $\theta_{m,t}$, $m = 1, \ldots, M$, which is affected by the global update $\theta^t_{m,d}$ and the channel quality $\theta^t_{m,c}$, and can be represented as

$$\theta_{m,t} = \theta^t_{m,d} + \theta^t_{m,c}. \tag{11}$$

The impact of the global update $\theta^t_{m,d}$ can be evaluated using the training data size and the model loss, as given by [36]

$$\theta^t_{m,d} = \frac{D_m}{\sum_{m=1}^{M} D_m} \times TL_m^{t-1},\ m = 1, \ldots, M. \tag{12}$$

which simplifies as $\theta^t_{m,d} = \frac{TL_m^{t-1}}{M}$ when $D_m = D$, $\forall m$, consistent with the uniform weighting used in (2). TL_m^{t-1} denotes the model loss of user m in training round $t-1$. Notably, we consider an IID data distribution across users. In this case, user selection does not introduce non-IID effects into the global model, as the gradient distributions remain statistically aligned regardless of which users are selected.

The impact of the channel quality $\theta^t_{m,c}$ can be evaluated based on the channel coefficients $h_{m,t}$, as given by

$$\theta^t_{m,c} = \left|h_{m,t}\right|^2,\ m = 1, \ldots, M. \tag{13}$$

We propose that the server sorts the users in ascending order of $\theta_{m,t}$, i.e., $\theta_{\pi_1,t} \geq \ldots \geq \theta_{\pi_M,t}$, with $1 \leq \pi_i, \pi_j \leq M$, $1 \leq i, j \leq M$, and $\pi_i \neq \pi_j$ for $i \neq j$. The server selects the users associated with the largest K evaluation parameters, i.e., $\theta_{\pi_p,t}$ with $p = 1, \ldots, K$, as the subset of users participating in the t-th training round, i.e., $\mathcal{K}_t = \{\pi_1, \ldots, \pi_K\}$. This user selection strategy precludes users with either catastrophic channel conditions ($|h_m|^2 \to 0$) or pathological local training loss ($TL_m \to \infty$) to preserve the integrity of the global learning objective, which helps accelerate convergence while preserving validity under the IID data distribution. The computational complexity of user selection scales linearly with the number of users (i.e., $\mathcal{O}(M)$ for M users). This allows efficient selection even at large scales.

C. Noisy Gradient Computation

Before transmitting the local gradients to the server, the selected users k, $k \in \mathcal{K}_t$, clip their local gradients. The noise level of DP is determined by the clipping threshold C. Let $\boldsymbol{g}_{k,t}$ denote the gradient $\boldsymbol{g}_k(\boldsymbol{w}_t)$ for brevity. After the selected user k updates its local model, it clips the updated local model according to

$$clip(\boldsymbol{g}_{k,t}, C) = \boldsymbol{g}_{k,t} \cdot \min\left\{1, \frac{C}{\|\boldsymbol{g}_{k,t}\|}\right\}. \tag{14}$$

The user injects the artificial noise $\boldsymbol{n}_{k,t} \sim \mathcal{CN}(0, \sigma^2_{k,t}\boldsymbol{I})$ to the local gradients in the t-th training round, generating a perturbed version of the local update $\hat{\boldsymbol{g}}_{k,t}$, as given by

$$\hat{\boldsymbol{g}}_{k,t} = \boldsymbol{g}_{k,t} + \boldsymbol{n}_{k,t}. \tag{15}$$

To analyze the impact of clipping and perturbation, we define $\tilde{\boldsymbol{g}}_{k,t}$ as the local update after clipping and before perturbation. We consider that the loss function $f_k(w_t)$ has G-bounded gradients. The expected normalized squared L_2-distance between $\hat{\boldsymbol{g}}_{k,t}$ and $\boldsymbol{g}_{k,t}$ can be upper bounded by

$$\begin{aligned}\mathbb{E}\left[\frac{1}{d}\left\|\hat{\boldsymbol{g}}_{k,t} - \boldsymbol{g}_{k,t}\right\|_2^2\right] &\leq \frac{1}{d}\left(\mathbb{E}\left\|\hat{\boldsymbol{g}}_{k,t} - \boldsymbol{g}_{k,t}\right\|_2^2 + \left\|\tilde{\boldsymbol{g}}_{k,t} - \boldsymbol{g}_{k,t}\right\|_2^2\right) \\ &= \frac{1}{d}\max(0, \left\|\boldsymbol{g}_{k,t}\right\| - C^2) + \frac{\sigma^2 C^2}{K},\end{aligned} \tag{16}$$

where the first term on the right-hand side (RHS) captures the clipping bias inherent to gradient truncation, and is nonzero when $\left\|\boldsymbol{g}_{k,t}\right\| > C^2$. The second term represents the noise-induced variance arising from Gaussian perturbations with variance σ^2, scaled by C^2 and inversely proportional to the number of participating clients K. It was established in [37] that for any clipping threshold $C \geq n_t EG$, the effect on the convergence upper bound is $\mathcal{O}\left(\frac{\eta\eta_t ETd\ln\left(\frac{1}{\delta}\right)}{K^2\epsilon^2}\right)$.[1]

D. Over-the-Air Model Aggregation

For conciseness, we omit the superscript "t" in this section. Consider a block fading channel with invariant channel coefficients throughout the FL training. During any training round t, the users synchronously transmit their gradients in I time slots, one element per slot. Let $x_k[i]$ denote the transmit signal from user k at time index i, which is a function of the locally computed update $\hat{\boldsymbol{g}}_k$. Let $y[i]$ denote the corresponding received signal at the server, and $h_k[i] \in \mathbb{C}$ denote the complex channel coefficient between user k and the server at time i. The received signal is given by

$$y[i] = \sum_{k \in \mathcal{K}} h_k[i]x_k[i] + m[i], \tag{17}$$

where $m[i] \in \mathbb{C}$ is the additive white Gaussian noise (AWGN) at the BS with mean 0 and variance σ_m^2. The additive Gaussian noise is the most commonly adopted (if not universal) at communication receivers. Leveraging this naturally occurring Gaussian noise for privacy protection, we repurpose a traditionally adverse effect into a beneficial one, enhancing privacy without compromising performance.

Employing AirComp, the selected users transmit the clipped and perturbed local updates $\{\hat{\boldsymbol{g}}_k : k \in \mathcal{K}\}$ in the same time-frequency resources. $\hat{g}_k[i]$ is the i-th entry of $\hat{\boldsymbol{g}}_k$ with $i = 1, \ldots, I$. User k converts $\hat{g}_k[i]$ to the transmit signals $\{x_k[i] : i = 1, \ldots, I, k \in \mathcal{K}\}$. To normalize $\hat{g}_k[i]$, we calculate the mean and variance of the updated gradient statistics as $\bar{g}_k = \frac{1}{I}\sum_{i=1}^{I}\hat{g}_k[i]$ and $\upsilon_k^2 = \frac{1}{I}\sum_{i=1}^{I}(\hat{g}_k[i] - \bar{g}_k)$, respectively. The selected users upload $\{\bar{g}_k, \upsilon_k^2\}$, $\forall k \in \mathcal{K}$ to the server. The transmit signal of user k can be designed as

$$x_k[i] = \alpha_k s_k[i] \quad \text{with} \quad s_k[i] \triangleq \frac{\hat{g}_k[i] - \bar{g}_k}{\upsilon_k}, \tag{18}$$

where $\alpha_k = \frac{c}{h_k[i]} \in \mathbb{C}$ is the transmit scalar. The common scaling factor c ensures that signal superposition requires consistent effective channel gain across users, achieving amplitude alignment, a design principle consistently adopted in AirComp literature, e.g., [19], [29], [30].

In (18), $\hat{g}_k[i]$ is normalized to a zero-mean unit-variance symbol $s_k[i]$ using $\bar{g}_k$ and υ_k, such that $\mathbb{E}[s_k[i]] = 0$ and $\mathbb{E}[s_k^2[i]] = 1$. This normalization step ensures that $\mathbb{E}[x_k^2[i]] = |\alpha_k|^2$, where $|\alpha_k|^2$ controls the transmit power. The heterogeneity in channel conditions directly controls the transmit power through the well-established channel inversion principle [30], as given by

$$\mathbb{E}\left[|x_k[i]|^2\right] = \left|\alpha_k\right|^2 = \left|\frac{c}{h_k[i]}\right|^2 \leq P_0, \tag{19}$$

[1] Recent studies proposed adaptive clipping mechanisms where C is adjusted based on the statistical distribution (e.g., median or percentile) of the gradient norms observed across users in each round [37], [38]. These methods track the underlying gradient magnitudes while maintaining DP guarantees, thereby reducing the risk of under- or over-clipping.

where $P_0 \geq 0$ is the maximum transmit power of a user, and c can be determined by the worst-case channel condition to satisfy all power constraints.

By substituting (18) into (17), the aggregated signal y in time slot i is obtained as

$$y[i] = \sum_{k\in\mathcal{K}} h_k[i]\alpha_k \left[\frac{\hat{g}_k[i] - \bar{g}_k}{\upsilon_k}\right] + m[i]. \tag{20}$$

Then, the estimated global update $\hat{r}[i]$ is given by

$$\begin{aligned}\hat{r}[i] &= \frac{1}{\sqrt{\lambda}c} y[i] + \bar{g} \\ &= \frac{1}{\sqrt{\lambda}} \left(\sum_{k\in\mathcal{K}} \frac{\hat{g}_k[i] - \bar{g}_k}{\upsilon_k} + \frac{m[i]}{c}\right) + \bar{g} \\ &= \frac{1}{\sqrt{\lambda}} \left(\sum_{k\in\mathcal{K}} \frac{g_k[i] - \bar{g}_k}{\upsilon_k} + \sum_{k\in\mathcal{K}} \frac{n_k[i]}{\upsilon_k} + \frac{m[i]}{c}\right) + \bar{g},\end{aligned} \tag{21}$$

where $\bar{g} = \sum_{k\in\mathcal{K}} D_k \bar{g}_k$, and λ is the normalization scalar on the server's side to cancel the impact of the normalization operation conducted at the users [39]. The server can restore the gradient by adding $\bar{g}$ in (21), as $\bar{g}_k$ is subtracted for normalization on the users' side before each selected user k transmits its local update; see (18).

After estimating $\hat{\boldsymbol{r}} = [\hat{r}[1], \ldots, \hat{r}[I]]$, the server performs post-processing on $\hat{\boldsymbol{r}}$, as follows:

$$\begin{aligned}\nabla\hat{F}(\boldsymbol{w}) &= \frac{1}{KD}\hat{\boldsymbol{r}} = \frac{1}{KD\sqrt{\lambda}c}\boldsymbol{y} + \frac{1}{KD}\bar{g}\mathbf{1}_I \\ &= \frac{1}{KD\sqrt{\lambda}} \left(\sum_{k\in\mathcal{K}} \frac{\boldsymbol{g}_k - \bar{g}_k\mathbf{1}_I}{\upsilon_k} + \sum_{k\in\mathcal{K}} \frac{\boldsymbol{n}_k}{\upsilon_k} + \frac{\boldsymbol{m}}{c}\right) \\ &= \underbrace{\frac{1}{KD\sqrt{\lambda}} \sum_{k\in\mathcal{K}} \frac{\boldsymbol{g}_k - \bar{g}_k\mathbf{1}_I}{\upsilon_k} + \frac{1}{KD}\bar{g}\mathbf{1}_I}_{\tilde{\boldsymbol{g}}} \\ &\quad + \underbrace{\frac{1}{KD\sqrt{\lambda}} \left(\sum_{k\in\mathcal{K}} \frac{\boldsymbol{n}_k}{\upsilon_k} + \frac{\boldsymbol{m}}{c}\right)}_{\boldsymbol{z}},\end{aligned} \tag{22}$$

where $\tilde{\boldsymbol{g}}$ and $\boldsymbol{z}$ are defined for conciseness.

Then, the server updates the global model parameters according to (10), i.e.,

$$\boldsymbol{w}_t = \boldsymbol{w}_{t-1} - \eta\nabla\hat{F}(\boldsymbol{w}_t) = \boldsymbol{w}_{t-1} - \eta(\tilde{\boldsymbol{g}}_t + \boldsymbol{z}_t). \tag{23}$$

To analyze the impact of the normalization in (18), we quantify the distortion between the global gradient $\mathbf{r}$ and its estimate $\hat{\mathbf{r}}$. The mean squared error (MSE) is computed as

$$\begin{aligned}&\mathbb{E}\left[\left\|\frac{\mathbf{r}_t}{\sum_{k\in\mathcal{K}} D_k} - \frac{\hat{\mathbf{r}}_t}{\sum_{k\in\mathcal{K}} D_k}\right\|_2^2\right] = \frac{1}{\left(\sum_{k\in\mathcal{K}} D_k\right)^2} \\ &\sum_{i=1}^{I} \mathbb{E}\left[\left|\sum_{k\in\mathcal{K}} \left(D_k - \frac{1}{\sqrt{\lambda}\upsilon_k}\right)(\hat{g}_k[i] - \bar{g}_k)\right|^2 + \frac{|m[i]|^2}{\lambda c^2}\right],\end{aligned} \tag{24}$$

which can be rewritten as a function of the server's normalization scalar λ, as given by

$$\begin{aligned}F(\lambda) &= \sum_{i=1}^{I} \mathbb{E}\left[\left|\sum_{k\in\mathcal{K}} \left(D_k - \frac{1}{\sqrt{\lambda}\upsilon_k}\right)(\hat{g}_k[i] - \bar{g}_k)\right|^2\right] + \frac{S_m}{\lambda} \\ &= I\sum_{k\in\mathcal{K}} \left(D_k - \frac{u}{\upsilon_k}\right)^2 \upsilon_k^2 + S_m u^2,\end{aligned} \tag{25}$$

where $u = \frac{1}{\sqrt{\lambda}}$ and $S_m = \frac{1}{c^2}\sum_{i=1}^{I} |m[i]|^2$. Expanding the quadratic form yields

$$F(u) = I\sum_{k\in\mathcal{K}} D_k^2\upsilon_k^2 - 2uI\sum_{k\in\mathcal{K}} D_k\upsilon_k + u^2(IK + S_m), \tag{26}$$

with its minimum taken at $\frac{\partial F}{\partial u} = 0$, i.e., $u^* = \frac{I\sum_{k\in\mathcal{K}} D_k\upsilon_k}{IK + S_m}$. The optimal solution of λ, denoted by λ^*, is given by

$$\begin{aligned}\lambda^* &= \left(\frac{IK + \frac{1}{c^2}\sum_{i=1}^{I} |m[i]|^2}{I\sum_{k\in\mathcal{K}} D_k\upsilon_k}\right)^2 \\ &= \left(\frac{K + \frac{1}{c^2}\sigma_m^2}{\sum_{k\in\mathcal{K}} D_k\upsilon_k}\right)^2 \overset{(a)}{\approx} \left(\frac{K}{\sum_{k\in\mathcal{K}} D_k\upsilon_k}\right)^2,\end{aligned} \tag{27}$$

where the approximation (a) is valid under typical AirCOMP settings with $K \gg \frac{1}{c^2}\sigma_m^2$, since the SNR scales with K, i.e., SNR $= \frac{|c|^2 K}{\sigma_m^2}$, and with user selection ($K \geq 1$), the multi-user beamforming gain dominates over the noise. To this end, selecting an appropriate normalization scalar λ at the server contributes to mitigating communication distortion.

Algorithm 1 describes this PEA-FL algorithm.

IV. Convergence and Privacy Analysis, and Loss-Leakage Optimization

In this section, we analyze the convergence bound of the proposed AirComp framework. We derive the privacy constraint from the server perspective by analyzing the effective noise perturbations in AirComp. Considering that the adversary can use a signal detection technique to extract signals sent by individual users, we derive a closed-form privacy leakage function from the adversary's perspective.

A. Convergence Analysis

We start with the following commonly adopted assumptions:

Assumption 1: (L-Lipschitz Continuous Gradient): ∇f is Lipschitz continuous with a constant $L > 0$, if $\|\nabla f_k(\boldsymbol{x}) - \nabla f_k(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^I$, $k \in \mathcal{K}_t$.

Assumption 2: (Bounded Local and Global Variance): The variance of each local gradient estimator is bounded by $\mathbb{E}\left[\left\|\nabla f_k(\boldsymbol{w}_t, \xi_t^k) - \nabla f_k(\boldsymbol{w}_t)\right\|^2\right] \leq \sigma_L^2$, $\forall k \in \mathcal{K}_t$, with the parameter $\sigma_L > 0$. The global variability of the local gradient of the cost function is bounded by $\mathbb{E}\left[\|\nabla f_k(\boldsymbol{w}_t) - \nabla F(\boldsymbol{w}_t)\|^2\right] \leq \sigma_G^2$ with the parameter $\sigma_G > 0$.

Under these assumptions, we analyze the errors of model updates resulting from DP perturbation and communication noise, leading to the establishment of the convergence bound for PEA-FL, as asserted in the following theorem.

Theorem 1: (Convergence Bound of PEA-FL): Suppose the constant local learning rate η_t and the global learning rate η satisfy $\eta_t < \frac{1}{2\sqrt{30}LE}$ and $\eta\eta_t \leq \frac{1}{LE}$, respectively. Under Assumptions 1 and 2, it holds that

$$\min_{t\in[1,T]} \mathbb{E}\left[\|\nabla F(\boldsymbol{w}_t)\|^2\right] \leq \frac{F_0 - F_*}{c'E\eta\eta_t T} + \Phi, \tag{28}$$

Algorithm 1 Proposed PEA-FL Algorithm

Initialize $\boldsymbol{w}_0$, number of local updates E, privacy budget ϵ, channel noise $\boldsymbol{m}_t$, channel response $\boldsymbol{h}_k$ from user k, clipping threshold C.
for $t = 0, 1, \ldots, T-1$ **do**
Each user calculates evaluation value by (12) and (13) and uploads the evaluation parameter to the server.
The server selects the largest K evaluation values $\theta_{k,t}$ to choose an adequate subset $\mathcal{K}_t$.
The server broadcasts $\boldsymbol{w}_t$ to users in the subset $\mathcal{K}_t$.
for each user $k \in \mathcal{K}_t$ **do**
$\boldsymbol{w}_{t,0}^k = \boldsymbol{w}_t$.
for $e = 0, \ldots, E-1$ **do**
$\boldsymbol{g}_{t,e}^k = \nabla f_k(\boldsymbol{w}_{t,e}^k; \boldsymbol{x}, y)$ of $\nabla f_k(\boldsymbol{w}_{t,e}^k)$.
$\boldsymbol{w}_{t,e+1}^k = \boldsymbol{w}_{t,e}^k - n_t \boldsymbol{g}_{t,e}^k$.
end for
$\boldsymbol{g}_{t,k} = \boldsymbol{w}_{t,E}^k - \boldsymbol{w}_{t,0}^k = -n_t \sum_{e=0}^{E-1} \boldsymbol{g}_{t,e}^k$.
Clip the local gradient and use optimized scaling factor $\boldsymbol{n}_{t,k}$ to perturb $\boldsymbol{g}_{t,k}$ by (14) and (15).
Compute the local gradient statistics of $\hat{\boldsymbol{g}}_{t,k}$ to normalize the model parameter and use optimized scaling factor c_t to transmit by (18).
end for
At the server:
Receive the aggregated signal by (20).
Compute the estimated global update by (22).
Update the global model parameters by (23).
end for

where $\Phi \triangleq \frac{1}{c'}\left[\frac{L\eta_t}{K}\sigma_L^2 + 5E\eta_t^2L^2(\sigma_L^2 + 6E\sigma_G^2)\right] + \frac{I}{E\eta_t c'}\left(\frac{1}{\eta_t E} + \frac{3L\eta}{2}\right)\frac{1}{(KD)^2\lambda}\left(\sum_{k\in\mathcal{K}_t}\frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2} + \frac{\sigma_{m,t}^2}{c_t^2}\right)$; *$F_0$ and F_* represent the initial and the optimal loss function values, respectively; c' is a constant. To satisfy (28), the constant c' must be a positive value strictly bounded by*

$$0 < c' < \frac{1}{2} - 30E^2\eta_t^2L^2, \tag{29}$$

where $\frac{1}{2} - 30E^2\eta_t^2L^2 > 0$ *is satisfied since* $\eta_t < \frac{1}{2\sqrt{30}LE}$.

Proof: See Appendix 1.■

As shown in Theorem 1, for finite T, $\{c_t, \sigma_{k,t}^2\}$ are important determinants of the convergence bound of PEA-FL. Specifically, the first term on the RHS of (28) is a constant independent of $\{c_t, \sigma_{k,t}^2\}$, while the second term is a function of these parameters. To explicitly capture and optimize the impact of $\{c_t, \sigma_{k,t}^2\}$ on the convergence of PEA-FL, we define the loss function $\mathcal{L}$ as

$$\mathcal{L}(c_t, \sigma_{k,t}^2) = \frac{I}{E\eta_t c'}\left(\frac{1}{\eta_t E} + \frac{3L\eta}{2}\right)\frac{1}{(KD)^2\lambda}\left(\sum_{k\in\mathcal{K}_t}\frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2} + \frac{\sigma_{m,t}^2}{c_t^2}\right). \tag{30}$$

Clearly, $\mathcal{L}(c_t, \sigma_{k,t}^2)$ captures the impact of $\{c_t, \sigma_{k,t}^2\}$ on the convergence of PEA-FL.

B. Privacy Leakage Analysis

With the consideration of an "honest-but-curious" server, we derive the privacy constraint satisfying RDP to prevent the privacy leakage of PEA-FL. Since $\{c_t, \sigma_{k,t}^2\}$ are fixed constants, we focus on the local gradients uploaded from the users to the server. Based on (15) and (20), we start by decoupling the received signal $\boldsymbol{y}_t$, as given by

$$\boldsymbol{y}_t = \sum_{k\in\mathcal{K}_t} c_t\left[\frac{\boldsymbol{g}_{k,t} - \bar{g}_{k,t}\mathbf{1}_I}{\upsilon_{k,t}}\right] + \sum_{k\in\mathcal{K}_t} c_t\left[\frac{\boldsymbol{n}_{k,t}}{\upsilon_{k,t}}\right] + \boldsymbol{m}_t \tag{31a}$$

$$= c_t\left[\frac{\boldsymbol{g}_{k,t} - \bar{g}_{k,t}\mathbf{1}_I}{\upsilon_{k,t}}\right] + \sum_{j\in\mathcal{K}_t, j\neq k} c_t\left[\frac{\boldsymbol{g}_{j,t} - \bar{g}_{j,t}\mathbf{1}_I}{\upsilon_{j,t}}\right] + \sum_{k\in\mathcal{K}_t} c_t\left[\frac{\boldsymbol{n}_{k,t}}{\upsilon_{k,t}}\right] + \boldsymbol{m}_t. \tag{31b}$$

The first term on the RHS of (31b) is the transmitted signal concerning user k's gradients $\boldsymbol{g}_{k,t}$, the second term is the transmitted signal concerning the other users' gradients, and the third term is the noise perturbation.

Note that the effective noise perturbation on PEA-FL offers the same degree of privacy protection to all transmitted signals, and stems from the artificial DP noises from all users and the inherent communication noise. Let ξ_t denote the variance of the effective noise. Then,

$$\xi_t = \sum_{k\in\mathcal{K}_t} c_t^2\left[\frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2}\right] + \sigma_{m,t}^2. \tag{32}$$

Also note in (9) that the DP level also depends on the sensitivity of the dataset. To deliver (α, ϵ)-RDP, we define a global sensitivity Δs dependent on the sensitivity of the noise-free transmitted signals. Consider two neighboring datasets, $\mathcal{D}_{k,t}$ and $\mathcal{D}'_{k,t}$. Let $\mathcal{M}$ be a random mechanism applied to $\mathcal{D}_{k,t}$ and $\mathcal{D}'_{k,t}$. The sensitivity Δs can be calculated as

$$\begin{aligned}\Delta s &= \max_{\mathcal{D}_{k,t}, \mathcal{D}'_{k,t}} \left\|\mathcal{M}(\mathcal{D}_{k,t}) - \mathcal{M}(\mathcal{D}'_{k,t})\right\| \\ &= \max_{\mathcal{D}_{k,t}, \mathcal{D}'_{k,t}} \left\|\boldsymbol{y}_t - \boldsymbol{y}'_t\right\| \\ &= \max_{\mathcal{D}_{k,t}, \mathcal{D}'_{k,t}} \left\|c_t\Big(\frac{\boldsymbol{g}_{k,t}(\boldsymbol{w}_t, \mathcal{D}_{k,t}) - \bar{g}_{k,t}\mathbf{1}}{\upsilon_{k,t}} - \frac{\boldsymbol{g}_{k,t}(\boldsymbol{w}_t, \mathcal{D}'_{k,t}) - \bar{g}_{k,t}\mathbf{1}_I}{\upsilon_{k,t}}\Big)\right\| \\ &= \frac{2\eta_t c_t}{\upsilon_{k,t}^2} \cdot \max \left\|\boldsymbol{g}_{k,t}(\boldsymbol{w}_t, \mathcal{D}_{k,t}) - \bar{g}_{k,t}\mathbf{1}_I\right\| \overset{(a)}{\leq} \frac{4\eta_t c_t C}{\upsilon_{k,t}^2},\end{aligned} \tag{33}$$

where (a) holds as the gradients are clipped before uploading.

Based on (9) in Definition 4 and (33), the privacy constraint for the selected users is provided in the ensuing lemma.

Lemma 1: For any fixed sequence of $\{c_t, \sigma_{k,t}^2\}_{t=1}^T$, *the PEA-FL system satisfies* (α, ϵ)*-RDP if*

$$\sum_{t=1}^{T} \frac{8\alpha(\eta_t c_t C)^2}{\upsilon_k^2 \xi_t} \leq \epsilon. \tag{34}$$

Both the wireless channel noise and artificial noise contribute to the privacy guarantee of the AirComp-FL system.

Proof: This lemma is obtained by plugging (33) into (9).■

We proceed to derive the privacy leakage from the adversary's perspective. As mentioned earlier, AirComp has

the waveform superposition property. Consequently, the adversary cannot obtain each user's model gradients directly to implement the DLG attack. Suppose that the adversary employs a ZF strategy [40] for signal detection to reconstruct the gradient of each individual user from the global aggregation. The adversary is equipped with N antennas, $N \geq K$, as the adversary can separate and recover all users' signals using the ZF strategy only if the number of antennas, N, is greater than or equal to the number of selected users, K.[2] The signal obtained at each antenna is a sum of the transmitted signals scaled by the channel coefficients.

Let $\boldsymbol{X}_t = \left[\boldsymbol{x}_{1,t}, \ldots, \boldsymbol{x}_{K,t}\right]^T \in \mathbb{C}^{K\times I}$ collect all signals transmitted by the selected users, $\boldsymbol{H}_t = \left[\boldsymbol{h}_{1,t}, \ldots, \boldsymbol{h}_{K,t}\right] \in \mathbb{C}^{N\times K}$ collect the channel responses from user k to the server, and $\boldsymbol{M}_t = \boldsymbol{m}_t \mathbf{1}_I^T$ collect the channel noise with $\boldsymbol{m}_t = \left[m_{1,t}, \ldots, m_{N,t}\right] \in \mathbb{C}^{N\times 1}$. The signal model $\boldsymbol{Y}_t = \left[\boldsymbol{y}_{1,t}, \ldots, \boldsymbol{y}_{N,t}\right]^T \in \mathbb{C}^{N\times I}$ is multi-dimensional, as given by

$$\boldsymbol{Y}_t = \boldsymbol{H}\boldsymbol{X}_t + \boldsymbol{M}_t. \tag{35}$$

After ZF detection, the detected signals $\hat{\boldsymbol{X}}_t$ can be written as

$$\begin{aligned}\hat{\boldsymbol{X}}_t &= \left(\boldsymbol{H}_t^H \boldsymbol{H}_t\right)^{-1} \boldsymbol{H}_t^H \boldsymbol{Y}_t \\ &= \left(\boldsymbol{H}_t^H \boldsymbol{H}_t\right)^{-1} \boldsymbol{H}_t^H \left(\boldsymbol{H}_t \boldsymbol{X}_t + \boldsymbol{M}_t\right) \\ &= \boldsymbol{X}_t + \left(\boldsymbol{H}_t^H \boldsymbol{H}_t\right)^{-1} \boldsymbol{H}_t^H \boldsymbol{M}_t. \end{aligned} \tag{36}$$

Define $\boldsymbol{J} = [\boldsymbol{j}_{1,t}, \ldots, \boldsymbol{j}_{K,t}]^T \in \mathbb{C}^{K\times N}$ with $\boldsymbol{J} \triangleq \left(\boldsymbol{H}^H \boldsymbol{H}\right)^{-1} \boldsymbol{H}^H$. Based on (36), the adversary can compute the estimated local update of user k, as given by

$$\begin{aligned}\nabla \hat{f}_k(\boldsymbol{w}_t) &= \frac{\upsilon_{k,t}}{\alpha_{k,t}} \hat{\boldsymbol{x}}_{k,t} + \bar{g}_{k,t} \\ &= \frac{\upsilon_{k,t}}{\alpha_{k,t}} \left(\alpha_{k,t} \left[\frac{\hat{\boldsymbol{g}}_{k,t} - \bar{g}_{k,t}}{\upsilon_{k,t}}\right] + \boldsymbol{j}_{k,t} \boldsymbol{m}_t \mathbf{1}_I^T\right) + \bar{g}_{k,t} \\ &= \boldsymbol{g}_{k,t} + \boldsymbol{n'}_{k,t}, \end{aligned} \tag{37}$$

where $\boldsymbol{n'}_{k,t} = \boldsymbol{n}_{k,t} + \frac{\upsilon_{k,t}}{\alpha_{k,t}} \boldsymbol{j}_{k,t} \boldsymbol{m}_t \mathbf{1}_I^T$ is the error caused by ZF and the added artificial DP noise. Here, $\frac{\upsilon_{k,t}}{\alpha_{k,t}} \boldsymbol{j}_{k,t} \boldsymbol{m}_t \mathbf{1}_I^T$ accounts for the ZF-induced estimation error, arising from imperfect channel inversion. This stems from the relation

$$\left(\boldsymbol{H}_t^H \boldsymbol{H}_t\right)^{-1} \boldsymbol{H}_t^H \boldsymbol{M}_t = \boldsymbol{J} \boldsymbol{m}_t \mathbf{1}_I^T, \tag{38}$$

where $\boldsymbol{j}_{k,t}$ is the k-th row of $\boldsymbol{J}$, and $\boldsymbol{m}_t$ captures the uplink channel noise. The inversion in (38) amplifies this noise via $\boldsymbol{j}_{k,t}$, introducing non-negligible distortion. The variance of $\boldsymbol{n'}_{k,t}$ can be expressed as

$$\begin{aligned}\hat{\sigma}_{k,t}^2 &= \mathbb{E}\left[\|\boldsymbol{n'}_{k,t}\|^2\right] \\ &= \underbrace{\mathbb{E}\left[\|\boldsymbol{n}_{k,t}\|^2\right]}_{\text{DP noise}} + \underbrace{\left(\frac{\upsilon_{k,t}}{\alpha_{k,t}}\right)^2 \|\boldsymbol{j}_{k,t}\|^2 \mathbb{E}\left[\|\boldsymbol{m}_t\|^2\right]}_{\text{ZF-amplified channel noise}}. \end{aligned} \tag{39}$$

The ZF-induced error is beyond the adversary's control and complements the artificial DP noise, enhancing privacy.

[2]When $N < K$, it is impossible to perfectly separate the K individual user signals. ZF involves computing the pseudo-inverse $\boldsymbol{H}_t^\dagger$ of the $N \times K$ channel matrix $\boldsymbol{H}_t$. Perfect signal separation requires $\boldsymbol{H}_t$ to have full column rank, i.e., $\text{rank}(\boldsymbol{H}_t) = K$. When $\text{rank}(\boldsymbol{H}_t) \leq \min(N, K) = N < K$, $\boldsymbol{H}_t$ is rank-deficient and hence not left-invertible.

The adversary obtains the perturbed local gradients from $\nabla \hat{f}_k(\boldsymbol{w}_t)$ using (37), which include the noise disturbances that perturb the local gradients and cannot be separated. Based on (3), the perturbed local gradients can be computed as

$$\begin{aligned}\nabla f_k(\boldsymbol{w}_t) &= -\frac{\nabla \hat{f}_k(\boldsymbol{w}_t)}{\eta_t E} \\ &= \frac{1}{E} \sum_{e=1}^{E} f_k(\boldsymbol{w}_t; \boldsymbol{x}, y) - \frac{\boldsymbol{n'}_{k,t}}{\eta_t E} \\ &= \nabla \bar{f}(\boldsymbol{w}_t) - \frac{\boldsymbol{n'}_{k,t}}{\eta_t E}, \end{aligned} \tag{40}$$

where $\bar{f}(\boldsymbol{w}_t)$ is the mean local gradient.

By applying (40) into (5), the gradient match loss is

$$\|\nabla f(\boldsymbol{w}') - \nabla f(\boldsymbol{w})\|^2$$

$$= \left\|\nabla f(\boldsymbol{w'}_t) - \left(\nabla \bar{f}(\boldsymbol{w}_t) - \frac{\boldsymbol{n'}_{k,t}}{\eta_t E}\right)\right\|^2 \tag{41a}$$

$$\overset{(a)}{\leq} 2\left\|\nabla f(\boldsymbol{w'}_t) - \nabla \bar{f}(\boldsymbol{w}_t)\right\|^2 + 2\left\|\frac{\boldsymbol{n'}_{k,t}}{\eta_t E}\right\|^2 \tag{41b}$$

$$= 2\left\|\nabla f(\boldsymbol{w'}_t) - \nabla \bar{f}(\boldsymbol{w}_t)\right\|^2 + \frac{2\hat{\sigma}_{k,t}^2}{\eta_t^2 E^2}, \tag{41c}$$

where $\hat{\sigma}_{k,t}^2$ is the noise scale. Here, (41b) comes from $\|\boldsymbol{a} + \boldsymbol{b}\|^2 \leq 2\|\boldsymbol{a}\|^2 + 2\|\boldsymbol{b}\|^2$.

The first term $2\left\|\nabla f(\boldsymbol{w'}_t) - \nabla \bar{f}(\boldsymbol{w}_t)\right\|^2$ in (41c) is the objective function of the adversary. It gradually tends to zero using the standard gradient-based method. The second term $\frac{2\hat{\sigma}_{k,t}^2}{\eta_t^2 E^2}$ in (41c) reflects the impact of ZF and added DP noise on the gradient match loss, which is non-negligible. The gradient match loss depends linearly on $\hat{\sigma}_{k,t}^2$. As $\hat{\sigma}_{k,t}^2 \to \infty$, the adversary is unable to minimize the gradient matching loss below a meaningful threshold, hence failing to recover user k's model updates. Since $\hat{\sigma}_{k,t}^2$ depends linearly on both c_t and $\sigma_{k,t}^2$, we posit that the privacy leakage is inversely proportional to $\{c_t, \sigma_{k,t}^2\}$. This assumption has been empirically and analytically verified in existing work, e.g., [31], where a negative correlation was demonstrated between the adversary's gradient matching loss and the privacy loss.

The privacy leakage level can be defined as the reciprocal of the gradient match loss. Based on (37), the privacy leakage $\mathcal{G}$ is defined as

$$\mathcal{G}_{k,t}(c_t, \sigma_{k,t}^2) \triangleq \frac{\eta_t^2 E^2}{2\hat{\sigma}_{k,t}^2} = \frac{\eta_t^2 E^2}{2\left(\sigma_{k,t}^2 + \left(\frac{\upsilon_{k,t}|h_{k,t}|}{c_t}\right)^2 \sigma_{j_{k,t}}^2 \sigma_{m,t}^2\right)}, \tag{42}$$

where $\sigma_{j_{k,t}}^2$ is the variance of $\boldsymbol{j}_{k,t}$.

C. Joint Loss-Leakage Optimization

The objective of PEA-FL is to minimize global loss and privacy leakage. Since the global loss and privacy leakage are functions of c_t and $\sigma_{k,t}^2$, $\forall k, t$, the overall loss function L concerning c_t and $\sigma_{k,t}^2$, $\forall k, t$, is given by

$$L(c_t, \sigma_{k,t}^2) = \sum_{k\in\mathcal{K}_t} \mathcal{G}_{k,t}(c_t, \sigma_{k,t}^2) + \mathcal{L}(c_t, \sigma_{k,t}^2). \tag{43}$$

By substituting (30) and (42) into (43), with the power constraint (19) and the privacy constraint (34), the joint loss-leakage optimization problem can be formulated as

$$\min_{c_t,\sigma_{k,t}^2,\xi_t} \quad L_1 + L_2 \tag{44a}$$

$$s.t. \quad \left(\frac{c_t}{h_{k,t}}\right)^2 \le P_0, \quad \forall k,t \tag{44b}$$

$$\xi_t \le \sum_{k\in\mathcal{K}_t} c_t^2 \frac{\sigma_{k,t}^2}{\upsilon_k^2} + \sigma_{m,t}^2, \tag{44c}$$

$$\sum_{t=1}^{T} \frac{8\alpha(\eta_t c_t C)^2}{\upsilon_{k,t}^2 \xi_t} \le \epsilon, \tag{44d}$$

$$0 \le \xi_t, 0 \le c_t, 0 \le \sigma_{k,t}^2, \tag{44e}$$

where $L_1 = \sum_{k\in\mathcal{K}_t} \frac{\Phi'_2}{\sigma_{k,t}^2 + (\upsilon_{k,t} h_{k,t}/c_t)^2 \sigma_{j_k,t}^2 \sigma_{m,t}^2}$ originates from the privacy leakage $\mathcal{G}$, and $L_2 = \Phi'_1 \left(\sum_{k\in\mathcal{K}_t} \frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2} + \frac{\sigma_{m,t}^2}{c_t^2}\right)$ stems from the global loss $\mathcal{L}$ with $\Phi'_1 = \frac{I}{E\eta_t c'} \left(\frac{1}{\eta_t E} + \frac{3L\eta}{2}\right) \frac{1}{(KD)^2 \lambda}$ and $\Phi'_2 = \frac{\eta_t^2 E^2}{2}$. Moreover, (44b) is a power constraint. (44c) and (44d) are the privacy constraints of the selected users, inherited from (32) and (34), respectively.

Leveraging variable substitution with $p_t = \frac{1}{c_t^2}$, $q_t = \frac{\xi_t}{c_t^2}$, and $u_{k,t} = \sigma_{k,t}^2$, $\forall k$, problem (44) can be rewritten as

$$\min_{\{p_t,q_t,u_{k,t},\forall k\}} \quad L'_1 + L'_2 \tag{45a}$$

$$\text{s.t. } p_t \ge \max_k \frac{1}{P_0 h_{k,t}^2}, \quad \forall t \tag{45b}$$

$$q_t \le \sum_{k\in\mathcal{K}_t} \frac{u_{k,t}}{\upsilon_{k,t}^2} + p_t \sigma_{m,t}^2, \quad \forall t \tag{45c}$$

$$q_t \ge \max_k \frac{8\alpha(\eta_t C)^2 T}{\upsilon_{k,t}^2 \epsilon}, \quad \forall t \tag{45d}$$

$$p_t \ge 0, \ q_t \ge 0, \ u_{k,t} \ge 0, \quad \forall k,t \tag{45e}$$

where $L'_1 = \Phi'_1 \left(\sum_{k\in\mathcal{K}_t} \frac{u}{\upsilon_{k,t}^2} + p_t \sigma_{m,t}^2\right)$, $L'_2 = \Phi'_2 \times \sum_{k\in\mathcal{K}_t} \frac{1}{u_{k,t} + \Phi'_3 \cdot p_t}$, and $\Phi'_3 = \left(\upsilon_{k,t} h_{k,t}\right)^2 \sigma_{j_k,t}^2 \sigma_{m,t}^2$. (45) is a convex problem with variables, p_t, q_t, and $u_{k,t}$, $\forall k$, and can be solved efficiently using standard gradient descent techniques.

D. Discussion and Extension

1) Complexity Analysis: Given its convexity, problem (45) can be efficiently solved using the interior-point method. The problem involves $K+2$ optimization variables, specifically, p_t, q_t, and $\mathbf{u}_t = \{u_{1,t}, \ldots, u_{K,t}\}$, and $3K+1$ linear inequality constraints. The number of iterations required to achieve an ε-optimal solution is bounded by $\mathcal{O}\left(\sqrt{\nu}\log(1/\varepsilon)\right)$, where the barrier parameter $\nu = 3K+1$ and thus $\sqrt{\nu} = \mathcal{O}(\sqrt{K})$ [41]. The dominant computational cost stems from solving the Newton system, with complexity $\mathcal{O}(K^3)$ per iteration. The total per-round complexity is $\mathcal{O}\left(K^{3.5}\log(1/\varepsilon)\right)$.

When privacy constraints are suppressed (i.e., (45c) and (45d) are dropped), the number of optimization variables reduces to $K+2$, and the total number of constraints becomes $3K-1$. The per-iteration computational cost of the interior-point method is $\mathcal{O}((3K-1)(K+2)^2 + (3K-1)^2(K+2) + (3K-1)^3) = \mathcal{O}(K^3)$ [41]. With a target accuracy tolerance of ε, the number of iterations required to reach an ε-optimal solution is $\mathcal{O}(\sqrt{3K-1}\log(1/\varepsilon)) = \mathcal{O}(\sqrt{K}\log(1/\varepsilon))$ [41]. As a result, the overall per-round computational complexity is $\mathcal{O}(K^{3.5}\log(1/\varepsilon))$, which is consistent with the asymptotic complexity with privacy considered.

While performing well for small-to-medium-scale problems, the interior-point method can be memory-intensive and computationally prohibitive for large-scale problems due to its requirement to compute and store the dense Hessian matrix of the barrier function. In the latter case, the interior-point method with preconditioned conjugate gradient and warm-start can deliver effective performance [42].

2) Extension to Attacks With Minimum MSE (MMSE) Receivers: The adversary can also take other signal detection strategies, such as MMSE, in which case the estimated signals, denoted by $\hat{\mathbf{X}}_t^{\text{MMSE}}$, can be written as

$$\begin{aligned}
\hat{\mathbf{X}}_t^{\text{MMSE}} &= \left(\mathbf{H}_t^H \mathbf{H}_t + \sigma_m^2 \mathbf{I}_K\right)^{-1} \mathbf{H}_t^H \mathbf{Y}_t \\
&= \left(\mathbf{H}_t^H \mathbf{H}_t + \sigma_m^2 \mathbf{I}_K\right)^{-1} \mathbf{H}_t^H (\mathbf{H}_t \mathbf{X}_t + \mathbf{M}_t) \\
&= \underbrace{\left(\mathbf{H}_t^H \mathbf{H}_t + \sigma_m^2 \mathbf{I}_K\right)^{-1} \mathbf{H}_t^H \mathbf{H}_t}_{\mathbf{W}} \mathbf{X}_t \\
&\quad + \underbrace{\left(\mathbf{H}_t^H \mathbf{H}_t + \sigma_m^2 \mathbf{I}_K\right)^{-1} \mathbf{H}_t^H}_{\mathbf{B}} \mathbf{M}_t \\
&= \mathbf{W}\mathbf{X}_t + \mathbf{B}\mathbf{M}_t,
\end{aligned} \tag{46}$$

where $\mathbf{W} \in \mathbb{C}^{K\times K}$ is the signal shrinkage matrix, and $\mathbf{B} \in \mathbb{C}^{K\times N}$ is the noise shaping matrix. The (k,p)-th element $w_{k,p}$ of $\mathbf{W}$ quantifies signal distortion: The diagonal elements $w_{k,k}$, $\forall k$ represent signal preservation gains; the off-diagonal elements $w_{k,p}$, $\forall k \ne p$ characterize inter-user interference.

Let $\mathbf{w}_k$ denote the k-th row vector of $\mathbf{W}$, and $\mathbf{b}_k$ denote the k-th row vector of $\mathbf{B}$. The MMSE estimate for user k's signal is given by

$$\hat{\mathbf{x}}_{k,t}^{\text{MMSE}} = \mathbf{w}_k \mathbf{X}_t + \mathbf{b}_k \mathbf{M}_t = w_{k,k}\mathbf{x}_{k,t} + \sum_{p\ne k} w_{k,p}\mathbf{x}_{p,t} + \mathbf{b}_k \mathbf{M}_t. \tag{47}$$

To recover $\mathbf{g}_{k,t}$ from the MMSE estimate, based on (18) and (47), the scaling factor for gradient reconstruction is given by

$$\beta_{k,t} \triangleq \frac{1}{w_{k,k}} \frac{\upsilon_{k,t}}{\alpha_{k,t}}. \tag{48}$$

Based on (47) and (48), the adversary can estimate the local update of user k, as given by

$$\begin{aligned}
\nabla \hat{f}_k(\mathbf{w}_t) &= \beta_{k,t} \left(w_{k,k}\mathbf{x}_{k,t} + \sum_{p\ne k} w_{k,p}\mathbf{x}_{p,t} + \mathbf{b}_k \mathbf{M}_t\right) + \bar{\mathbf{g}}_{k,t} \\
&= \underbrace{\beta_{k,t} w_{k,k} \frac{\alpha_{k,t}}{\upsilon_{k,t}} \mathbf{g}_{k,t}}_{\mathbf{g}_{k,t}} + \underbrace{\beta_{k,t} w_{k,k} \frac{\alpha_{k,t}}{\upsilon_{k,t}} \mathbf{n}_{k,t}}_{\mathbf{n}_{k,t}} \\
&\quad + \underbrace{\beta_{k,t} \sum_{p\ne k} w_{k,p}\mathbf{x}_{p,t} + \beta_{k,t}\mathbf{b}_k \mathbf{M}_t}_{\mathbf{n}_{k,t}^{\text{total}}}.
\end{aligned} \tag{49}$$

In other words, $\nabla \hat{f}_k(\mathbf{w}_t)$ comprises the true gradient $\mathbf{g}_{k,t}$, artificial DP noise $\mathbf{n}_{k,t}$, multi-user interference $\beta_{k,t}\sum_{p\neq k} w_{k,p}\mathbf{x}_{p,t}$, and MMSE-filtered channel noise $\beta_{k,t}\mathbf{b}_k\mathbf{M}_t$. As $\sigma_m^2 \to 0$ (high SNR), we have

$$w_{k,k} \to 1,\ w_{k,p\neq k} \to 0,\ \mathbf{b}_k \to \mathbf{j}_{k,:}^{\mathrm{ZF}},\ \beta_{k,t} \to \frac{\upsilon_{k,t}}{\alpha_{k,t}}. \tag{50}$$

As a result, (49) degenerates to

$$\begin{aligned}\nabla \hat{f}_k(\mathbf{w}_t) &\to \mathbf{g}_{k,t} + \mathbf{n}_{k,t} + \frac{\upsilon_{k,t}}{\alpha_{k,t}}\sum_{p\neq k} 0\cdot \mathbf{x}_{p,t} + \frac{\upsilon_{k,t}}{\alpha_{k,t}}\mathbf{j}_{k,:}^{\mathrm{ZF}}\mathbf{M}_t \\ &= \mathbf{g}_{k,t} + \mathbf{n}_{k,t} + \frac{\upsilon_{k,t}}{\alpha_{k,t}}\mathbf{j}_{k,t}\mathbf{M}_t,\end{aligned} \tag{51}$$

which concurs with the ZF-based attack in (37).

The proposed PEA-FL framework remains robust against attacks facilitated with the MMSE receivers: In high SNR regimes ($\sigma_m^2 \to 0$), the MMSE detector asymptotically approaches the ZF detector in performance, as shown in (51), thereby requiring comparable levels of local noise injection to ensure privacy. In low SNR regimes, MMSE's superior decoupling capability reduces the impact of channel noise on the adversary's effective observation ($\mathbf{n}_{k,t}^{\mathrm{total}}$), necessitating stronger artificial noise at the clients to preserve privacy guarantees. While more sophisticated attacks, e.g., MMSE, can diminish the inherent "free privacy" benefits offered by ambient channel noise and interference, PEA-FL remains effective through adaptively calibrated local noise injection.

V. Numerical Evaluation

In this section, numerical tests are reported to compare the proposed PEA-FL algorithm with state-of-the-art methods in classification accuracy and privacy preservation.

A. Experimental Setting

Consider a uniform distribution of users at a distance of $d_{KS} = 50$ m from the server. The path losses of the user-server channels follow $G_S G_K \left(\frac{3\times 10^8}{4\pi f_c d_{KS}}\right)^{\gamma}$, where G_S and G_K are the antenna gains at the server and users, respectively; γ denotes the path loss exponent; and f_c is the carrier frequency. By default, we set $G_S = 5$ dBi, $G_K = 0$ dBi, $f_c = 915$ MHz, and $\gamma = 3.76$. The small-scale fading coefficients of the user-server channels adhere to the independent and identically distributed (i.i.d.) Rayleigh fading channel model. The size of the local dataset is $D_k = 600$ at user k.

1) ML Model: We train a convolutional neural network (CNN) with three convolution layers, followed by max pooling. The first convolution layer transforms the single input channel to 32 feature maps using 3×3 kernels (with $32\times 3\times 3\times 1 + 32 = 320$ parameters), followed by ReLU activation and 2×2 max pooling. The second layer expands to 64 feature maps with 3×3 convolutions (with $64\times 3\times 3\times 32 + 64 = 18,496$ parameters), followed by ReLU and 2×2 pooling. The third convolution layer maintains 64 feature maps with 3×3 kernels (with $64\times 3\times 3\times 64 + 64 = 36,928$ parameters), ReLU activation and 2×2 pooling. The network flattens to a fully connected layer with 64 units (with $64\times 64 + 64 = 4,160$ parameters), followed by another fully connected layer with 64 units (with $64\times 64 + 64 = 4,160$ parameters) and ReLU activation. Finally, a linear layer with 10 output units (with $64\times 10 + 10 = 650$ parameters) produces the classification logits for the classes. The total number of trainable parameters is 64,714. Cross-entropy serves as the loss function. The learning rate is 0.1 at the local users and 1 at the server.

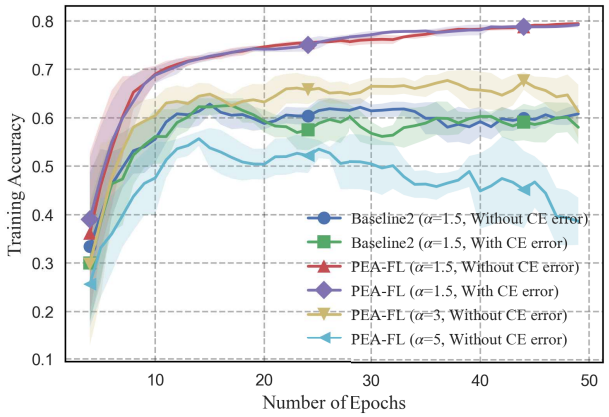
2) Datasets: We consider the MNIST and FASHIONMNIST datasets. In the former, each example is a 28×28 pixel grayscale image of handwritten digits (0-9). In the latter, each example is a 28×28 grayscale image associated with a label from 10 classes of fashion items. Both datasets have a training set of 60,000 examples and a test set of 10,000 examples. We follow the standard train/test splits, with 60,000 training samples and 10,000 testing samples for both MNIST and FashionMNIST datasets, consistent with all baselines. Each communication round consists of five local training epochs.

3) Benchmarks: The first baseline (**Baseline 1**) was developed in [19]. This baseline introduces an FL framework leveraging both OMA and NOMA transmissions with AirComp, harnessing intrinsic channel noise to enhance user privacy. The second baseline (**Baseline 2**) was developed in [30], which enhances an RIS-assisted FL system with a DP mechanism to safeguard user privacy while boosting learning performance. This baseline employs a two-step alternating minimization framework to jointly optimize transmission power, artificial noise, and RIS reflection coefficients. Another baseline (**Baseline 3**) was developed in [43], which is a DP mechanism with time-varying noise amplitude for FL. It also derived the bounds for loss functions and optimize the aggregation number to balance the privacy and utility of FL.

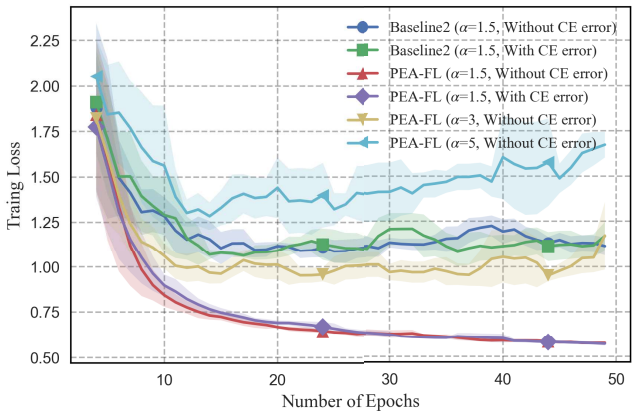
4) Performance Indicators: The privacy leakage function is employed to measure the vulnerability of AirComp to DLG attacks, where a higher leakage value indicates greater privacy exposure, and a lower value indicates stronger privacy preservation. Moreover, the loss function is used to assess the convergence of AirComp, serving as a measure of its effectiveness in achieving the intended learning objectives.

B. Impact of Channel Estimation (CE) Error and α

Figs. 2a and 2b evaluate system performance under SNR= 0 dB and $\epsilon = 5$, with phase noise standard deviation of 0.1 and estimation error standard deviation σ_{error} of 0.05. It is noticed that σ_{error} is approximately 16.67 times larger than the standard deviation of the ideal channel gain (0.003). Nonetheless, the performance of PEA-FL remains largely unaffected under moderate error levels, owing to its inherent gradient alignment mechanism within AirComp. This mechanism exploits coherent signal superposition to partially mitigate the impact of phase distortions. Such resilience enables stable performance in the presence of channel imperfections. However, when $\sigma_{\mathrm{error}} > 0.1$, estimation errors begin to dominate the effective channel characteristics, resulting in significant performance degradation. PEA-FL exhibits significant performance degradation at larger α values in the RDP mechanism, with accuracy dropping from 0.7938 at $\alpha = 1.5$ to 0.52 at $\alpha = 3$ (a reduction of 34.5%) and further to 0.43 at $\alpha = 5$ (a total reduction of 45.8%). This degradation occurs because larger α values require larger noise injections under the same privacy budget, with noise variance approaching that of standard DP as $\alpha \to \infty$, thereby increasingly distorting model updates.

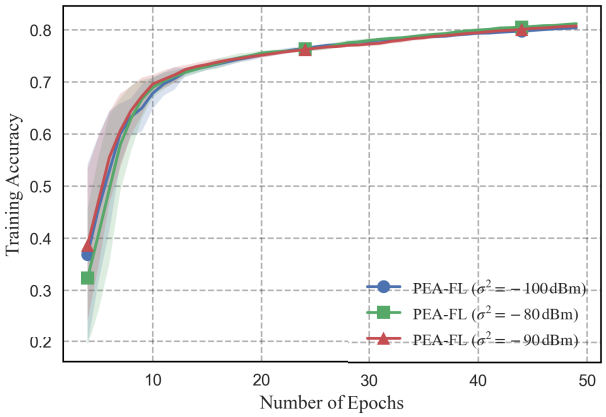


(a) The accuracy of the global model test from FASHIONMNIST under different α and CE error.

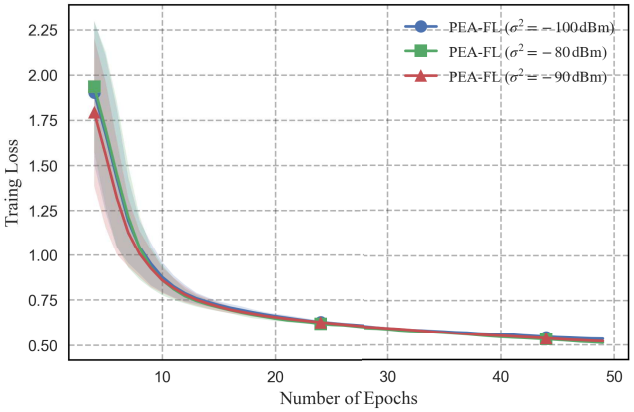


(b) The loss of the global model test from FASHIONMNIST under different α and CE error.

Fig. 2. Different metrics of the considered methods on the different datasets.



(a) The accuracy of the global model test from FASHIONMNIST under different σ^2.



(b) The loss of the global model test from FASHIONMNIST under different σ^2.

Fig. 3. Different metrics of the considered methods on the different σ^2.

C. Impact of Noise Estimation Uncertainty

We experimentally validate our solution's robustness by examining noise variances ranging from −100 dBm to −80 dBm. Figs. 3a and 3b demonstrate that model accuracy remains stable across this 20 dB dynamic range. Less than 1% variation is observed in test accuracy when comparing the minimum and maximum noise power conditions, confirming the robustness to noise estimation uncertainties. For real-time adaptation to changing environments, we can implement a dynamic calibration protocol that adjusts artificial noise parameters based on measured channel conditions. Specifically, we can select the lower bound of noise variance as the baseline, and inject additional artificial noise to ensure privacy guarantees.

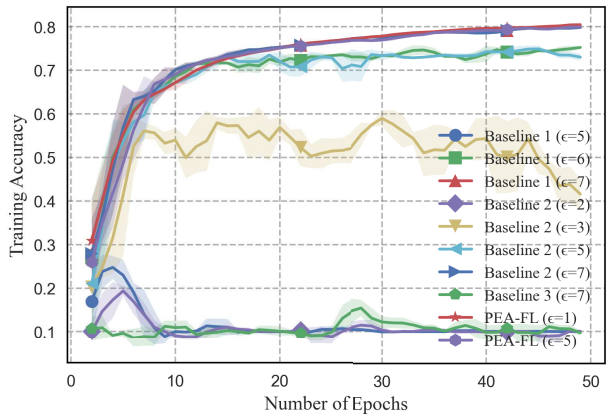
D. Convergence Performance

Fig. 4 compares PEA-FL and the baselines across different datasets (i.e., MNIST and FASHIONMNIST) and privacy levels, where we evaluate model accuracy and loss. In the FASHIONMNIST experiments shown in Figs. 4a and 4b, PEA-FL achieves comparable model convergence and enhanced privacy protection, compared to Baseline 1. Under relaxed privacy constraints (e.g., $\epsilon = 7$), Baselines 1 and 2 achieve comparable convergence performance. Baseline 3 fails to converge even under the relaxed privacy constraints, due to its lack of consideration of channel noise utilization.

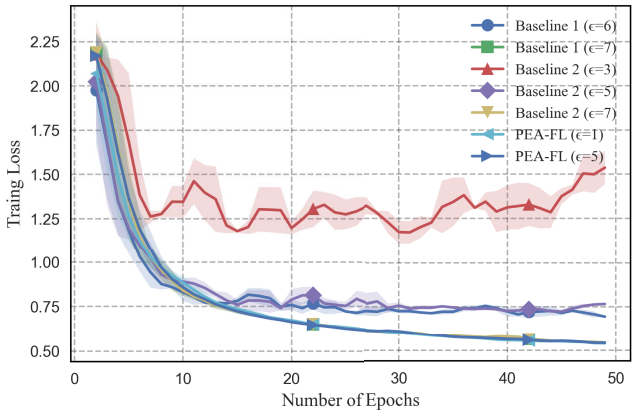
As privacy requirements become more stringent (e.g., $\epsilon = 5$ or 6), Baseline 1 experiences slight accuracy degradation when $\epsilon = 6$ or fails to converge when $\epsilon = 5$; Baseline 2 maintains favorable loss performance. Under strict privacy settings (e.g., $\epsilon = 1$ or 2), Baseline 2 experiences accuracy degradation when $\epsilon = 2$, or fails to converge when $\epsilon = 1$. By contrast, PEA-FL exhibits robust performance and convergence. The MNIST experiments shown in Figs. 4c and 4d further validate the effectiveness of PEA-FL.

E. Impact of Privacy Level

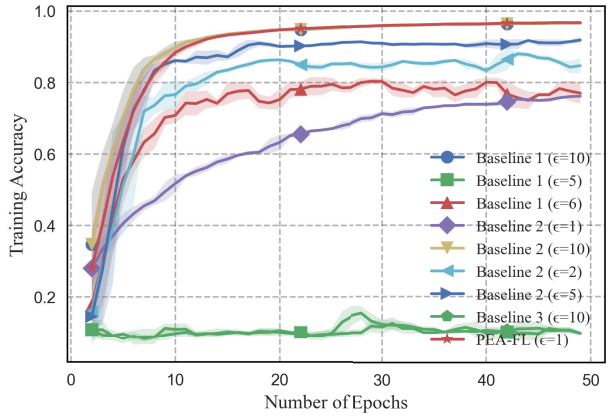
Fig. 5 compares PEA-FL and the baselines, where $1 \leq \epsilon \leq 6$. As ϵ increases, both PEA-FL and Baseline 2 display gradually declining model losses. In contrast, Baseline 1 undergoes a rapid drop from an initial high loss of 1, primarily due to the additional noise injection required to achieve



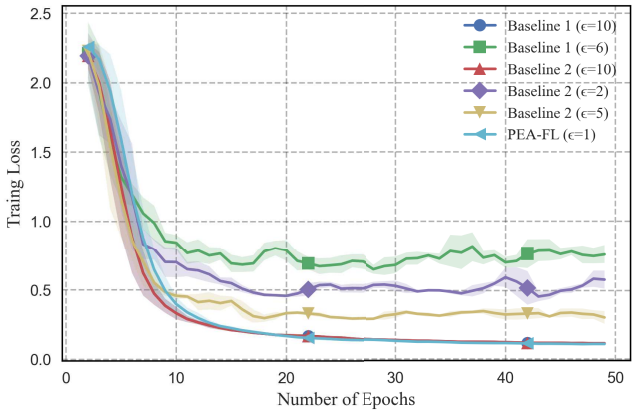
(a) The accuracy of the global model test from FASHIONMNIST under different privacy levels.



(b) The loss of the global model test from FASHIONMNIST under different privacy levels.



(c) The accuracy of the global model test from MNIST under different privacy levels.



(d) The loss of the global model test from MNIST under different privacy levels.

Fig. 4. Different metrics of the considered methods on the different datasets.

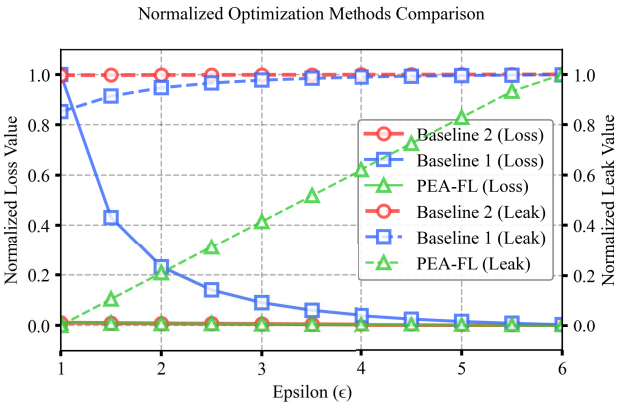


Fig. 5. Model loss and privacy leakage vs. privacy level.

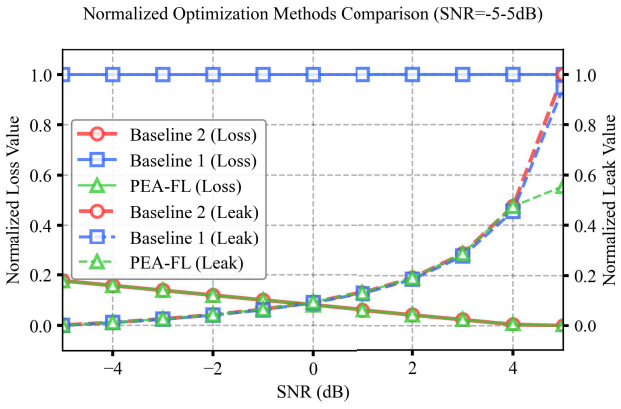
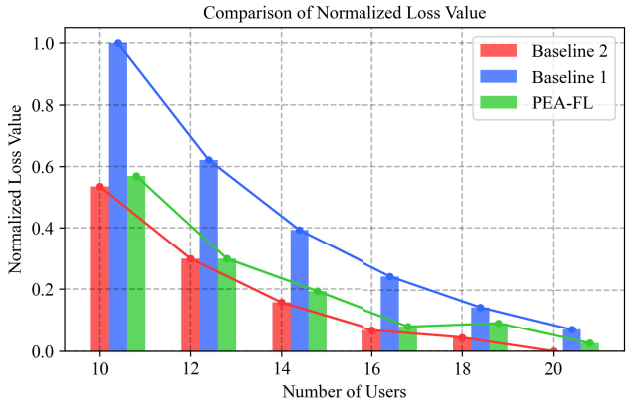


Fig. 6. Model loss and privacy leakage under different SNRs.

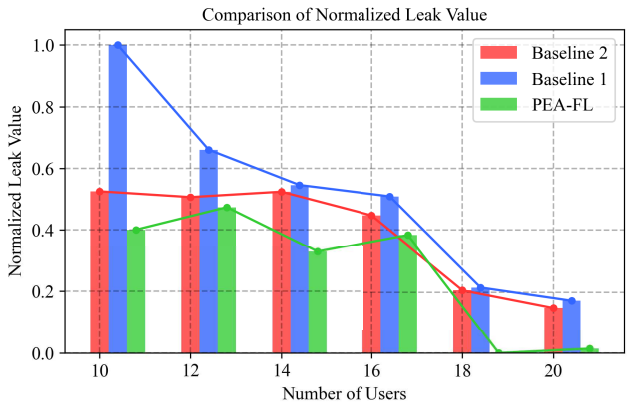
comparable privacy protection. PEA-FL achieves remarkably low privacy leakage at $\epsilon = 1$, compared to Baselines 1 and 2. As privacy constraints are relaxed, the privacy leakage gradually increases, until convergence with those of Baselines 1 and 2. Moreover, PEA-FL exhibits robust learning performance under stringent privacy requirements. When $\epsilon = 1$, the algorithm achieves comparable model loss with Baseline 2 while reducing the privacy leakage by 80%.

F. Impact of SNR

Fig. 6 illustrates the performance comparison between the proposed PEA-FL and the baselines in model loss and privacy



(a) Model loss under different numbers of users.



(b) Privacy leakage under different numbers of users.

Fig. 7. Model loss and privacy leakage vs. the number of users.

leakage, where the SNR ranges from −5 dB to 5 dB and $\epsilon = 5$. Under low SNRs, comparable privacy leakage performances are noticed across all three schemes, attributed to the inherent channel noise, which perturbs the transmitted signals. Under high SNRs, Baselines 1 and 2 demonstrate inferior privacy protection due to their lack of consideration of DLG attacks. By contrast, PEA-FL achieves significantly lower privacy leakage than these two baselines with a 40% reduction in leakage value at the SNR of 5 dB.

This substantial improvement of PEA-FL in privacy protection can be attributed to two key factors. On the one hand, PEA-FL leverages AirComp's waveform superposition property to introduce advantageous noise during model aggregation. On the other hand, it inherently accounts for potential DLG attacks by active adversaries, guiding the design of robust artificial noise injection mechanisms. Moreover, PEA-FL demonstrates comparable convergence performance to Baseline 2 under low SNR conditions. As the SNR increases, PEA-FL exhibits significantly improved convergence and outperforms Baseline 1 by approximately 80%.

G. Impact of User Number

Fig. 7 presents the system behavior with a varying number of users, where we evaluate two metrics: model loss and privacy leakage. The number of users ranges from 10 to 20. Fig. 7a shows that PEA-FL only incurs a marginal increase in model loss (about 2% higher) across different numbers of users while maintaining approximately 62.86% lower loss than Baseline 1. Notably, with 20 users, the algorithm maintains nearly identical loss performance while substantially enhancing privacy protection. On the other hand, Fig. 7b shows that PEA-FL consistently maintains the lowest privacy leakage across different user numbers. When the number of users reaches 20, our approach reduces privacy leakage by 20% compared to Baselines 1 and 2.

While Figs. 5 and 6 display the superiority of PEA-FL across different SNRs and privacy levels, Fig. 7 further validates its scalability and robustness with varying numbers of participating users. This comprehensive evaluation across multiple dimensions (i.e., privacy level, loss value, and user number) substantiates the practical applicability and effectiveness of our privacy-preserving AirComp FL framework.

VI. Conclusion

In this paper, we have proposed a new AirComp-FL system that leverages the waveform superposition and channel propagation to balance model accuracy and privacy. We have derived a mathematical model for privacy leakage, and established convergence upper bounds to account for errors induced by noise. Accordingly, optimal power scaling and noise parameters have been obtained through joint privacy-accuracy optimization. Simulations have validated the proposed framework, demonstrating that under stringent privacy requirements, the AirComp-FL system reduces privacy leakage by up to 80% compared to baselines. The system exhibited superior robustness under low SNRs, achieving up to 40% lower privacy leakage under given privacy budgets.

Appendix
Proof of Theorem Theorem 1

Based on (22), we use $\Delta_t = \tilde{\boldsymbol{g}}_t + \boldsymbol{z}_t$ to represent the estimated global update. Under the smoothness assumption in Assumption 1, taking the expectation of $F\left(\boldsymbol{w}_{t+1}\right)$ at communication round t, we have

$$\begin{aligned}\mathbb{E}_t[F(\boldsymbol{w}_{t+1})] &\leq F\left(\boldsymbol{w}_t\right) + \left\langle \nabla F\left(\boldsymbol{w}_t\right), \mathbb{E}_t[\boldsymbol{w}_{t+1} - \boldsymbol{w}_t] \right\rangle \\ &\quad + \frac{L}{2}\mathbb{E}_t\left[\left\|\boldsymbol{w}_{t+1} - \boldsymbol{w}_t\right\|^2\right] \\ &= F\left(\boldsymbol{w}_t\right) - \eta\eta_t E \left\|\nabla F(\boldsymbol{w}_t)\right\|^2 + \frac{L}{2}\eta^2 \mathbb{E}_t\left[\|\Delta_t\|^2\right] \\ &\quad + \eta\left\langle \nabla F\left(\boldsymbol{w}_t\right), \mathbb{E}_t\left[\Delta_t + \eta_t E \nabla F\left(\boldsymbol{w}_t\right)\right]\right\rangle. \end{aligned} \tag{52}$$

Define $A'_1 = \left\langle \nabla F\left(\boldsymbol{w}_t\right), \mathbb{E}_t\left[\Delta_t + \eta_t E \nabla F(\boldsymbol{w}_t)\right]\right\rangle$, and $A'_2 = \mathbb{E}_t[\|\Delta_t\|^2]$. Next, we bound the term A'_1 as

$$\begin{aligned} A'_1 = \Bigg\langle \sqrt{\eta_t E}\nabla F(\boldsymbol{w}_t), -\frac{\sqrt{\eta_t}}{K\sqrt{E}} \cdot \mathbb{E}_t\Big[\sum_{k\in\mathcal{K}_t} \sum_{e=1}^{E} \left(\nabla f_k(\boldsymbol{w}_{t,e}^k)\right. \\ \left. - \nabla f_k(\boldsymbol{w}_t)\right)\Big] + \frac{1}{\sqrt{\eta_t E}}\boldsymbol{z}_t \Bigg\rangle. \end{aligned} \tag{53}$$

Since $\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \frac{1}{2}\left[\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 - \|\boldsymbol{a} - \boldsymbol{b}\|^2\right]$, we have

$$A'_1 = \frac{\eta_t E}{2}\|\nabla F\left(\boldsymbol{w}_t\right)\|^2 + \frac{\eta_t}{2K^2 E} \times$$

$$\left\|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)-\nabla f_k\left(\boldsymbol{w}_t\right)\right)+\frac{K}{\eta_t}\boldsymbol{z}_t\right]\right\|^2 - \frac{\eta_t}{2K^2E}\left\|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)\right)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\right\|^2. \quad (54)$$

Based on (54) and using $\mathbb{E}_t\left[\|\boldsymbol{x}_1+\ldots+\boldsymbol{x}_n\|^2\right] \le n\mathbb{E}_t\left[\|\boldsymbol{x}_1\|^2+\ldots+\|\boldsymbol{x}_n\|^2\right]$, we have

$$A'_1 \overset{(a)}{\le} \frac{\eta_t E}{2}\|\nabla F(\boldsymbol{w}_t)\|^2+\frac{\eta_t}{K^2E}\mathbb{E}_t\left\|\frac{K}{\eta_t}\boldsymbol{z}_t\right\|^2 + \frac{\eta_t}{K}\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\mathbb{E}_t\|\nabla f_k(\boldsymbol{w}_{t,e}^k)-\nabla f_k(\boldsymbol{w}_t)\|^2 - \frac{\eta_t}{2K^2E}\left\|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\nabla f_k(\boldsymbol{w}_{t,e}^k)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\right\|^2, \quad (55)$$

where (a) also follows from $\mathbb{E}_t\left[\|\boldsymbol{x}_1+\ldots+\boldsymbol{x}_n\|^2\right] \le n\mathbb{E}_t\left[\|\boldsymbol{x}_1\|^2+\ldots+\|\boldsymbol{x}_n\|^2\right]$. Note that $F(\boldsymbol{w})$ is L-Lipschitz smooth. Under Assumption 1, (55) can be converted to

$$A'_1 \le \frac{\eta_t E}{2}\|\nabla F(\boldsymbol{w}_t)\|^2+\frac{\eta_t L^2}{K}\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\mathbb{E}_t\|\boldsymbol{w}_{t,e}^k-\boldsymbol{w}_t\|^2 + \frac{\eta_t}{K^2E}\mathbb{E}_t\left[\|\frac{K}{\eta_t}\boldsymbol{z}_t\|^2\right]-\frac{\eta_t}{2K^2E}\times \|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)\right)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\|^2. \quad (56)$$

Next, we prove (56) by starting with the following lemma.

Lemma 2: [44, Lemma 4]. If $\eta_t < \frac{1}{2\sqrt{30}LE}$, *then*

$$\frac{1}{K}\sum_{k\in\mathcal{K}_t}\mathbb{E}_t\left[\|\boldsymbol{w}_{t,e}^k-\boldsymbol{w}_t\|^2\right] \le 5E\eta_t^2(\sigma_L^2+6E\sigma_G^2) +30E^2\eta_t^2\|\nabla F(\boldsymbol{w}_t)\|^2. \quad (57)$$

Based on Lemma 2 and (56), we have

$$A'_1 \le \eta_t E\left(\frac{1}{2}+30E^2\eta_t^2L^2\right)\|\nabla F(\boldsymbol{w}_t)\|^2+5E^2\eta_t^3L^2\cdot (\sigma_L^2+6E\sigma_G^2)+\frac{\eta_t}{K^2E}\mathbb{E}_t\left[\|\frac{K}{\eta_t}\boldsymbol{z}_t\|^2\right]-\frac{\eta_t}{2K^2E}\cdot \|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)\right)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\|^2. \quad (58)$$

Next, we can bound the term A'_2 as

$$A'_2 \overset{(b)}{\le} \frac{2\eta_t{}^2}{K^2}\mathbb{E}_t\left[\left\|\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\boldsymbol{g}_{t,e}^k\right\|^2+\left\|\frac{K}{\eta_t}\boldsymbol{z}_t\right\|^2\right] + \frac{2\eta_t{}^2}{K^2}\mathbb{E}_t\left[\left\|\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)\right\|^2\right]+2\mathbb{E}_t\|\boldsymbol{z}_t\|^2 \overset{(c)}{\le} \frac{2E\eta_t{}^2}{K}\sigma_L^2+\frac{2\eta_t{}^2}{K^2}\mathbb{E}_t\left[\left\|\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\nabla f_k\left(\boldsymbol{w}_{t,e}^k\right)\right\|^2\right] + 2\mathbb{E}_t\|\boldsymbol{z}_t\|^2, \quad (59)$$

where $\boldsymbol{g}_{t,e}^k$ denotes the gradient which user k uses the training sample $(\boldsymbol{x},y)\in\mathcal{D}_k$ to update its local model in E epochs at the t-th training round. (b) is derived from the arithmetic and geometric mean inequality and the triangle inequality. (c) is because of **Assumption** 2 about bounded variance and $\mathbb{E}_t\left[\|\boldsymbol{x}_1+\ldots+\boldsymbol{x}_n\|^2\right] \le n\mathbb{E}_t\left[\|\boldsymbol{x}_1\|^2+\ldots+\|\boldsymbol{x}_n\|^2\right]$.

Substituting (58) and (59) into (52), it follows that

$$\begin{aligned}\mathbb{E}_t[F(w_{t+1})] &\overset{(d)}{\le} F(\boldsymbol{w}_t)-\eta\eta_t E\left(\frac{1}{2}-30E^2\eta_t^2L^2\right)\|\nabla F(\boldsymbol{w}_t)\|^2 \\ &+\frac{EL\eta^2\eta_t^2}{K}\sigma_L^2+5E^2\eta\eta_t^3L^2(\sigma_L^2+6E\sigma_G^2)+\frac{\eta\eta_t}{K^2E}\mathbb{E}_t\left[\left\|\frac{K}{\eta_t}\boldsymbol{z}_t\right\|^2\right] \\ &-\frac{\eta\eta_t}{2K^2E}\left\|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k(\boldsymbol{w}_{t,e}^k)\right)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\right\|^2 \\ &+\frac{\eta^2\eta_t^2L}{K^2}\mathbb{E}_t\left[\left\|\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k(\boldsymbol{w}_{t,e}^k)\right)\right\|^2\right]+L\eta^2\mathbb{E}_t\|\boldsymbol{z}_t\|^2 \\ &\overset{(e)}{\le} F(\boldsymbol{w}_t)-\eta\eta_t E\left(\frac{1}{2}-30E^2\eta_t^2L^2\right)\|\nabla F(\boldsymbol{w}_t)\|^2 \\ &+\frac{\eta\eta_t}{K^2E}\mathbb{E}_t\left[\left\|\frac{K}{\eta_t}\boldsymbol{z}_t\right\|^2\right]+\frac{\eta^2\eta_t^2L}{2K^2}\mathbb{E}_t\left\|\frac{K}{\eta_t}\boldsymbol{z}_t\right\|^2+L\eta^2\mathbb{E}_t\|\boldsymbol{z}_t\|^2 \\ &-\frac{\eta\eta_t}{2K^2E}\left\|\mathbb{E}_t\left[\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k(\boldsymbol{w}_{t,e}^k)\right)-\frac{K}{\eta_t}\boldsymbol{z}_t\right]\right\|^2 \\ &+\frac{\eta^2\eta_t^2L}{2K^2}\mathbb{E}_t\left[\left\|\sum_{k\in\mathcal{K}_t}\sum_{e=1}^{E}\left(\nabla f_k(\boldsymbol{w}_{t,e}^k)\right)\right\|^2-\frac{K}{\eta_t}\boldsymbol{z}_t\right] \\ &\overset{(f)}{\le} F(\boldsymbol{w}_t)-\eta\eta_t E\left(\frac{1}{2}-30E^2\eta_t^2L^2\right)\|\nabla F(\boldsymbol{w}_t)\|^2+\frac{EL\eta^2\eta_t^2}{K} \\ &\times\sigma_L^2+5E^2\eta\eta_t^3L^2(\sigma_L^2+6E\sigma_G^2)+\left(\frac{I\eta}{\eta_t E}+\frac{3IL\eta^2}{2}\right)\sigma_{z,t}^2,\end{aligned} \quad (60)$$

where (d) originates from $\mathbb{E}[\|\boldsymbol{x}\|^2] = \mathbb{E}[\|\boldsymbol{x}-E[\boldsymbol{x}]\|^2]+\|\mathbb{E}[\boldsymbol{x}]\|^2$, (e) follows from $\mathbb{E}_t\left[\|\boldsymbol{x}_1+\ldots+\boldsymbol{x}_n\|^2\right] \le n\mathbb{E}_t\left[\|\boldsymbol{x}_1\|^2+\ldots+\|\boldsymbol{x}_n\|^2\right]$, and (f) is derived from $\left(\frac{\eta\eta_t}{2K^2E}-\frac{\eta^2\eta_t^2L}{2K^2}\right) \ge 0$ and $\mathbb{E}_t\|\boldsymbol{z}_t\|^2 = I\sigma_{z,t}^2$ with $\sigma_{z,t}^2=\frac{1}{(KD)^2\lambda_t}\sum_{k\in\mathcal{K}_t}\frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2}+\frac{1}{(KDc_t)^2\lambda_t}\sigma_{m,t}^2$.

Rearranging and summing from $t=0,1,\ldots,T-1$ yields

$$\min_{t\in[1,T]}\mathbb{E}\left[\|\nabla F(\boldsymbol{w}_t)\|^2\right] \le \frac{F_0-F_*}{c'E\eta\eta_t T}+\Phi$$

$$\Phi=\frac{1}{c'}\left[\frac{L\eta\eta_t}{K}\sigma_L^2+5E\eta_t^2L^2\left(\sigma_L^2+6E\sigma_G^2\right)\right]+\frac{I}{E\eta_t c'}\cdot \left(\frac{1}{\eta_t E}+\frac{3L\eta}{2}\right)\frac{1}{(KD)^2\lambda_t}\left(\sum_{k\in\mathcal{K}_t}\frac{\sigma_{k,t}^2}{\upsilon_{k,t}^2}+\frac{1}{(c_t)^2}\sigma_{m,t}^2\right), \quad (61)$$

where c' is a constant and satisfies $\left(\frac{1}{2}-30E^2\eta_t^2L^2\right)>c'>0$.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 60, 2017, pp. 84–90.

[2] J. Zheng, K. Li, N. Mhaisen, W. Ni, E. Tovar, and M. Guizani, "Exploring deep-reinforcement-learning-assisted federated learning for online resource allocation in privacy-preserving EdgeIoT," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21099–21110, Nov. 2022.

[3] B. Yang et al., "Edge intelligence for autonomous driving in 6G wireless system: Design challenges and solutions," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 40–47, Apr. 2021.

[4] Y. Wan, Y. Qu, W. Ni, Y. Xiang, L. Gao, and E. Hossain, "Data and model poisoning backdoor attacks on wireless federated learning, and the defense mechanisms: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1861–1897, 3rd Quart., 2024.

[5] K. Li, J. Zheng, X. Yuan, W. Ni, O. B. Akan, and H. V. Poor, "Data-agnostic model poisoning against federated learning: A graph autoencoder approach," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 3465–3480, 2024.

[6] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[7] B. Xiao, X. Yu, W. Ni, X. Wang, and H. V. Poor, "Over-the-air federated learning: Status quo, open challenges, and future directions," *Fundam. Res.*, vol. 5, no. 4, pp. 1710–1724, Jul. 2025.

[8] X. Yu, B. Xiao, W. Ni, and X. Wang, "Optimal adaptive power control for over-the-air federated edge learning under fading channels," *IEEE Trans. Commun.*, vol. 71, no. 9, pp. 5199–5213, Sep. 2023.

[9] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, Jan. 2020.

[10] J. Zheng, H. Tian, W. Ni, W. Ni, and P. Zhang, "Balancing accuracy and integrity for reconfigurable intelligent surface-aided over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10964–10980, Dec. 2022.

[11] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[12] 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.

[13] L. Lyu et al., "Privacy and robustness in federated learning: Attacks and defenses," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 8726–8746, Jul. 2024.

[14] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1333–1345, May 2018.

[15] C. Li, S. Long, H. Liu, Y. Choi, H. Sekiya, and Z. Li, "Enhancing sparse mobile CrowdSensing with manifold optimization and differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 6070–6083, 2024.

[16] G. Yu et al., "IronForge: An open, secure, fair, decentralized federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 1, pp. 354–368, Jan. 2025.

[17] X. He, L. Li, H. Peng, and F. Tong, "An efficient image privacy preservation scheme for smart city applications using compressive sensing and multi-level encryption," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 10, pp. 14958–14972, Oct. 2024.

[18] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.

[19] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 170–185, Jan. 2021.

[20] G. Zhu, Y. Du, D. Gündüz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2120–2135, Mar. 2021.

[21] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.

[22] P. Sun et al., "Reconfigurable intelligent surface-assisted wireless federated learning with imperfect aggregation," *IEEE Trans. Commun.*, vol. 73, no. 2, pp. 1058–1071, Feb. 2024.

[23] J. Du, T. Lin, C. Jiang, Q. Yang, C. F. Bader, and Z. Han, "Distributed foundation models for multi-modal learning in 6G wireless networks," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 20–30, Jun. 2024.

[24] X. Liu, H. Li, G. Xu, S. Liu, Z. Liu, and R. Lu, "PADL: Privacy-aware and asynchronous deep learning for IoT applications," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6955–6969, Aug. 2020.

[25] M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security*, 2016, pp. 308–318.

[26] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[27] N. Wu, F. Farokhi, D. Smith, and M. A. Kaafar, "The value of collaboration in convex machine learning with differential privacy," in *Proc. IEEE Symp. Secur. Privacy (SP)*, San Francisco, CA, USA, May 2020, pp. 304–317.

[28] N. Agarwal, A. T. Suresh, F. Yu, S. Kumar, and H. B. McMahan, "CpSGD: Communication-efficient and differentially-private distributed SGD," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 7575–7586.

[29] B. Hasircioglu and D. Gündüz, "Private wireless federated learning with anonymous over-the-air computation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 5195–5199.

[30] Y. Yang, Y. Zhou, Y. Wu, and Y. Shi, "Differentially private federated learning via reconfigurable intelligent surface," *IEEE Internet Things J.*, vol. 9, no. 20, pp. 19728–19743, Oct. 2022.

[31] T. Liu, B. Di, P. An, and L. Song, "Privacy-preserving incentive mechanism design for federated cloud-edge learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2588–2600, Jul. 2021.

[32] T. Liu, B. Di, B. Wang, and L. Song, "Loss-privacy tradeoff in federated edge learning," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 546–558, Apr. 2022.

[33] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 8792–8802.

[34] I. Mironov, "Renyi differential privacy," in *Proc. IEEE 30th Comput. Secur. Found. Symp. (CSF)*, Aug. 2017, pp. 263–275.

[35] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, vol. 54, A. Singh and J.Zhu., Eds., Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

[36] L. Zhu, X. Liu, Y. Li, X. Yang, S.-T. Xia, and R. Lu, "A fine-grained differentially private federated learning against leakage from gradients," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 11500–11512, Jul. 2022.

[37] A. Cheng, P. Wang, X. S. Zhang, and J. Cheng, "Differentially private federated learning with local regularization and sparsification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10122–10131.

[38] X. Yang, W. Huang, and M. Ye, "Dynamic personalized federated learning with adaptive differential privacy," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 72181–72192.

[39] H. Liu, X. Yuan, and Y.-J. A. Zhang, "Reconfigurable intelligent surface enabled federated learning: A unified communication-learning design approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7595–7609, Nov. 2021.

[40] W. Ni, R. P. Liu, J. Biswas, X. Wang, I. B. Collings, and S. K. Jha, "Multiuser MIMO scheduling for mobile video applications," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5382–5395, Oct. 2014.

[41] F. A. Potra and S. J. Wright, "Interior-point methods," *J. Comput. Appl. Math.*, vol. 124, nos. 1–2, pp. 281–302, 2000.

[42] K. Koh, S. J. Kim, and S. P. Boyd, "An interior-point method for large-scale l_1-regularized logistic regression," *J. Mach. Learn. Res.*, pp. 1519–1555, 2007.

[43] X. Yuan, W. Ni, M. Ding, K. Wei, J. Li, and H. V. Poor, "Amplitude-varying perturbation for balancing privacy and utility in federated learning," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1884–1897, 2023.

[44] S. Reddi et al., "Adaptive federated optimization," 2020, *arXiv:2003.00295*.

Hexin Feng (Student Member, IEEE) was born in China in 1996. He is currently pursuing the Ph.D. degree with the College of Electronics and Information Engineering, Tongji University, Shanghai, China. His research interests include privacy-preserving federated learning and decentralized federated learning.

Rui Wang (Senior Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, China, in 2013.

From 2012 to 2013, he was a Visiting Ph.D. Student with the University of California at Riverside. From 2013 to 2014, he was a Post-Doctoral Research Associate with the Institute of Network Coding, CUHK. From 2014 to 2016, he was an Assistant Professor with Tongji University, where he is currently a Full Professor. He has published more than 60 articles. His research interests include wireless communications, artificial intelligence, and wireless positioning. He received Shanghai Excellent Doctor Degree Dissertation Award in 2015 and the ACM Shanghai Rising Star Nomination Award in 2016. He is an Associate Editor of IEEE ACCESS and an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS.

Erwu Liu (Senior Member, IEEE) received the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2001. From 2001 to 2007, he was with Alcatel-Lucent Bell Laboratories. From 2007 to 2011, he was with the Imperial College London. He is currently a Professor with Tongji University. He leads the PAI Research Center with Tongji University and Shanghai Engineering Research Center for Blockchain Applications and Services (SERCBAAS). He is an IET Fellow. He won the Microsoft Indoor Localization Competition (IPSN) in 2016 and 2018, and developed the indoor navigation system for China International Import Expo (CIIE). He is the TPC Co-Chair of the IEEE Future Networks World Forum (FNWF) and the Community Development Co-Chair of the IEEE Blockchain Technical Community (BCTC), and leads the local group development of IEEE BCTC in Asia/China. He is also the Founding Chair of the IEEE Global Blockchain Conference (GBC). He is the Founding Editor-in-Chief of *IET Blockchain*.

Wei Ni (Fellow, IEEE) received the B.E. and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He was the Deputy Project Manager with Bell Laboratories, Alcatel/Alcatel-Lucent, from 2005 to 2008; a Senior Research Engineer with Nokia from 2008 to 2009; and a Senior Principal Research Scientist and the Group Leader with the Commonwealth Scientific and Industrial Research Organization (CSIRO) from 2009 to 2025. He is currently the Associate Dean (Research) with the School of Engineering, Edith Cowan University, Perth, and a Conjoint Professor with the University of New South Wales, Sydney, Australia. He served as the Secretary, the Vice-Chair, and the Chair for the IEEE VTS NSW Chapter from 2015 to 2022, the Track Chair for VTC-Spring 2017, the Track Co-Chair for IEEE VTC-Spring 2016, the Publication Chair for BodyNet 2015, and the Student Travel Grant Chair for WPMC 2014. He has been an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2018; IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY since 2022; IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS since 2024; and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING since 2025.

Dusit Niyato (Fellow, IEEE) received the B.E. degree from the King Mongkuk's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a President's Chair Professor with the College of Computing and Data Science (CCDS), Nanyang Technological University, Singapore. His research interests include distributed collaborative machine learning, the Internet of Things (IoT), edge intelligent generative AI and AI-generated content (AIGC), mobile and distributed computing, and wireless networks. He is a fellow of IET. He was a recipient of the IEEE Vehicular Technology Society Stuart Meyer Memorial Award, the IEEE Communications Society (ComSoc) Best Survey Paper Award, and the IEEE Asia–Pacific Board (APB) Outstanding Paper Award. He is serving as the Editor-in-Chief for the IEEE Communications Surveys and Tutorials (Impact Factor of 34.4 for 2023) and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He is an Area Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE WIRELESS COMMUNICATIONS, *IEEE Network*, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE DATA DESCRIPTIONS, IEEE TRANSACTIONS ON SERVICES COMPUTING, *IEEE Communications Magazine*, and *ACM Computing Surveys*. He was also a Guest Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is the Members-at-Large of the Board of Governors of the IEEE Communications Society for 2024–2026. He was named the 2017–2023 Highly Cited Researcher in computer science.

Abbas Jamalipour (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Nagoya University, Nagoya, Japan, in 1996. He is currently a Professor of ubiquitous mobile networking with The University of Sydney. He has authored nine technical books, 11 book chapters, more than 550 technical articles, and five patents, all in the area of wireless communications and networking. He is a fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electrical, Information, and Communication Engineers (IEICE), and the Institution of Engineers Australia, an ACM Professional Member, and an IEEE Distinguished Speaker. He was a recipient of the number of prestigious awards, such as the 2019 IEEE ComSoc Distinguished Technical Achievement Award in Green Communications, the 2016 IEEE ComSoc Distinguished Technical Achievement Award in Communications Switching and Routing, the 2010 IEEE ComSoc Harold Sobol Award, the 2006 IEEE ComSoc Best Tutorial Paper Award, and more than 15 best paper awards. He has been the General Chair or Technical Program Chair for several prestigious conferences, including IEEE ICC, GLOBECOM, WCNC, and PIMRC. He was the President of the IEEE Vehicular Technology Society from 2020 to 2021. Previously, he held the positions of the Executive Vice-President and the Editor-in-Chief of VTS Mobile World and has been an elected member of the Board of Governors of the IEEE Vehicular Technology Society since 2014. He was the Editor-in-Chief of IEEE WIRELESS COMMUNICATIONS, the Vice President-Conferences, and a member of Board of Governors of the IEEE Communications Society. He sits on the Editorial Board of IEEE ACCESS and several other journals and is a member of the Advisory Board of IEEE INTERNET OF THINGS JOURNAL. Since January 2022, he has been the Editor-in-Chief of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.