This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

Secure and Private Over-the-air Federated Learning: Biased and Unbiased Aggregation Design

Na Yan, Kezhi Wang, Kangda Zhi, Cunhua Pan, Kok Keong Chai, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—Over-the-air federated learning (OTA-FL) presents a promising paradigm that improves the efficiency of local update aggregation by leveraging the superposition property of wireless multiple access channels (MACs). However, it faces significant security and privacy concerns that demand careful consideration. To address these threats associated with OTA-FL, we develop a secure and private over-the-air federated learning (SP-OTA-FL) framework, which can realize the secure and private aggregation for both OTA-FL with unbiased aggregation (UB-OTA-FL) and OTA-FL with biased aggregation (B-OTA-FL). In this framework, a subset of devices participate in training, while another subset functions as jammers, emitting jamming signals to enhance the security and privacy of the OTA-FL process. In particular, we measure the privacy leakage of users' data using differential privacy (DP) and introduce an innovative application of mean squared error security (MSE-security) to evaluate the security of the OTA-FL system. We conduct convergence analyses for both convex and non-convex loss functions. Building on these analytical results, we separately formulate optimization problems for UB-OTA-FL and B-OTA-FL to enhance the learning performance of SP-OTA-FL by strategically optimizing the scheduling of training participants and jammers. The effectiveness of the proposed schemes is verified through simulations.**

***Index Terms*—Federated learning (FL), differential privacy (DP), mean square error security (MSE-security).**

I. Introduction

The ever-growing volume of valuable data generated at the edge of wireless networks has paved the way for many artificial intelligence (AI) services catering to end-users by exploiting deep learning [1]. In various applications like the Internet of Things (IoT) and unmanned aerial vehicles (UAVs), data from sensors normally needs to be constantly collected and processed. The development and refinement of machine learning (ML) algorithms play a crucial role in extracting meaningful knowledge, improving system performance, and unlocking the full potential of the collected data. Traditionally, many of these ML algorithms have followed a centralized approach, wherein all raw data is transmitted to and aggregated at a powerful central server for model training [2]. However, centralized ML approaches become less practical and desirable due to privacy concerns and limitations related to latency, bandwidth, and power. This is particularly evident as the concerns regarding privacy and the volume of data grow. To address these challenges, decentralized and privacy-enhancing approaches have gained significant attention.

Federated learning (FL) [3] is one of the techniques that enable intelligent edges to train learning models collaboratively without uploading their local data to the server. By training models locally, FL not only makes full use of the computing power of the edge devices, but also effectively reduces the power consumption, latency, and privacy exposure due to the transmission of raw data. However, despite these promising benefits, wireless FL still faces significant challenges in the form of communication bottlenecks. These bottlenecks arise when a large number of devices attempt to upload gradients via a resource-limited wireless multiple access channel (MAC). This leads to considerable upload latency as the allocated bandwidth for each participant decreases with the increasing number of devices involved [3], [4].

Over-the-air FL(OTA-FL) [5], [6] emerges as a promising solution for enhancing the wireless aggregation of FL, capitalizing on the remarkable communication efficiency of over-the-air computation (AirComp). OTA-FL allows devices to upload their local updates simultaneously through a MAC, where gradients are typically transmitted via analog signals, enabling the central controller, often a base station (BS), to directly receive an aggregated gradient through waveform superposition. In OTA-FL, the assigned bandwidth for each participant is independent of the number of devices. This feature ensures high energy and bandwidth efficiency, particularly when dealing with a large number of devices [5], [7]. However, analog OTA-FL faces privacy and security challenges arising from the wireless nature of communication and the decentralized nature of data processing.

Although FL provides basic privacy preservation by avoiding the sharing of raw data, numerous studies [8]–[13] have demonstrated that privacy concerns still exist due to attacks targeting exchanged gradients. These concerns refer to the risks associated with the unintended exposure or

Na Yan, Kangda Zhi and Kok Keong Chai are with School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: n.yan, k.zhi, michael.chai@qmul.ac.uk). Kezhi Wang is with Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, U.K. (email: kezhi.wang@brunel.ac.uk). Cunhua Pan is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (email: cpan@seu.edu.cn). H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA. (email: poor@princeton.edu)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

inference of sensitive information contained in local model updates or gradients during the aggregation and transmission processes. For instance, [10] proposed an optimization algorithm capable of recovering both raw inputs and labels by attacking gradients. Despite the aggregation of gradients before reaching the BS in OTA-FL, the possibility of privacy leakage still exists [14]–[20]. One notable risk arises when a single participant contributes to a communication round, allowing attackers to access individual gradients. In this case, by leveraging well-established techniques [8]–[11], attackers can infer sensitive properties of the training data, thereby compromising privacy. In our considered synchronous FL scenarios, such a case can be caused by some device selection algorithms [21], [22], where only one device satisfies the selection criteria due to its outstanding capabilities such as channel quality, power constraints, or computational capacity. Such a case may also occur in certain cluster-based asynchronous FL approaches [23]–[25] where devices within each cluster update their models or gradients synchronously, while updates between clusters occur asynchronously [1]. In this context, a device with notably higher computational power can complete its local training much faster than other devices. To prevent this fast device from being delayed by slower devices, it maybe placed in a cluster by itself. As a result, the device operates independently, participating alone in the aggregation process. If an attacker identifies this device as the sole member of its cluster and persistently monitors its communication with the BS, it can repeatedly observe and analyze the gradients transmitted. Over time, this enables the attacker to continuously extract sensitive information from the gradients, ultimately resulting in significant privacy leakage. Furthermore, even in a more common and pratical scenario involving "multiple participants", where the attacker only has access to the aggregated global model and not the individual local models or their sources, [26] have demonstrated that the adversary is able to extract individual data records from federated NLP models as well as the device identity. To counter these privacy risks effectively, one viable strategy is to employ differential privacy (DP) [27]. It helps safeguard individual privacy in FL by introducing a controlled amount of random noise into the local update to randomize the disclosed statistics, ensuring that the contribution of any individual data sample to the model remains statistically indistinguishable. In [17], artificial Gaussian noise was added to each gradient before transmitting if channel noise cannot provide sufficient privacy protection and a static power allocation scheme was proposed to determine the scale of the artificial noise. Instead of introducing artificial noise, the work of [18] proposed a more energy-efficient strategy to guarantee DP by adjusting the transmit power. The authors of [19] investigated differentially private FL in both orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) channels and proposed adaptive power allocation schemes. The aforementioned studies primarily focused on the scenarios of OTA-FL with unbiased aggregation (UB-OTA-FL) [6], [28], where gradients are aligned by a constant coefficient during the aggregation. In this way, the impact of the fading channel becomes a constant and can be easily removed by performing an inverse operation of the pre-processing at the BS. However, this coefficient is limited by the device with the poorest channel condition due to the peak transmit power constraint. Consequently, this limitation can result in a notably poor signal-to-noise ratio (SNR) [17]. The authors in [29] have endeavored to address the limitation on the coefficient to enhance learning performance. That was achieved by optimizing participant selection, and carefully examining the delicate balance between the number of scheduled devices and the coefficient. Another approach to address the issue of low SNR is the OTA-FL with biased aggregation (B-OTA-FL) [30], [31], where gradients are aggregated using different weights. While this approach may lead to a biased gradient estimate, it has the potential to improve the overall SNR. A differentially private B-OTA-FL (DP-B-OTA-FL) was studied in [31] where the power allocation of the gradient and the added noise were optimized. However, it is worth noting that these studies primarily addressed privacy risks without placing adequate emphasis on security concerns.

In practical FL scenarios, the broadcast nature of wireless channels exposes FL to eavesdropping attacks, which is often the initial step for malicious third parties to launch security attacks. The adversary can then target various types of personal information using model inversion attacks [11], membership attacks [8], attribute inference attacks [32], or replace the exchanged models with malicious models. Hence, it is important to bolster the security of FL communications. In pursuit of this goal, the authors of [33], [34] adopted a covert communication (CC) technique with which a friendly jammer transmits jamming signals to prevent an eavesdropper from detecting the update transmission of the local model from mobile devices in FL. The work of [35] utilized power control to improve the security of FL in the internet of drones (IoD) networks where security rate was employed to measure the security of wireless communications. In [36], [37], homomorphic encryption was employed as an effective measure to secure the process of FL. However, it should be emphasized that these security measures in FL predominantly focus on safeguarding and protecting digitally coded FL from eavesdropping. To the best of our knowledge, the security of analog OTA-FL over a wiretap channel has remained largely unexplored, presenting a significant research gap within this field. Analog signals are often considered more vulnerable to attacks compared to digital signals due to a lack of error detection and correction and challenges in encryption. They can be intercepted easily through methods like eavesdropping. Eavesdropping on the aggregated gradients exchanged between the BS and the

[1]This example is provided to enhance the motivation for considering privacy risk issues. While our work focuses on synchronous aggregation process in this paper, the underlying theoretical insights and framework are adaptable to asynchronous scenarios, as detailed in Section VII.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

participant not only poses the risk of leaking sensitive data, but also exposes information about the model structure or parameters, which could potentially empower attackers to launch more severe security attacks. Moreover, eavesdropping also enables attackers to obtain model updates without contributing, effectively benefiting from improvements at no cost. Hence, investigating the security of analog transmission is both highly challenging and of great significance, particularly given the growing prevalence of over-the-air computation (AirComp), which typically adopts analog transmission, offering a promising solution for alleviating latency issues and improving energy efficiency.

A. Contributions

Inspired by these identified research gaps, this paper introduces a secure and private OTA-FL (SP-OTA-FL) framework. This framework leverages jamming and channel noise as a dual-purpose countermeasure: firstly, to prevent privacy leakage at the "honest but curious" BS, and secondly, to mitigate the security risk associated with eavesdropping on analog OTA-FL. Notably, our study represents the first effort in evaluating the security of analog OTA-FL. In this framework, a subset of devices actively engages in the training, while another subset serves as jammers, transmitting jamming signals to enhance the security and privacy of the OTA-FL process. The device scheduling and transmission design (i.e, power scaling factor and the post-processing factor) are separately optimized for two typical scenarios, i.e., UB-OTA-FL and B-OTA-FL. The main contributions of the paper are summarized as follows:

- ***Novel SP-OTA-FL framework.*** We present an innovative SP-OTA-FL framework, initiating the exploration of security in analog OTA-FL. This framework offers dual protection by preventing privacy breaches at the BS and mitigating the security threat posed by eavesdropping. This is achieved by jointly designing the scheduling of both learning participants and jammers. Notably, it can enhance security in a more effective manner by strategically designating devices closer to potential eavesdroppers as jammers rather than relying on adding noise or reducing the power of the gradient transmission, which could result in a more distorted aggregated gradient at the BS.
- ***Theoretical analysis of privacy, security, and learning performance.*** To evaluate the impact of device scheduling and transmission design on the privacy and security of OTA-FL, we perform privacy and security analyses utilizing DP and mean squared error security (MSE-security) introduced in [38]. Then, we derive closed-form expressions for the optimality gap and the average-squared gradient to demonstrate the convergence for SP-OTA-FL in the cases of convex and non-convex loss functions, respectively. These results characterize how the design of the learning participants, jammers, post-processing factor, and power scaling factor (in UB-OTA-FL scenario) can affect both the privacy and security protection, as well as the overall performance of SP-OTA-FL.
- ***Optimal design for UB-OTA-FL.*** We formulate an optimization problem with the aim of improving the learning performance by optimizing the power scaling factor, the post-processing factor, and the device scheduling of learning participants and jammers in UB-OTA-FL. We derive a closed-form solution for the power scaling factor and establish the relationship between the post-processing factor and the device scheduling of learning participants with a given jammer set. This enables us to effectively find the optimal solution within a limited search space. A sequential update (SU) algorithm is developed to address the scheduling of jammers. Furthermore, we introduce a security and privacy-guided successive jammer removal (SP-SJR) algorithm, which finds the optimal jammer set with given learning participants and transmission design. Building upon the SP-SJR algorithm, we develop a low-complexity (LC) scheme to tackle this problem effectively.
- ***Optimal design for B-OTA-FL.*** In B-OTA-FL, we optimize the post-processing factor and device scheduling for learning participants and jammers while transmitting the gradient with the maximum transmission power. We derive a closed-form solution for the post-processing factor by initially specifying the learning participants and jammers. Subsequently, we derive the optimal solution by employing the SU algorithm and alternate optimization (AO) techniques. Specifically, we employ SU algorithm to enumerate potential optimal jammer sets, and then we employ AO to identify the optimal learning participants and the post-processing factor corresponding to the specified jammer sets. Furthermore, we present an efficient LC scheme based on SP-SJR algorithm to tackle the problem in B-OTA-FL.

II. System Model and Preliminaries

In this section, we introduce the proposed SP-OTA-FL and provide the definitions of DP and MSE-security.

A. SP-OTA-FL framework

In the considered system, there are N edge devices, denoted by the set $\mathcal{N} = \{1, 2, \cdots, N\}$, collaborating to train a deep neural network model with the assistance of a BS. Each device of index $n \in \mathcal{N}$ is assumed to have a local dataset $\mathcal{D}_n$ which contains D_n pairs of training samples $(\boldsymbol{u}, v)$ where $\boldsymbol{u}$ is the raw data and v is the corresponding label. For simplicity, we assume that $D_1 = \cdots = D_N$[2].

1) General FL: The purpose of an FL task is to obtain the model parameter that can minimize the loss function. Mathematically, the goal of the learning can be given as:

$$\min_{\boldsymbol{m}} L(\boldsymbol{m}) = \frac{1}{N} \sum_{n=1}^{N} L_n(\boldsymbol{m}), \tag{1}$$

[2]The assumption that the local datasets of the devices are equivalent makes it easier to focus on the core mechanisms of the SP-OTA-FL framework and the fundamental characteristics of the learning process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

where $\boldsymbol{m} \in \mathbb{R}^d$ is the model parameter to be optimized. More specifically, the objective function of device n is:

$$L_n(\boldsymbol{m}) = \frac{1}{D_n} \sum_{(\boldsymbol{u},v)\in\mathcal{D}_n} l(\boldsymbol{m};(\boldsymbol{u},v)), \tag{2}$$

where $l(\boldsymbol{m};(\boldsymbol{u},v))$ is an empirical loss function defined by the learning task, quantifying the loss of $\boldsymbol{m}$ at sample $(\boldsymbol{u},v)$.

To address problem (1), a typical approach known as gradient descent (GD) can be applied[3]. In the context of FL, the initial step involves the BS selecting a subset of devices to actively participate in the training round. Taking round t as an example, the BS broadcasts the latest global model parameter $\boldsymbol{m}^t$ to these scheduled devices. Then, each participant utilizes the received global model parameter to initialize its local model parameter i.e., $\boldsymbol{m}_n^t = \boldsymbol{m}^t$. Subsequently, each device individually computes the gradient using its local dataset, executing the process of:

$$\boldsymbol{g}_n^t \triangleq \nabla L_n\left(\boldsymbol{m}_n^t\right) = \frac{1}{D_n} \sum_{(\boldsymbol{u},v)\in\mathcal{D}_n} \nabla l\left(\boldsymbol{m}_n^t;(\boldsymbol{u},v)\right). \tag{3}$$

Then, the scheduled devices simultaneously communicate their gradients to the BS via a shared MAC [4]. The MAC's waveform-superposition property allows efficient aggregation of gradients over the air.

2) Threat model: A potential security concern arises from the presence of an eavesdropper whose objective is to intercept the transmitted gradients, potentially inferring sensitive information or disrupting the training process. Following similar arguments in [39]–[41], we assume that eavesdropper is an active user but it is un-trusted by the BS, which indicates that the perfect channel state information (CSI) of eavesdropper's channel is available at the BS [5].

Furthermore, the BS primarily acts as a coordinator, managing the training process and aggregating model updates. However, the BS is considered "honest but curious", meaning it may attempt to extract private information from the received gradients during regular training, which poses a potential privacy risk. Our goal is to ensure that any such privacy and security risks, even in edge cases, can be effectively mitigated by the secure and privacy-enhancing mechanisms implemented in this work.
[6]

3) SP-OTA-FL: To mitigate the risks of privacy breaches and potential wiretap security attacks, SP-OTA-FL strategically selects a subset of devices, denoted as $\mathcal{K}^t$, where $\mathcal{K}^t \subseteq \mathcal{N}$ and $\mathcal{K}^t \neq \varnothing$, to actively participate as learners in the training process. Simultaneously, a portion of the remaining devices are scheduled as jammers, identified by $\mathcal{J}^t \subseteq \mathcal{N}\backslash\mathcal{K}^t$. These jammers generate and transmit Gaussian artificial noise, effectively serving as a jamming signal.

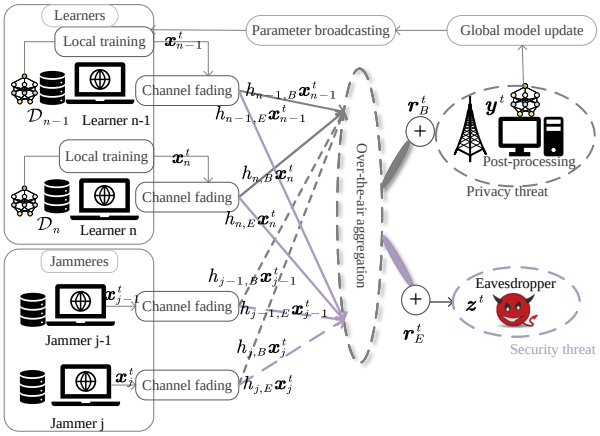


Fig. 1: Jamming-aided SP-OTA-FL.

Assume that P_n is the maximum transmission power of device n and $\boldsymbol{e}_n^t \sim \mathcal{N}\left(0, \frac{1}{d}\mathbf{I}_d\right)$ denotes the artificial noise produced by device n when it is scheduled as a jammer. Then, the signal from device n is given by [7]

$$\boldsymbol{x}_n^t = \begin{cases} \frac{\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\boldsymbol{g}_n, & n \in \mathcal{K}^t, \\ \sqrt{P_n}\boldsymbol{e}_n^t, & n \in \mathcal{J}^t, \end{cases} \tag{4}$$

where $\eta_n \in [0,1]$ is the power scaling factor, ensuring $\mathbb{E}\left[\|\boldsymbol{x}_n^t\|_2^2\right] \leq P_n$. Assume that $h_{n,B}^t \in \mathbb{R}^+$ and $h_{n,E}^t \in \mathbb{R}^+$ are the channel gain coefficients linking device n with the BS and the eavesdropper, respectively. Based on the superposition of MAC, the received signals at the BS and the eavesdropper are respectively given by

$$\boldsymbol{y}^t = \sum_{n\in\mathcal{K}^t} \frac{h_{n,B}^t\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\boldsymbol{g}_n^t + \sum_{n\in\mathcal{J}^t} h_{n,B}^t\sqrt{P_n}\boldsymbol{e}_n^t + \boldsymbol{r}_B^t, \tag{5}$$

$$\boldsymbol{z}^t = \sum_{n\in\mathcal{K}^t} \frac{h_{n,E}^t\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\boldsymbol{g}_n^t + \sum_{n\in\mathcal{J}^t} h_{n,E}^t\sqrt{P_n}\boldsymbol{e}_n^t + \boldsymbol{r}_E^t, \tag{6}$$

where $\boldsymbol{r}_B^t \sim \mathcal{N}\left(0, \frac{\sigma_B^2}{d}\mathbf{I}_d\right)$ and $\boldsymbol{r}_E^t \sim \mathcal{N}\left(0, \frac{\sigma_E^2}{d}\mathbf{I}_d\right)$ represent the received noise at the BS and eavesdropper, respectively. To obtain the average gradient, i.e., $\boldsymbol{g}_{ave}^t = \frac{1}{|\mathcal{K}^t|}\sum_{n\in\mathcal{K}^t}\boldsymbol{g}_n^t$, the BS processes the received signal by

$$\boldsymbol{g}_{esti}^t = \frac{1}{|\mathcal{K}^t|\,\varphi^t}\boldsymbol{y}^t, \tag{7}$$

[3]Stochastic Gradient Descent (SGD) is effective in the SP-OTA-FL framework as well. However, we are using GD to avoid the randomness that comes with SGD. This helps us focus more easily on the main processes of the SP-OTA-FL and renders the mathematical derivations less complex.

[4]We are using a conventional FL algorithm instead of federated averaging to simplify our analysis of the core aspects of the SP-OTA-FL framework. Investigating the impact of multiple local training rounds on security and privacy enhancement will be addressed in future work.

[5]In our future work, we plan to further investigate the scenario with passive mode and imperfect CSI concerning the eavesdropper.

[6]The one participant cases were intended to illustrate potential privacy risks under extreme conditions.

[7]We simplify by not considering power control of jammers; instead, the total power of the jamming signal is managed through device scheduling.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

where φ^t is a post-processing factor and $|\mathcal{K}^t|$ is for averaging. Given that the desired gradient for the global update is $\nabla L\left(\boldsymbol{m}^t\right)=\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{g}_n^t$, the introduced error is given by

$$\Delta \boldsymbol{g}_{err}^t=\underbrace{\boldsymbol{g}_{esti}^t-\boldsymbol{g}_{ave}^t}_{\Delta \boldsymbol{g}_{com}^t}+\underbrace{\boldsymbol{g}_{ave}^t-\nabla L\left(\boldsymbol{m}^t\right)}_{\Delta \boldsymbol{g}_{ds}^t}. \tag{8}$$

This error primarily stems from two sources. The first source is the distortion induced during communication, while the second source is the impact of device scheduling. Therefore, we can minimize the error by optimizing the device scheduling and transmission design, including the design of power scaling factor $\eta_n, n\in\mathcal{K}^t$ and the post-processing factor φ. Finally, the global update is performed using learning rate τ by $\boldsymbol{m}^{t+1}=\boldsymbol{m}^t-\tau\boldsymbol{g}_{esti}^t$.

B. Differential Privacy and MSE-Security

DP [27] provides a formal and mathematically rigorous framework for safeguarding the sensitive information of individuals contained within datasets. It ensures that the contribution of any individual data sample to the model remains statistically indistinguishable by injecting controllable noise and randomness into the disclosed message. The definition of (ϵ,ζ)-DP is given as follows.

Definition 1. *(ϵ,ζ)-DP [27]: A randomized mechanism $\mathcal{O}$ guarantees (ϵ,ζ)-DP if for two adjacent datasets $\mathcal{D},\mathcal{D}'$ differing in one sample, and measurable output space $\mathcal{Q}$ of $\mathcal{O}$, it satisfies* $\Pr\left[\mathcal{O}\left(\mathcal{D}\right)\in\mathcal{Q}\right]\leqslant e^{\epsilon}\Pr\left[\mathcal{O}\left(\mathcal{D}'\right)\in\mathcal{Q}\right]+\zeta$.

The Gaussian DP mechanism which guarantees privacy by adding artificial Gaussian noise is introduced as follows.

Definition 2. *Gaussian mechanism [27]: Gaussian mechanism $\mathcal{O}$ alters the output of another algorithm $\mathcal{L}:\mathcal{D}\rightarrow\mathcal{Q}$ by adding Gaussian noise, i.e., $\mathcal{O}\left(\mathcal{D}\right)=\mathcal{L}\left(\mathcal{D}\right)+\mathcal{N}\left(0,\sigma^2\mathbf{I}_d\right)$. It guarantees (ϵ,ζ)-DP with $\epsilon=\frac{\kappa\Delta S}{\sigma}$ where $\kappa=\sqrt{2\ln\left(\frac{1.25}{\zeta}\right)}$ and $\Delta S\triangleq\max_{\mathcal{D},\mathcal{D}'}\left\|\mathcal{L}\left(\mathcal{D}\right)-\mathcal{L}\left(\mathcal{D}'\right)\right\|_2$ stands for the sensitivity of the algorithm $\mathcal{L}$ signifying the extent to which the algorithm's output varies when a single data sample is altered.*

In [38], MSE-security was proposed to assess the security of analog messages. In this context, MSE-security involves the deliberate introduction of noise, acting as jamming, to degrade the SNR at the eavesdropper. This degradation in SNR serves to thwart the eavesdropper's ability to obtain a low-noise estimate of the obtained message. The details of MSE-security are introduced as follows.

Definition 3. *$(\mathcal{E},\gamma)$-MSE-security [38]: A uniform distributed mechanism $\mathcal{E}:\mathcal{G}\rightarrow\mathcal{Y}$, where $\mathcal{Y}$ is a measureable and bounded output space, guarantees $(\mathcal{E},\gamma)$-MSE-security if under a uniform distribution of $\mathcal{E}\left(\{\boldsymbol{g}_n^t\}_{n\in\mathcal{K}^t}\right)$, for any eavesdropper's estimator $e:\mathcal{Z}\rightarrow\mathcal{Y}$, there is a real number $\gamma\geqslant 0$ satisfying $\mathbb{E}\left[\left(e\left(\boldsymbol{z}^t\right)-\mathcal{E}\left(\{\boldsymbol{g}_n^t\}_{n\in\mathcal{K}^t}\right)\right)^2\right]\geqslant\gamma$.*

In statistical terms, a scheme guarantees $(\mathcal{E},\gamma)$-MSE-security means that all the estimators that the eavesdropper can apply have MSE at least γ.

III. Theoretical analysis

In this section, we present theoretical insights into the impact of device scheduling and power control on the learning process of SP-OTA-FL. We specifically quantify privacy leakage using DP and evaluate security levels with MES-security [8]. Additionally, we conduct a comprehensive convergence analysis of SP-OTA-FL, considering both convex and non-convex machine learning models.

A. Privacy analysis

Even though the gradients are already aggregated before reaching the BS, there remains a potential for privacy leakage in OTA-FL. Privacy can be compromised in specific circumstances where only the gradient from a particular device is changed, while the gradients from other devices remain fixed, as adopted in the privacy analyses of [17], [19]. This creates optimal conditions for malicious attackers to intercept information, representing a worst-case scenario. By evaluating the maximum potential privacy leakage in OTA-FL under these conditions, we ensure that the proposed privacy-enhancing mechanisms are robust even in the most challenging scenarios. We utilize the Gaussian mechanism to quantify the privacy leakage concerning the alteration of a single data point in the dataset of each device in the following.

Let us consider device a as an example. Suppose that we have two adjacent datasets, denoted as $\mathcal{D}_a$ and $\mathcal{D}_a^{'}$, where only one different sample exists between them. Correspondingly, we have two gradients: $\boldsymbol{g}_a^t$ based on $\mathcal{D}_a$ and $\left(\boldsymbol{g}_a^t\right)'$ based on $\mathcal{D}_a^{'}$. The received signals at the BS corresponding to datasets $\mathcal{D}_a$ and $\mathcal{D}_a^{'}$ are respectively given by (5) and

$$\begin{aligned}\left(\boldsymbol{y}^t\right)'=&\sum_{n\in\mathcal{K}^t,n\neq a}\frac{h_{n,B}^t\sqrt{\eta_nP_n}}{\left\|\boldsymbol{g}_n^t\right\|_2}\boldsymbol{g}_n^t+\frac{h_{a,B}^t\sqrt{\eta_aP_a}}{\left\|\left(\boldsymbol{g}_a^t\right)'\right\|_2}\left(\boldsymbol{g}_a^t\right)'\\&+\sum_{n\in\mathcal{J}^t}h_{n,B}^t\sqrt{P_n}\boldsymbol{e}_n^t+\boldsymbol{r}_B^t,\end{aligned} \tag{9}$$

which only differ in the gradient from device a. Following the definition of sensitivity in Definition 2, we have $\Delta S_a^t\triangleq\max_{\mathcal{D}_a,\mathcal{D}_a^{'}}\left\|\boldsymbol{y}^t-\left(\boldsymbol{y}^t\right)'\right\|_2$ and then we have the following results.

Lemma 1. *The sensitivity of SP-OTA-FL to the alterations in a single data within the dataset of device n is given by $\Delta S_n^t\leq 2h_{n,B}^t\sqrt{\eta_nP_n}$, and each learner n achieves (ϵ_n^t,ζ)-DP in round t when the following condition is satisfied,*

$$\epsilon_n^t=\frac{2\kappa\sqrt{d}h_{n,B}^t\sqrt{\eta_nP_n}}{\sqrt{\sum_{n\in\mathcal{J}^t}\left(h_{n,B}^t\right)^2P_n+\sigma_B^2}}. \tag{10}$$

[8]The distribution of local data and the frequency of the aggregation do not impact the analyses of privacy and security in this work. Therefore, privacy and security protection proposed in this work are applicable to the systems considering federated averaging and Non-IID datasets.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

Proof: Please refer to Appendix A. ■

Lemma 1 provides a significant revelation: devices that have better channel quality are more vulnerable to privacy disclosure. This vulnerability arises from the fact that concealing information within noise becomes more challenging when dealing with gradients of larger amplitude. To counteract the risk of privacy leakage within the system, two effective strategies emerge. One option is to enhance the strength of noise by optimizing the jammer set $\mathcal{J}$. Alternatively, one can reduce the power of gradients by adjusting the power scaling factor η_n.

B. Security analysis

In this subsection, we delve into the security analysis of SP-OTA-FL. Let us consider a scenario where the eavesdropper's goal is to reconstruct an averaged estimation of the gradients, denoted as $\boldsymbol{g}_{ave}^t$. This estimation holds the potential for further security attacks.

Lemma 2. *Assume that the elements of $\boldsymbol{g}_n^t$ are distributed uniformly in [a, b]. The aggregation mechanism $\mathcal{E}^t : (\boldsymbol{g}_n^t)_{n\in\mathcal{K}^t} \to \boldsymbol{z}^t \in \mathcal{Z}$ guarteens $\left(\mathcal{E}^t, (\varpi^t)^2 \Xi\left(\frac{b-a}{\varpi^t}\right)\right)$-MSE-security in training round t. Specifically, $\Xi(t) = \int_0^t \int_{-\infty}^{+\infty} \left(v + \frac{\alpha(-v)-\alpha(t-v)}{\beta(t-v)-\beta(-v)} - u\right)^2 \cdot \frac{1}{t}\alpha(u-v)\,dvdu$ where $\alpha(\cdot)$ and $\beta(\cdot)$ denote the probability density function and the cumulative distribution function of the standard normal distribution, and*

$$\varpi^t = \frac{1}{\sqrt{d}\,|\mathcal{K}^t|\,\gamma^t}\sqrt{\sum_{n\in\mathcal{J}^t}\left(h_{n,E}^t\right)^2 P_n + \sigma_E^2}, \tag{11}$$

where $\gamma^t = \max\limits_{n\in\mathcal{K}^t}\left\{\frac{h_{n,B}^t\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\right\}$.

Proof: Please refer to Appendix B. ■

According to Definition 3, $\left(\mathcal{E}^t, (\varpi^t)^2 \Xi\left(\frac{b-a}{\varpi^t}\right)\right)$-MSE-security ensures that the gradient estimates derived from $\boldsymbol{z}^t$ have a MSE of at least $(\varpi^t)^2 \Xi\left(\frac{b-a}{\varpi^t}\right)$. It has been validated that $(\varpi^t)^2 \Xi\left(\frac{b-a}{\varpi^t}\right)$ increases with ϖ^t in [38]. Therefore, a higher value of ϖ^t is associated with an increased level of system security, thus designating ϖ^t as the security coefficient, an indicator of the system's security level. (11) proves that one way to secure the FL process is to increase the aggregated noise at the evasdropper or design a smaller γ^t by device scheduling and power control.

C. Convergence analysis

We here present convergence analysis in the cases of convex and non-convex loss functions. Our subsequent convergence analyses of SP-OTA-FL rely on several assumptions on the loss function and gradients as follows, which are widely adopted for distributed optimization [19], [42]–[44].

Assumption 1. *The expected squared norm of the gradients is bounded by $\mathbb{E}[\|\boldsymbol{g}_n^t\|_2] \leqslant G$, which can be gauranteened by gradient clipping [45], [46].*

Assumption 2. *$L_n(\cdot)$ is θ-smooth, i.e., $L_n(\boldsymbol{\iota}') - L_n(\boldsymbol{\iota}) \leqslant (\boldsymbol{\iota}'-\boldsymbol{\iota})^{\mathrm{T}}\nabla L_n(\boldsymbol{\iota}) + \frac{\theta}{2}\|\boldsymbol{\iota}'-\boldsymbol{\iota}\|_2^2$.*

According to the properties of linear functions and (1), we can conclude that $L(\cdot)$ is also θ-smooth. Then, we have the following result.

Lemma 3. *An upper bound of the expected loss function after training round t is given by*

$$\mathbb{E}\left[L\left(\boldsymbol{m}^{t+1}\right)\right] \leq \mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right] - \frac{\tau}{2}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right] + \tau\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t}, \varphi^t, \mathcal{K}^t, \mathcal{J}^t\right), \tag{12}$$

where $\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t}, \varphi^t, \mathcal{K}^t, \mathcal{J}^t\right) = \mathbb{E}\left[\|\Delta\boldsymbol{g}_{com}^t\|_2^2\right] + \mathbb{E}\left[\|\Delta\boldsymbol{g}_{ds}^t\|_2^2\right]$ denotes the impact of the transmission design and device scheduling. Specifically, the communication MSE $\mathbb{E}\left[\|\Delta\boldsymbol{g}_{com}^t\|_2^2\right]$ is bounded by

$$\mathbb{E}\left[\left\|\Delta\boldsymbol{g}_{com}^t\right\|_2^2\right] \leq \frac{1}{|\mathcal{K}^t|}\sum_{n\in\mathcal{K}^t}\left(\frac{h_{n,B}^t\sqrt{\eta_n P_n}}{\varphi^t} - \left\|\boldsymbol{g}_n^t\right\|_2\right)^2 + \frac{1}{\left(|\mathcal{K}^t|\,\varphi^t\right)^2}\left(\sum_{n\in\mathcal{J}^t}\left(h_{n,B}^t\right)^2 P_n + \sigma_B^2\right), \tag{13}$$

and the device scheduling MSE is bounded by

$$\mathbb{E}\left[\left\|\Delta\boldsymbol{g}_{ds}^t\right\|_2^2\right] \leq 4\left(1 - \frac{|\mathcal{K}^t|}{N}\right)^2 G^2. \tag{14}$$

Proof: Please refer to Appendix C. ■

With Lemma 3, we establish an upper bound on the loss function $L(\cdot)$ concerning the communication MSE and the device selection MSE. The communication MSE in (13) consists of two terms. The first term characterizes the error stemming from the over-the-air aggregation, while the second term characterizes the error introduced by the jamming signal and channel noise. Letting $\Lambda = \max_{t\in T}\left\{\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t}, \varphi^t, \mathcal{K}^t, \mathcal{J}^t\right)\right\}$ and $\Gamma = \mathbb{E}\left[L\left(\boldsymbol{m}^0\right) - L\left(\boldsymbol{m}^*\right)\right]$, we present the convergence analysis results in the following.

1) Convex setting: We begin by considering the most benign setting, wherein the loss function $L(\cdot)$ is assumed to be strongly convex. We formalize a strong convexity assumption as below.

Assumption 3. *$L_n(\cdot)$ is ρ-strongly convex, i.e., $L_n(\boldsymbol{\iota}') - L_n(\boldsymbol{\iota}) \geqslant (\boldsymbol{\iota}'-\boldsymbol{\iota})^{\mathrm{T}}\nabla L_n(\boldsymbol{\iota}) + \frac{\rho}{2}\|\boldsymbol{\iota}'-\boldsymbol{\iota}\|_2^2$.*

The convergence theorem of the SP-OTA-FL with a strongly convex objective function is given as follows.

Theorem 1. *Assume that $\boldsymbol{m}^*$ is the optimal model and $\tau \leq \frac{1}{\theta}$. The optimality gap $\mathbb{E}\left[L\left(\boldsymbol{m}^T\right) - L\left(\boldsymbol{m}^*\right)\right]$ is bounded by*

$$\mathbb{E}\left[L\left(\boldsymbol{m}^T\right) - L\left(\boldsymbol{m}^*\right)\right] \leq (1-\rho\tau)^T\Gamma + \frac{1-(1-\rho\tau)^T}{\rho}\Lambda. \tag{15}$$

Proof: Please refer to Appendix D. ■

The optimality gap represented on the right-hand side of (15) can be decomposed into two distinct components.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

The first term pertains to the initial gap Γ, which gradually diminishes towards 0 as the parameter T tends towards infinity since $1-\rho\tau \leq 1$. The second term illustrates the impact of aggregation error on the learning process. As T approaches infinity, the second term converges to $\frac{\Lambda}{\rho}$. This means that to enhance learning accuracy, it is necessary to minimize Λ.

2) Non-convex setting: Given that various useful machine learning models are characterized by non-convex objective functions, we next investigate the convergence properties of SP-OTA-FL in the non-convex setting. Different from the convex case where the convergence rate is typically quantified using the expected optimality gap, in the case of non-convex loss function $L(\cdot)$, the algorithm converging to a global minimum cannot generally be guaranteed. A reasonable substitute is to adopt the average expected squared gradient norm as the convergence indicator, which is widely used in convergence analysis for non-convex loss functions [47]–[50].

Theorem 2. *The average expected squared gradient norm after T training rounds is bounded as follows:*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right] \leq \frac{2\Gamma}{\tau T}+2\Lambda. \tag{16}$$

Proof: Please refer to Appendix E. ■

Similar to the result in the convex case, the upper bound is determined by the initial optimality gap Γ and Λ. Hence, in both convex and non-convex settings, the minimization of Λ is crucial for improving convergence performance. This minimization is equivalent to minimizing $\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t},\varphi^t,\mathcal{K}^t,\mathcal{J}^t\right)$ in each training round t.

To demonstrate the applicability of the proposed SP-OTA-FL utilizing federated averaging (FedAvg) across non-independent and identically distributed (Non-IID) datasets, we also provide the convergence behavior of the convex loss function of FedAvg on Non-IID datasets below. We define the optimal solution as

$$\boldsymbol{m}^* = \arg\min_{\boldsymbol{m}} L(\boldsymbol{m}), \tag{17}$$

and the optimal solution for device n as

$$\boldsymbol{m}_n^* = \arg\min_{\boldsymbol{m}} L_n(\boldsymbol{m}). \tag{18}$$

We use the term

$$\Omega = L(\boldsymbol{m}^*) - \frac{1}{N}\sum_{n=1}^{N} L_n(\boldsymbol{m}_n^*), \tag{19}$$

for quantifying the degree of Non-IID as in [51], [52]. If the data are IID, then the value of Ω approaches zero as the number of samples increases. Conversely, if the data are Non-IID, Ω is nonzero, and its magnitude reflects the heterogeneity of the data distribution.

Then, we have the following results.

Lemma 4. *Given the number of the local training rounds in FedAvg is E, an upper bound of the gap between the model communication round t, i.e., $\boldsymbol{m}^{t+1}$, and the optimal model $\boldsymbol{m}^*$ can be given by*

$$\mathbb{E}\left[\left\|\boldsymbol{m}^{t+1}-\boldsymbol{m}^*\right\|_2^2\right] \leq \alpha\mathbb{E}\left[\left\|\boldsymbol{m}^{t}-\boldsymbol{m}^*\right\|_2^2\right]+\beta, \tag{20}$$

where

$$\alpha = 2\left(1-\rho\tau\left(E-\tau\left(E-1\right)\right)\right), \tag{21}$$

and

$$\begin{aligned}\beta =& 4\tau\left(E-1\right)\Omega + 2\tau^2\left(E-1\right)G^2 + 4\tau^2\Lambda \\ &+\tau^2\left(1+\rho\left(1-\tau\right)\right)\frac{E\left(E-1\right)\left(2E-1\right)}{3}G^2.\end{aligned} \tag{22}$$

Proof: The proof follows a similar process as in [52] and is omitted here. ■

Following the θ-smothness in Assumption 2, we can derive the optimality gap as follows:

$$\begin{aligned}&\mathbb{E}\left[L\left(\boldsymbol{m}^T\right)-L\left(\boldsymbol{m}^*\right)\right] \leq \frac{\theta}{2}\mathbb{E}\left[\left\|\boldsymbol{m}^T-\boldsymbol{m}^*\right\|_2^2\right] \\ \leq&\frac{\theta}{2}\left(\alpha^T\left\|\boldsymbol{m}^0-\boldsymbol{m}^*\right\|_2^2+\frac{1-\alpha^T}{1-\alpha}\Lambda\right).\end{aligned} \tag{23}$$

From Lemma 4 and equation (23), it can be observed that even though the local training round E of FedAvg and the degree of Non-IID Ω influence the optimality gap, they are independent of the term Λ, which relates to device scheduling and transmission design. Under the conditions that (1) $\rho \leq \frac{1}{2}$ while $E \geq \max\left\{1, \frac{2+\sqrt{\Delta_2}}{2\rho}\right\}$ or $1 \leq E \leq \frac{2-\sqrt{\Delta_2}}{2\rho}$ or (2) $\rho \geq \frac{1}{2}$ where $\Delta_1 = \rho^2E^2 - 2\rho\left(E-1\right)$ and $\Delta_2 = 4-8\rho$, $\alpha \leq 1$ holds, therefore, the first term will approach zero as T goes infinity. Consistent with the basic scenario involving one local training round algorithm and IID datasets, the learning performance can be improved by minimizing $\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t},\varphi^t,\mathcal{K}^t,\mathcal{J}^t\right)$. Therefore, the following optimal designs are also applicable to the FedAvg algorithm and Non-IID scenarios.

Motivated by the above analytical insights, in the following sections, we formulate optimization problems to strike a balance between the enhanced privacy and security and the compromised accuracy by minimizing $\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}^t},\varphi^t,\mathcal{K}^t,\mathcal{J}^t\right)$. We will explore two primary OTA-FL scenarios: SP-OTA-FL with unbiased aggregation (UB-SP-OTA-FL) and SP-OTA-FL with biased aggregation (B-SP-OTA-FL). The privacy budget of device n and the security requirement for implementing SP-OTA-FL are respectively denoted by (ϵ_n,ζ) and ϖ, and index t will be omitted for brevity. Detailed explanations of the optimal designs for UB-SP-OTA-FL and B-SP-OTA-FL are presented in Section IV and Section V, respectively.

IV. Optimization for UB-SP-OTA-FL

In this section, we formulate an optimization problem with the goal of enhancing the learning performance of SP-OTA-FL through the optimization of the learner $\mathcal{K}$ and jammer $\mathcal{J}$, as well as the design of two critical factors: the power scaling factor $\eta_n, n\in\mathcal{K}$ and the post-processing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

factor φ, where $0 \leq \eta_n \leq 1$, as outlined in II-B. For ease of presentation, we define $\sigma_{B,\mathcal{J}}^{tot} = \sum_{n\in\mathcal{J}} h_{n,B}^2 P_n + \sigma_B^2$ and $\sigma_{E,\mathcal{J}}^{tot} = \sum_{n\in\mathcal{J}} h_{n,E}^2 P_n + \sigma_E^2$. Then, the optimization problem can be expressed as follows:

$$\mathbf{P1}. \min_{\{\eta_n\}_{n\in\mathcal{K}},\varphi,\mathcal{K},\mathcal{J}} \left\{\Lambda\left(\{\eta_n\}_{n\in\mathcal{K}}, \varphi, \mathcal{K}, \mathcal{J}\right)\right\} \tag{24}$$

$$\textbf{s.t.}\ 0 \leq \eta_n \leq 1, n \in \mathcal{K}, \tag{24a}$$

$$\mathcal{K} \neq \varnothing, \mathcal{K} \subseteq \mathcal{N}, \mathcal{J} \subseteq \mathcal{N}/\mathcal{K}, \tag{24b}$$

$$\frac{2\kappa\sqrt{d}h_{n,B}\sqrt{\eta_n P_n}}{\sqrt{\sigma_{B,\mathcal{J}}^{tot}}} \leqslant \epsilon_n, n \in \mathcal{K}, \tag{24c}$$

$$\frac{1}{\sqrt{d}\left|\mathcal{K}\right|\gamma}\sqrt{\sigma_{B,\mathcal{J}}^{tot}} \geqslant \varpi. \tag{24d}$$

The constraint (24a) establishes a constraint on the peak transmit power. The constraint (24b) implies that a device cannot serve as a learner and jammer concurrently and at least one device should be selected as a learner to participate in the training. The privacy of each learner is guaranteed in the constraint (24c) while (24d) denotes the security constraint of the system. To tackle the optimization problem presented in (24), let us simplify it first. By observing the objective function, the optimal $\eta_n^*, n \in \mathcal{K}$ can be determined by the following lemma.

Lemma 5. *Given any $\mathcal{K}$ and φ, the optimal solution of $\eta_n^*, n \in \mathcal{K}$ to problem $\mathbf{P1}$ is given by*

$$\eta_n^* = \frac{\varphi^2 \left\|\boldsymbol{g}_n\right\|_2^2}{h_{n,B}^2 P_n}, n \in \mathcal{K}. \tag{25}$$

Proof: This result follows by setting the first term of (24) as zero, i.e., $\frac{h_{n,B}\sqrt{\eta_n P_n}}{\varphi} - \left\|\boldsymbol{g}_n\right\|_2 = 0$. ■

With the aid of Lemma 5, the received signal at the BS in (5) can be simplified as

$$\boldsymbol{y} = \varphi \sum_{n\in\mathcal{K}} \boldsymbol{g}_n + \sum_{n\in\mathcal{J}} h_{n,B}\sqrt{P_n}\boldsymbol{e}_n + \boldsymbol{r}_B, \tag{26}$$

which is referred to as UB-OTA-FL and adopted in [17], [19]. In this UB-OTA-FL scenario, we refer to φ as the alignment coefficient. Consequently, the privacy leakage in (10) and the MSE-security in (11) can be re-expressed as

$$\epsilon_n = \frac{2\kappa\sqrt{d}\varphi\left\|\boldsymbol{g}_n\right\|_2}{\sqrt{\sigma_{B,\mathcal{J}}^{tot}}} \text{ and } \varpi = \frac{1}{\sqrt{d}\left|\mathcal{K}\right|\varphi}\sqrt{\sigma_{E,\mathcal{J}}^{tot}}. \tag{27}$$

By converting the constraint (24a) as $0 \leq \varphi \leq \frac{h_{n,B}\sqrt{P_n}}{\left\|\boldsymbol{g}_n\right\|_2}, n \in \mathcal{K}$ and defining $\Omega\left(\varphi, \mathcal{K}, \mathcal{J}\right) = \frac{\sigma_{B,\mathcal{J}}^{tot}}{(|\mathcal{K}|\varphi)^2} + 4\left(1 - \frac{|\mathcal{K}|}{N}\right)^2 G^2$, the optimization problem in UB-SP-OTA-FL is established as follows:

$$\mathbf{P2}. \min_{\varphi,\mathcal{K},\mathcal{J}} \left\{\Omega\left(\varphi, \mathcal{K}, \mathcal{J}\right)\right\} \tag{28}$$

$$\textbf{s.t.}\ \mathcal{K} \neq \varnothing, \mathcal{K} \subseteq \mathcal{N}, \mathcal{J} \subseteq \mathcal{N}/\mathcal{K}, \tag{28a}$$

$$0 \leq \varphi \leq \frac{h_{n,B}\sqrt{P_n}}{\left\|\boldsymbol{g}_n\right\|_2}, n \in \mathcal{K}, \tag{28b}$$

$$4d\kappa^2\varphi^2 \leq \frac{\epsilon_n^2}{\left\|\boldsymbol{g}_n\right\|_2^2}\sigma_{B,\mathcal{J}}^{tot}, n \in \mathcal{K}, \tag{28c}$$

$$d\varpi^2\left|\mathcal{K}\right|^2\varphi^2 \leq \sigma_{E,\mathcal{J}}^{tot}. \tag{28d}$$

In the following, we first present an algorithm to decouple the optimization variables and address the challenging non-convex optimization problem $\mathbf{P2}$. Our approach commences by offering a closed-form solution for the optimal φ^* and $\mathcal{K}^*$ under a given set $\mathcal{J}$ in Section IV-A. Subsequently, we introduce an algorithm called SU for obtaining optimal $\mathcal{J}$ in Section IV-B.

A. Closed-form Solution for Optimal φ^* and $\mathcal{K}^*$ Given $\mathcal{J}$

Given $\mathcal{J}$, $\mathbf{P2}$ can be simplified as

$$\mathbf{P3}. \min_{\varphi,\mathcal{K}} \left\{\Omega\left(\varphi, \mathcal{K}, \mathcal{J}\right)\right\} \tag{29}$$

$$\textbf{s.t.}\ \mathcal{K} \neq \varnothing, \mathcal{K} \subseteq \mathcal{R}_{\mathcal{J}}, \tag{29a}$$

$$(28\text{b}), (28\text{c}), (28\text{d}), \tag{29b}$$

where $\mathcal{R}_{\mathcal{J}} = \mathcal{N}/\mathcal{J}$. Then, the optimal value of φ can be transformed into a function of $\mathcal{K}$ according to the following lemma.

Lemma 6. *For any given $\mathcal{K}$ and $\mathcal{J}$, the optimal solution of φ to problem $\mathbf{P3}$ is given by*

$$\varphi^* = f\left(\mathcal{K}\right) \triangleq \min\left\{\varphi_{cc}, \varphi_{pr}, \varphi_{se}\right\}, \tag{30}$$

where $\varphi_{cc} = \min_{n\in\mathcal{K}}\left\{\frac{h_{n,B}\sqrt{P_n}}{\left\|\boldsymbol{g}_n\right\|_2}\right\}, \varphi_{se} = \frac{1}{\sqrt{d}\varpi|\mathcal{K}|}\sqrt{\sigma_{E,\mathcal{J}}^{tot}}$, *and* $\varphi_{pr} = \frac{1}{2\kappa}\sqrt{\frac{\sigma_{B,\mathcal{J}}^{tot}}{d}}\min_{n\in\mathcal{K}}\left\{\frac{\epsilon_n}{\left\|\boldsymbol{g}_n\right\|_2}\right\}$.

Proof: Following $(28\text{b}), (28\text{c}), (28\text{d})$, we have $0 \leq \varphi \leq \varphi_{cc}$, $0 \leq \varphi \leq \varphi_{pr}$, and $0 \leq \varphi \leq \varphi_{se}$, respectively. To achieve the minimal value of the objective function with given $\mathcal{K}$ and $\mathcal{J}$, the largest φ within the feasible region is required, leading to $\varphi^* = \min\left\{\varphi_{cc}, \varphi_{pr}, \varphi_{se}\right\}$. ■

Remark 1. *It can be seen from Lemma 6 that there is a tradeoff between the alignment coefficient φ and the number of scheduled devices $|\mathcal{K}|$. Specifically, a larger $|\mathcal{K}|$ may lead to a smaller φ. For example, if $|\mathcal{K}| = N$, that means all the devices are involved in the training and $\varphi_{cc}, \varphi_{pr}$, and φ_{se}, achieve the minimal values, respectively, which will lead to the minimal φ. On the contrary, φ can achieve its largest value by scheduling the device with the largest $\frac{h_{n,B}\sqrt{P_n}}{\left\|\boldsymbol{g}_n\right\|_2}$ or the largest $\frac{\epsilon_n}{\left\|\boldsymbol{g}_n\right\|_2}$ when $|\mathcal{K}| = 1$.*

Building upon Lemma 6, $\mathbf{P3}$ can be simplified as

$$\mathbf{P4}. \min_{\mathcal{K}} \left\{\frac{\sigma_{B,\mathcal{J}}^{tot}}{\left(|\mathcal{K}| f\left(\mathcal{K}\right)\right)^2} + 4\left(1 - \frac{|\mathcal{K}|}{N}\right)^2 G^2\right\} \tag{31}$$

$$\textbf{s.t.}\ (29a). \tag{31a}$$

By observing (31), we can see that the largest $f\left(\mathcal{K}\right)$ is required to minimize the objective function for a fixed $|\mathcal{K}|$. Taking $|\mathcal{K}| = i, 1 \leq i \leq |\mathcal{R}_{\mathcal{J}}|, i \in \mathcal{Z}$, where $\mathcal{Z}$ is the set of integers, as an example, the largest $f\left(\mathcal{K}\right)$ satisfying $|\mathcal{K}| = i$ can be achieved by finding the i devices that yield the largest

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

$\min\{\varphi_{cc}, \varphi_{pr}, \varphi_{se}\}$ according to (30). It is equivalent to achieve the largest $\min\{\varphi_{cc}, \varphi_{pr}\}$ since φ_{se} is fixed when $|\mathcal{K}| = i$. By sorting the devices in descending order of $\varphi^n_{ccpr} = \min_{n\in\mathcal{K}} \left\{ \frac{h_{n,B}\sqrt{P_n}}{\|\boldsymbol{g}_n\|_2}, \frac{1}{2\kappa}\sqrt{\frac{\sigma^{tot}_{B,\mathcal{J}}}{d}}\frac{\epsilon_n}{\|\boldsymbol{g}_n\|_2} \right\}$, the $\mathcal{K}$ that achieves the largest $f(\mathcal{K})$ is given by the following lemma. For ease of presentation, we introduce a mechanism denoted as $S(\mathcal{N}; q)$, which organizes the elements in the set $\mathcal{N}$ in descending order according to the parameter q.

Lemma 7. *For any $|\mathcal{K}| = i, 1 \le i \le |\mathcal{R}_{\mathcal{J}}|, i \in \mathcal{Z}$, the largest $f(\mathcal{K})$, denoted as $f(\mathcal{K}^*_i)$, is achieved when*

$$\mathcal{K}^*_i = \{\boldsymbol{s}_{ccpr}[n] \,|n < i\}, \tag{32}$$

where $\boldsymbol{s}_{ccpr} = S\left(\mathcal{R}_{\mathcal{J}}; \varphi^n_{ccpr}\right)$.

Proof: As the elements in $\boldsymbol{s}_{ccpr}$ are sorted in descending order of φ^n_{ccpr}, the first i elements will contribute to the largest $\min\{\varphi_{cc}, \varphi_{pr}, \varphi_{se}\}$, i.e., the largest $f(\mathcal{K})$, in the case that $|\mathcal{K}| = i$. ■

Based on Lemma 7, the optimal solution $\mathcal{K}^*$ to problem $\mathbf{P3}$ and $\mathbf{P4}$ is given by $\mathcal{K}_{i^*}$ where

$$i^* = \arg\min_{1\le i\le|\mathcal{R}_{\mathcal{J}}|} \{\Omega(f(\mathcal{K}^*_i), \mathcal{K}^*_i, \mathcal{J})\}. \tag{33}$$

Consequently, the optimal φ^* for problem $\mathbf{P3}$ can be obtained using Lemma 6.

B. SU Algorithm for Optimal $\mathcal{J}$

The optimal $\mathcal{J}$ can be derived by exhaustively considering all possible sets of jammers. For each set of jammers, the optimal solutions for φ and $\mathcal{K}$ outlined in Section IV-A can be implemented. However, this exhaustive enumeration introduces a computational complexity on the order of $\binom{N}{c}$ when c devices are designated as jammers. Moreover, this complexity escalates with the increase in the parameter N. Thus, obtaining the optimal solution for $\mathcal{J}$ becomes considerably challenging with a large value of N. To mitigate the high enumeration complexity, we develop an SU algorithm inspired by [53] to obtain $\mathcal{J}$. In the SU algorithm, we systematically consider different jammer quantities. For each specific jammer count, we generate an initial jammer set and proceed to iteratively optimize an individual jammer while keeping the others fixed. For ease of presentation, we introduce the notation $\mathcal{J}^{ls}$ denoting the corresponding list of the set $\mathcal{J}$. Then, the overall procedure based on the SU algorithm for solving $\mathbf{P}2$ is summarized in Algorithm 1.

Given that we derive φ_c and $\mathcal{K}_c$ using Lemma 6 and Lemma 7 with a computational complexity of $\mathcal{O}(N)$, the overall complexity of the SU algorithm is $\mathcal{O}(N^4)$. While the SU algorithm reduces the enumeration complexity, it remains substantial, especially with larger values of N.

C. SP-SJR-based LC Solution for UB-SP-OTA-FL

By examining constraints (28c) and (28d) in $\mathbf{P2}$, we see that there are three ways to guarantee the privacy and security: (1) decrease φ; (2) reduce $\mathcal{K}$; or (3) increase the power of jamming. Inspired by this, we propose an LC solution for addressing $\mathbf{P2}$.

Algorithm 1 SU-based Solution for Solving $\mathbf{P2}$

Input: Given $\{(\epsilon_n, \zeta)\}_{n=1}^N$ and ϖ. Initialize $\Omega^{min} = +\infty$.
Output: $\mathcal{K}^*$, $\mathcal{J}^*$ and φ^*.

```
for c ∈ [0, N − 1] do
    Initialize J_c and let R_c = N/J_c.
    for id ∈ [0, c − 1] do
        for j ∈ R_c ∪ J_c^ls [id] do
            Let J_c^ls [id] = j and compute φ_c and K_c using Lemma 6 and Lemma 7.
            if Ω(φ_c, K_c, J_c) < Ω^min then
                Let φ* = φ_c, K* = K_c, J* = J_c, and Ω^min = Ω(φ_c, K_c, J_c).
            end if
        end for
        Let J_c = J*.
    end for
end for
```

We begin by enumerating the potential number of scheduled learners and obtain the corresponding optimal $\mathcal{K}$ and φ without considering privacy and security constraints. Then, we design the optimal set of jammers that guarantees privacy and security under the obtained $\mathcal{K}$ and φ. Specifically, for given $\mathcal{K}$ and φ, we have the following results:

Proposition 1. *For given $\mathcal{K}$ and φ, no jammers are required when the privacy of each participant and the security of the system satisfy the following conditions:*

$$\max_{n\in\mathcal{K}} \left\{ \frac{\|\mathbf{g}_n\|_2}{\epsilon_n} \right\} \le \frac{\sigma_B}{2\sqrt{d}\kappa\varphi} \text{ and } \varpi \le \frac{\sigma_E}{\sqrt{d}\,|\mathcal{K}|\,\varphi}. \tag{34}$$

Conversely, no set of jammers is qualified to guarantee privacy and security when

$$\min_{n\in\mathcal{K}} \left\{ \frac{\|\mathbf{g}_n\|_2}{\epsilon_n} \right\} > \frac{1}{2\kappa\varphi}\sqrt{\frac{\sigma^{tot}_{B,\mathcal{R}_{\mathcal{K}}}}{d}} \text{ and } \varpi > \frac{1}{|\mathcal{K}|\,\varphi}\sqrt{\frac{\sigma^{tot}_{E,\mathcal{R}_{\mathcal{K}}}}{d}}, \tag{35}$$

where $\mathcal{R}_{\mathcal{K}} = \mathcal{N}/\mathcal{K}$.

Proof: Based on $4d\kappa^2\varphi^2\frac{\|\boldsymbol{g}_n\|_2^2}{\epsilon_n^2} \le \sigma_B^2, n \in \mathcal{K}$ and $d\varpi^2|\mathcal{K}|^2\varphi^2 \le \sigma_E^2$, we establish the validity of (34). Similarly, when we set $4d\kappa^2\varphi^2\frac{\|\boldsymbol{g}_n\|_2^2}{\epsilon_n^2} \ge \sigma^{tot}_{B,\mathcal{R}_{\mathcal{K}}}, n \in \mathcal{K}$ and $d\varpi^2|\mathcal{K}|^2\varphi^2 \ge \sigma^{tot}_{E,\mathcal{R}_{\mathcal{K}}}$, we conclude the proof of (35). ■

In the scenario described by (35), where privacy and security cannot be guaranteed by the remaining devices, we schedule all the remaining devices as jammers, and then update φ using Lemma 7. From Proposition 1, we can also see that in the case that $\frac{\sigma_B}{2\sqrt{d}\kappa\varphi} < \left\{\frac{\|\mathbf{g}_n\|_2}{\epsilon_n}\right\}_{n\in\mathcal{K}} < \frac{1}{2\kappa\varphi}\sqrt{\frac{\sigma^{tot}_{B,\mathcal{R}_{\mathcal{K}}}}{d}}$ and $\frac{\sigma_E}{\sqrt{d}|\mathcal{K}|\varphi} < \varpi < \frac{1}{|\mathcal{K}|\varphi}\sqrt{\frac{\sigma^{tot}_{E,\mathcal{R}_{\mathcal{K}}}}{d}}$, some of the remaining devices are required to be scheduled as jammers. To find the optimal jammer set that satisfies privacy and security constraints, we propose an SP-SJR algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

Algorithm 2 The procedure of the SP-SJR algorithm

Input: Given $\mathcal{K}$, φ.
Output: $\mathcal{J}$.
1: Let $\mathcal{J}=\mathcal{R}_{\mathcal{K}}$ and compute σ^{pr} and σ^{se}.
2: **repeat**
3: $j=\begin{cases} n \text{ with the largest } h_{n,B},\ \sigma^{pr}>\sigma^{se}, \\ n \text{ with the largest } h_{n,E},\ \sigma^{pr}<=\sigma^{se}. \end{cases}$
4: Let $\mathcal{J}=\mathcal{J}/\{j\}$, $\sigma^{pr}=\sigma^{pr}-h_{j,B}^2P_j$ and $\sigma^{se}=\sigma^{se}-h_{j,E}^2P_j$.
5: **until** $\sigma^{pr}<=0$ and $\sigma^{se}<=0$ or $\mathcal{J}=\varnothing$.

Algorithm 3 SP-SJR-based LC scheme for solving $\mathbf{P2}$.

Input: Given $\{(\epsilon_n,\zeta)\}_{n=1}^N$ and ϖ.
Output: $\mathcal{K}^*$, $\mathcal{J}^*$ and φ^*.
1: **for** $i\in[1,\cdots,N]$ **do**
2: Let $\mathcal{K}_i^*=\{\boldsymbol{s}_{\mathcal{N}}[n]\,|n<i\}$, $\mathcal{R}_i=\mathcal{N}/\mathcal{K}_i^*$ and $\varphi_i^*=\min_{n\in\mathcal{K}_i^*}\left\{\frac{h_{n,B}\sqrt{P_n}}{\|\boldsymbol{g}_n\|_2}\right\}$.
3: **if** $\min_{n\in\mathcal{K}}\left\{\frac{\|\mathbf{g}_n\|_2}{\epsilon_n}\right\} > \frac{1}{2\kappa\varphi}\sqrt{\frac{\sigma_{B,\mathcal{R}_{\mathcal{K}}}^{tot}}{d}}$ and $\varpi > \frac{1}{|\mathcal{K}|\varphi}\sqrt{\frac{\sigma_{E,\mathcal{R}_{\mathcal{K}}}^{tot}}{d}}$, **then**
4: Let $\mathcal{J}_i^*=\mathcal{R}_i$ and update φ_i^* using Lemma 7.
5: **else**
6: Obtain $\mathcal{J}_i^*$ by applying Algorithm 2.
7: **end if**
8: **end for**
9: $\mathcal{K}^*=\mathcal{K}_i^*$, $\mathcal{J}^*=\mathcal{J}_i^*$ and $\varphi^*=\varphi_i^*$ where $i=\arg\min_{1\le i\le N}\{\Omega(\mathcal{K}_i^*,\mathcal{J}_i^*,\varphi_i^*)\}$.

With SP-SJR, we first set all the remaining devices as jammers, i.e., $\mathcal{J}=\mathcal{R}_{\mathcal{K}}$ and calculate the powers of the jamming signals needed for enhancing privacy and security, respectively, denoted as σ^{pr} and σ^{se}. Then, we remove one jammer at a time accroding to privacy and securiy requirements until the privacy and security constraints are no longer met. Following this, we obtain an optimal jammer set satisfying privacy and security. The procedure of the SP-SJR algorithm to find the optimal jammer set is summarized in Algorithm 2, where $\sigma^{pr}=4d\kappa^2\varphi^2\max_{n\in\mathcal{K}}\left\{\frac{\|\boldsymbol{g}_n\|_2^2}{\epsilon_n^2}\right\}-\sigma_B^2$ and $\sigma^{se}=d\varpi^2|\mathcal{K}|^2\varphi^2-\sigma_E^2$.

The optimal solution to Problem $\mathbf{P2}$ can be obtained by searching these enumerated potential solutions. The overall procedure of the LC solution for solving $\mathbf{P2}$ is presented in Algorithm 3 where $\boldsymbol{s}_{\mathcal{N}}=S\left(\mathcal{N};\frac{h_{n,B}\sqrt{P_n}}{\|\boldsymbol{g}_n\|_2}\right)$. Given that the SP-SJR algorithm has a computational complexity of $\mathcal{O}(N)$, the proposed LC solution operates at a complexity of $\mathcal{O}\left(N^2\right)$, which can efficiently address Problem $\mathbf{P2}$.

V. OPTIMIZATION FOR B-SP-OTA-FL

In this section, we investigate the optimization problem in B-OTA-FL systems [30], [31] where the gradients do not need to be aligned. Then, the overall SNR will not be limited by the alignment coefficient. We consider that all the gradients are transmitted with maximum transmission power, i.e., $\eta_n=1$. Then, the minimization problem is given as

$$\mathbf{P6}.\ \min_{\varphi,\mathcal{K},\mathcal{J}}\left\{\Lambda\left(\{1\}_{n\in\mathcal{K}},\varphi,\mathcal{K},\mathcal{J}\right)\right\} \tag{36}$$
$$\text{s.t. } \mathcal{K}\neq\varnothing,\mathcal{K}\subseteq\mathcal{N},\mathcal{J}\subseteq\mathcal{N}/\mathcal{K}, \tag{36a}$$
$$\frac{2\kappa\sqrt{d}h_{n,B}\sqrt{P_n}}{\epsilon_n}\le\sqrt{\sigma_{B,\mathcal{J}}^{tot}},n\in\mathcal{K}, \tag{36b}$$
$$d\varpi^2|\mathcal{K}|^2\gamma^2\le\sigma_{E,\mathcal{J}}^{tot}. \tag{36c}$$

We first employ the alternating optimization (AO) algorithm to obtain the optimal φ and $\mathcal{K}$ for given $\mathcal{J}$.

A. SU-based AO (AOSU) Algorithm

The optimal $\mathcal{J}$ can be obtained by applying the SU algorithm proposed in IV-B[9]. Then, the optimal solutions for φ and $\mathcal{K}$ can be obtained by AO with given $\mathcal{J}$. Specifically, the details for obtaining the optimal optimal φ and $\mathcal{K}$ in each iteration in AO are elaborated as follows.

1) Optimal φ^*: By giving $\mathcal{K}$ and $\mathcal{J}$, $\mathbf{P6}$ can be decomposed as follows:

$$\mathbf{P7}.\ \min_{\varphi}\left\{\Lambda\left(\{1\}_{n\in\mathcal{K}},\varphi,\mathcal{K},\mathcal{J}\right)\right\}. \tag{37}$$

The optimal solution to problem $\mathbf{P7}$ is given as follows.

Lemma 8. *With any given $\mathcal{K}$ and $\mathcal{J}$, the optimal solution for φ^* to problem $\mathbf{P7}$ is given by*

$$\varphi^*=\frac{|\mathcal{K}|\sum_{n\in\mathcal{K}}h_{n,B}^2P_n+\sigma_{B,\mathcal{J}}^{tot}}{|\mathcal{K}|\sum_{n\in\mathcal{K}}h_{n,B}\sqrt{P_n}\|\boldsymbol{g}_n\|_2}. \tag{38}$$

Proof: By introducing $\hat{\varphi}=\frac{1}{\varphi}$, problem $\mathbf{P7}$ is recast as the following convex quadratic problem:

$$\mathbf{P8}.\ \min_{\hat{\varphi}}\left\{\frac{1}{|\mathcal{K}|}\sum_{n\in\mathcal{K}}\left(h_{n,B}\sqrt{P_n}\hat{\varphi}-\|\boldsymbol{g}_n\|_2\right)^2+\frac{\sigma_{B,\mathcal{J}}^{tot}}{|\mathcal{K}|^2}\hat{\varphi}^2\right\}.$$

By setting the first derivative of the objective function in problem $\mathbf{P8}$ to zero, we can obtain the optimal solution to problem $\mathbf{P8}$, and accordingly get the optimal solution to problem $\mathbf{P7}$ with $\varphi^*=\frac{1}{\hat{\varphi}^*}$. ■

2) Optimal $\mathcal{K}$: With given $\mathcal{J}$ and φ, $\mathbf{P6}$ can be re-expressed as

$$\mathbf{P9}.\ \min_{\mathcal{K}}\left\{\Lambda\left(\{1\}_{n\in\mathcal{K}},\varphi,\mathcal{K},\mathcal{J}\right)\right\} \tag{39}$$
$$\text{s.t. } (36a),(36b),(36c). \tag{39a}$$

By defining $\mathcal{R}_{\mathcal{J}}=\mathcal{N}/\mathcal{J}$ and introducing $k_n\in\{0,1\},n\in\mathcal{R}_{\mathcal{J}}$ to indicate whether or not device $n\in\mathcal{R}_{\mathcal{J}}$ is scheduled as a learner, the optimization can be re-expressed as

$$\mathbf{P10}.\ \min_{\{k_n\}_{n\in\mathcal{R}_{\mathcal{J}}}}\left\{\frac{1}{\sum_{n\in\mathcal{R}_{\mathcal{J}}}k_n}\sum_{n\in\mathcal{R}_{\mathcal{J}}}k_nA_n^2\right.$$

[9]Considering the complexity of the SU algorithm, we will introduce an alternative solution with low complexity in the next subsection. The effectiveness of this alternative solution will be demonstrated through a comparison with the SU algorithm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

$$+\frac{B}{\left(\sum_{n\in\mathcal{R}_\mathcal{J}} k_n\right)^2}+4\big(1-\frac{\sum_{n\in\mathcal{R}_\mathcal{J}} k_n}{N}\big)^2G^2\Bigg\} \tag{40}$$

$$\textbf{s.t. } k_n\in\{0,1\}, n\in\mathcal{R}_\mathcal{J}, \tag{40a}$$

$$4d\kappa^2\frac{h_{n,B}^2P_n}{\epsilon_n^2}k_n\le\sigma_{B,\mathcal{J}}^{tot}, n\in\mathcal{R}_\mathcal{J}, \tag{40b}$$

$$d\varpi^2\frac{k_nh_{n,B}^2P_n}{\|\boldsymbol{g}_n\|_2^2}\big(\sum_{n\in\mathcal{R}_\mathcal{J}}k_n\big)^2\le\sigma_{E,\mathcal{J}}^{tot}, n\in\mathcal{R}_\mathcal{J}, \tag{40c}$$

where $A_n=\frac{h_{n,B}\sqrt{P_n}}{\varphi}-\|\boldsymbol{g}_n\|_2$ and $B=\frac{\sigma_{B,\mathcal{J}}^{tot}}{\varphi^2}$. Next, we can decompose $\mathbf{P10}$ into a set of $|\mathcal{R}_\mathcal{J}|$ optimization sub-problems. This decomposition is achieved by introducing a crucial constraint: for each i within the range $1\le i\le|\mathcal{R}_\mathcal{J}|$, we add the constraint $\sum_{n\in\mathcal{R}_\mathcal{J}}k_n=i$ to the i-th sub-problem. Specifically, the i-th, $1\le i\le|\mathcal{R}_\mathcal{J}|$ sub-problem can be given by

$$\mathbf{P10}\,(i).\min_{\{k_n^i\}_{n\in\mathcal{R}_\mathcal{J}}}\Bigg\{\frac{1}{i}\sum_{n\in\mathcal{R}_\mathcal{J}}k_n^iA_n^2+\frac{B}{i^2}+4\big(1-\frac{i}{N}\big)^2G^2\Bigg\} \tag{41}$$

$$\textbf{s.t. } k_n^i\in\{0,1\}, n\in\mathcal{R}_\mathcal{J}, \tag{41a}$$

$$4d\kappa^2\frac{h_{n,B}^2P_n}{\epsilon_n^2}k_n^i\le\sigma_{B,\mathcal{J}}^{tot}, n\in\mathcal{R}_\mathcal{J}, \tag{41b}$$

$$d\varpi^2\frac{h_{n,B}^2P_n}{\|\boldsymbol{g}_n\|_2^2}k_n^i\le\frac{\sigma_{E,\mathcal{J}}^{tot}}{i^2}, n\in\mathcal{R}_\mathcal{J}, \tag{41c}$$

$$\sum_{n\in\mathcal{R}_\mathcal{J}}k_n^i=i. \tag{41d}$$

By relaxing constraint (41a) as $0\le k_n^i\le 1, n\in\mathcal{R}_\mathcal{J}$, $\mathbf{P10}\,(i)$ can be transformed to a convex problem and can be solved by convex optimization tools. Upon solving the set of $|\mathcal{R}_\mathcal{J}|$ convex problems, we acquire $|\mathcal{R}_\mathcal{J}|$ optimal solutions for the corresponding sub-problems, denoted as $\mathcal{K}_i^*$ where $1\le i\le|\mathcal{R}_\mathcal{J}|$. Clearly, the optimal solution for the original problem P10 is hidden in these solutions. By defining $\Theta\left(\left\{k_n^i\right\}_{n\in\mathcal{R}_\mathcal{J}}\right)=\frac{1}{i}\sum_{n\in\mathcal{R}_\mathcal{J}}k_n^iA_n^2+\frac{B}{i^2}+4\left(1-\frac{i}{N}\right)^2G^2$, we obtain the optimal solution to problem $\mathbf{P10}$ by $\{k_n^*\}_{n\in\mathcal{R}_\mathcal{J}}=\left\{k_n^i\right\}_{n\in\mathcal{R}_\mathcal{J}}$ where $i=\arg\min_{1\le i\le|\mathcal{R}_\mathcal{J}|}\left\{\Theta\left(\left\{k_n^i\right\}_{n\in\mathcal{R}_\mathcal{J}}\right)\right\}$.

B. SP-SJR-based LC Solution for B-SP-OTA-FL

In this subsection, we introduce an approach of reduced complexity to address problem $\mathbf{P6}$ by systematically exploring all potential values for the size of $\mathcal{K}$. For a given size of $\mathcal{K}$, we can first jointly design $\mathcal{K}$ and $\mathcal{J}$ with the consideration of privacy and security. Then, based on the $\mathcal{K}$ and $\mathcal{J}$, the optimal φ can be obtained by applying Lemma 8. The optimal solution can be obtained by searching the enumerated potential solutions. Specifically, we consider the size of $\mathcal{K}$ from 1 to N. For each value, we schedule the $|\mathcal{K}|$ devices that result in the least stringent privacy and security constraints. Subsequently, if the remaining devices can meet the privacy and security requirements, we derive the optimal jammer set by applying the SP-SJR algorithm. Finally, the optimal solution can be obtained by exhaustively searching through all possible solution cases. Let $Pr=[pr_1,\cdots pr_n,\cdots pr_N]$ and $Se=[se_1,\cdots se_n,\cdots se_N]$ where $pr_n=4d\kappa^2\frac{h_{n,B}^2P_n}{\epsilon_n^2}$ and $se_n=d\varpi^2\frac{h_{n,B}^2P_n}{\|\boldsymbol{g}_n\|_2^2}$. Then, we define $Prse=[prse_1,\cdots,prse_n,\cdots prse_N]$ where $prse_n=\max\{pr_n,se_n\}$. The details of the proposed LC solution are summarized in Algorithm 4 where $\boldsymbol{s}_\mathcal{N}^{prse}=S(\mathcal{N};prse_n)$.

Algorithm 4 SP-SJR-based LC scheme for solving $\mathbf{P6}$.

Input: Given $\{(\epsilon_n,\zeta)\}_{n=1}^N$ and ϖ.
Output: $\mathcal{K}^*$, $\mathcal{J}^*$, φ^*.

```
for i in {1, ..., N} do
    Let K_i^* = {s_N^prse[n] | n >= N - i + 1}, R_i = N/K_i^*, J_i = R_i, and w = s_N^prse[N - i + 1].
    if Constraints (36b) and (36c) can be satisfied then
        Obtain J_i^* by employing Algorithm 2 where sigma^pr = pr_w - sigma_B^2 and sigma^se = se_w - sigma_B^2.
        Obtain varphi_i^* using Lemma 8.
    end if
end for
Obtain K*, J*, varphi* = K_i^*, J_i^*, varphi_i^* where i = arg min_{1<=i<=N} {Lambda({1}_{n in K_i^*}, varphi_i^*, K_i^*, J_i^*)}.
```

Considering that the SP-SJR algorithm has a computational complexity of $\mathcal{O}(N)$, the proposed LC solution, with a complexity of $\mathcal{O}(N^2)$, provides an efficient method for solving Problem $\mathbf{P6}$.

VI. Simulation Results

This section presents simulation results to assess the impact of privacy and security on the learning process and to compare the performance of the proposed jamming-aided SP-OTA-FL schemes with state-of-the-art approaches that do not incorporate jamming.

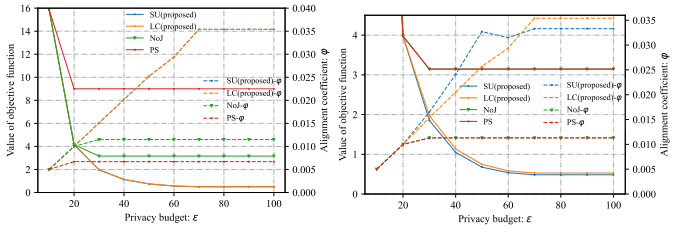
A. Simulation Setting

We evaluate our proposed scheme by training a convolutional neural network (CNN) on the popular MNIST dataset used for handwritten digit classification. In particular, CNN consists of two 5×5 convolution layers with the rectified linear unit (ReLU) activation. The two convolution layers have 10 and 20 channels respectively, and each layer has 2×2 max pooling, a fully connected layer with 50 units and ReLU activation, and a log-softmax output layer, in which case $d=21840$. The learning rate is set as $\eta=0.1$. We set σ_B^2 and σ_E^2 both to 1, and each device's maximum transmission power is limited to 30dBm.

B. Evaluation of jamming-aided UB-SP-OTA-FL

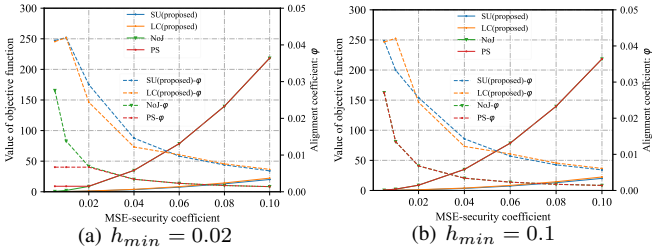
In this section, we present simulation results to show how privacy and security constraints affect the learning process of UB-OTA-FL and validate the performance of the proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications



(a) $h_{min} = 0.02$ (b) $h_{min} = 0.1$

Fig. 2: The optimal value of the objective function and the alignment coefficient φ under varying privacy budgets ϵ.

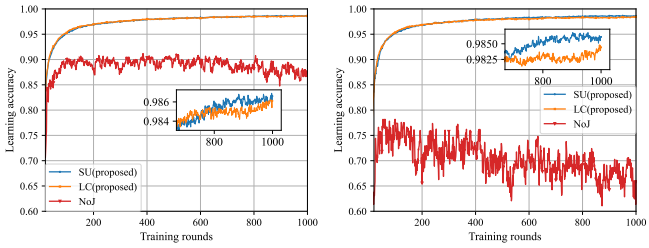


(a) $h_{min} = 0.02$ (b) $h_{min} = 0.1$

Fig. 3: The optimal value of the objective function and the alignment coefficient φ under varying MSE-security levels.

schemes by comparing it with the two schemes: no jamming-aided (NoJ) scheme [29] and the power scaling (PS) scheme [18]. Notably, both the NoJ and PS schemes operate without jammers, relying on the reduction of the alignment coefficient to ensure privacy and security. Specifically, the alignment coefficient and the device scheduling are jointly optimized in NoJ while only the alignment coefficient is optimized in PS.

In Fig. 2, we depict the alignment coefficient φ and the objective function value in $\mathbf{P}2$ concerning the privacy budget ϵ, with different worst channel conditions $h_{\min}$. To highlight the impact of privacy, we maintain ϖ at a value of 0.012. Firstly, the proposed LC and SU algorithm consistently achieve nearly identical performance and outperform the benchmarks. Generally, the alignment coefficient φ is constrained by the channel condition, privacy and security constraint as shown in (30). In the scenario where $\epsilon = 10$, these schemes achieve the same performance where the alignment coefficient is constrained by the strict privacy requirement. It is readily understandable that all the devices in NoJ and PS are scheduled to participate in the training process to achieve a better learning performance when φ is limited to a small value for satisfying the stringent privacy constraint. Regarding the proposed jamming-aided schemes, it is possible to increase the alignment coefficient φ by involving certain devices as jammers, resulting in a reduced number of active learners. However, this increase in φ has a limited contribution to decrease the objective function when weighed against the reduction in the number of participants. Consequently, the proposed jamming-aided schemes yield comparable results to those achieved by the NoJ and PS strategies. In the cases of $\epsilon \geq 20$, the alignment coefficient φ in PS is limited by the worst channel condition as the alignment coefficient and objective function do not show any improvement with the increased privacy budgets. While it is possible to enhance the alignment coefficient with decreased privacy requirements and involving fewer participants, a tradeoff exists between achieving an improved alignment coefficient φ and having a reduced number of participants to attain the optimal value of the objective function. Consequently, when $\epsilon \geq 30$, both the alignment coefficient and the objective function in NoJ do not demonstrate any improvement, even with a larger privacy budget. Considering



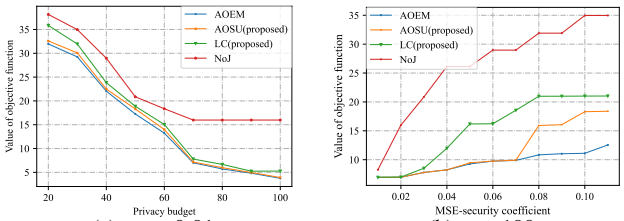
(a) $\epsilon = 40, \varpi = 0.05$ (b) $\epsilon = 20, \varpi = 0.06$

Fig. 4: The learning performance with different schemes.

the two proposed jamming-aided schemes, the advantages of enhancing alignment through the incorporation of jamming techniques outweighed the drawback of having fewer learners. As a result, performance can be further enhanced by increasing the privacy budget when $\varphi \leq 70$. However, in scenarios where $\varphi \geq 70$, the outcomes are constrained by security requirements, which implies that no additional improvement can be achieved through an increased privacy budget.

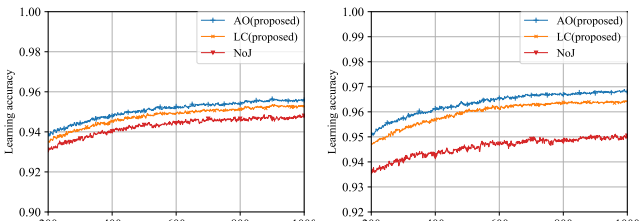
Fig. 3 illustrates the alignment coefficient φ and the objective function value in $\mathbf{P}2$ with different security requirements where the privacy budget is set to $\epsilon = 80$. The two proposed jamming-aided schemes consistently outperform the benchmarks, with significant performance superiority observed when the security requirement is stringent. It is worth noting that in pursuit of stringent security objectives, a substantial reduction in the alignment coefficient φ within NoJ and PS scenarios can have a noteworthy adverse impact on the utility of the aggregated gradient at the BS. In contrast, the utility of the aggregated gradient at the BS may not experience such a substantial decline when security requirements are met through the strategic design of jammers, emphasizing their impact on eavesdroppers rather than on the BS. This underscores the extra effectiveness of the proposed jamming framework when addressing strict security constraints.

In Fig. 4, we present a comparison of learning accuracy between the proposed schemes and the NoJ scheme. For this analysis, we exclude the consideration of PS, as it often yields identical outcomes to NoJ in the majority of privacy and security constraint cases. In the context of UB-OTA-FL scenarios, the proposed schemes ensure the fulfillment



(a) $\varpi = 0.01$ (b) $\epsilon = 100$

Fig. 5: The optimal value of the objective function under varying privacy and security requirements.



(a) $\epsilon = 30, \varpi = 0.01$ (b) $\epsilon = 100, \varpi = 0.1$

Fig. 6: The learning performance with different schemes.

of privacy and security requirements through a combined optimization of reducing the alignment coefficient φ and adjusting the scheduling of jammers. However, in scenarios without jammers, privacy and security are achieved primarily by reducing the alignment coefficient φ. Consequently, our two proposed jamming-aided schemes demonstrate a marginal decrease in accuracy as privacy and security requirements intensify, while the NoJ scenario experiences a substantial drop in accuracy under the same conditions, aligning with the findings in Fig. 3.

C. Evaluation of jamming-aided B-SP-OTA-FL

We evaluate the performance of the jamming-aided schemes by comparing it with exhaustive search-based alternative optimization (AOEM) and the NoJ scheme.

In Fig. 5, we plot the value of the objective function in B-OTA-FL with varying privacy budgets and security requirements in Fig. 5(a) and Fig. 5(b), respectively. In Fig. 5(a), the security coefficient is set to $\varpi = 0.01$ and the privacy budget is set to $\epsilon = 100$ in Fig. 5(b). In Fig. 5(a) and Fig. 5(b), the proposed AOSU scheme achieves nearly identical results as AOES, which demonstrates the effectiveness of the proposed SU scheme. In B-OTA-FL scenarios, privacy and security in the NoJ scheme are maintained by scheduling eligible devices that take advantage of channel noise for protection. In contrast, the jamming-aided schemes ensure privacy and security by orchestrating the joint scheduling of learners and jammers. Both of the proposed schemes outperform the NoJ scheme, with the performance gap being especially pronounced in Fig. 5(b) when the security requirements are stringent. This enhanced performance can be attributed to the inclusion of jamming techniques. Just as in the case of UB-SP-OTA-FL, the reduced participation of devices can directly impact the contribution of the gradient at the BS, while jamming can be strategically designed to introduce more distortion to eavesdroppers, enhancing the security of the system.

Fig. 6 depicts the learning accuracy for both the proposed jamming-aided schemes and the NoJ scheme. In Fig. 6(a), we set the privacy budget and security requirement to $\epsilon = 30$ and $\varpi = 0.01$, while in Fig. 6(b), these values are set to $\epsilon = 100$ and $\varpi = 0.1$. Fig. 6(a) illustrates scenarios with more stringent privacy constraints, while Fig. 6(b) describes scenarios with stricter security constraints. In both scenarios, our jamming-aided schemes consistently outperform the non-jamming approach, demonstrating the effectiveness of our proposed schemes. Specifically, in Fig. 6(a), the proposed schemes outperform NoJ schemes because they can schedule more devices to participate in the training process by leveraging jamming as a privacy protection measure. In contrast, NoJ schemes only schedule devices whose privacy can be protected by the channel noise, thereby compromising the learning performance, especially in cases where only a very limited number of devices meet the privacy requirement. This superiority of our proposed jamming-aided approach is particularly pronounced in scenarios with heightened security requirements as in Fig. 6 (b). This is attributed to our strategic jamming design. Specifically, NoJ approaches ensure security by scheduling the devices with poor channel conditions, as demonstrated in security analysis. However, this action inevitably degrades the gradient quality at both the BS and the eavesdropper. As the security requirement increases, the gradient quality will be significantly compromised. In contrast, our jamming-aided approach strategically schedules devices close to potential eavesdroppers as jammers. This effectively degrades gradient quality at the eavesdropper while causing less distortion to the gradient at the BS, thereby being particularly effective for strict security constraint.

VII. Extension to Asynchronous FL Scenarios

In asynchronous FL, devices update and transmit their local gradients at varying times due to differences in computational capabilities, communication latency, or energy constraints. There are mainly two types of asynchronous FL in terms of the number of the devices participating during one aggregation:

- Traditional asynchronous FL algorithms, where only one device updates its gradients to the server in each aggregation round;
- Semi-asynchronous FL algorithms, where devices are grouped into clusters to perform synchronous aggregation within each cluster, while clusters themselves operate asynchronously.

A. Case 1: Traditional Asynchronous FL

For traditional asynchronous FL algorithms, consider the t-th aggregation round, where only one device a is ac-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TWC.2025.3550159, IEEE Transactions on Wireless Communications

tively involved in transmitting its gradients to the server. When the gradient is computed on $\mathcal{D}_a$, the signal received at the BS can be expressed as: $\boldsymbol{y}^t = \frac{h_{a,B}^t\sqrt{\eta_a P_a}}{\|\boldsymbol{g}_a^t\|_2}\boldsymbol{g}_a^t + \sum_{n\in\mathcal{J}^t} h_{n,B}^t\sqrt{P_n}\boldsymbol{e}_n^t + \boldsymbol{r}_B^t$. Now, assume there is a data sample changed in the dataset $\mathcal{D}_a$, resulting in a modified dataset $\mathcal{D}_a'$ and corresponding gradients $(\boldsymbol{g}_a^t)'$. Consequently, the signal received at the BS in this case becomes: $(\boldsymbol{y}^t)' = \frac{h_{a,B}^t\sqrt{\eta_a P_a}}{\|(\boldsymbol{g}_a^t)'\|_2}(\boldsymbol{g}_a^t)' + \sum_{n\in\mathcal{J}^t} h_{n,B}^t\sqrt{P_n}\boldsymbol{e}_n^t + \boldsymbol{r}_B^t$. Then, the sentivitiy of the propsoed algorithms is given by $\Delta S_a^t = \boldsymbol{y}^t - (\boldsymbol{y}^t)' = h_{a,B}^t\sqrt{\eta_a P_a}\left|\frac{\boldsymbol{g}_a^t}{\|\boldsymbol{g}_a^t\|_2} - \frac{(\boldsymbol{g}_a^t)'}{\|(\boldsymbol{g}_a^t)'\|_2}\right| \leq 2h_{a,B}^t\sqrt{\eta_a P_a}$. This result demonstrates that the privacy analysis for traditional asynchronous FL represents a specific case of the analysis presented in Section III-A.

B. Case 2: Semi-Asynchronous FL

In semi-asynchronous FL algorithms, devices within each cluster aggregate their gradients synchronously, while clusters communicate asynchronously. Since the aggregation process within each cluster is still synchronous, it is directly compatible with the proposed SP-OTA-FL framework and its theoretical analyses.

VIII. Conclusion

In this paper, we have introduced a novel framework called SP-OTA-FL, aimed at strengthening privacy and security of OTA-FL by allocating certain devices as jammers. Leveraging theoretical insights from privacy, security and convergence analyses, we have devised optimization problems to delve into the optimal design for SP-OTA-FL in two common scenarios: UB-OTA-FL and B-OTA-FL. We have also developed efficient schemes to tackle these problems. Compared to schemes that do not take jamming into account, the proposed jamming-aided SP-OTA-FL can significantly improve security. This improvement is achieved by strategically designating devices located closer to potential eavesdroppers as jammers, as opposed to reducing the power of the gradient, which might lead to a more distorted aggregated gradient at the BS.

Appendix A
Proof of Lemma 1

We use index a instead of n to avoid confusion between the specific index of device n and the notation n in the summation. Based on the definition of sensitivity, one has

$$\Delta S_a^t = h_{a,B}^t\sqrt{\eta_a P_a}\left|\frac{\boldsymbol{g}_a^t}{\|\boldsymbol{g}_a^t\|_2} - \frac{(\boldsymbol{g}_a^t)'}{\|(\boldsymbol{g}_a^t)'\|_2}\right| \leq 2h_{a,B}^t\sqrt{\eta_a P_a}, \tag{42}$$

where the last inequality stems from Triangular Inequality. The variance of the aggregated noise at the BS is given by $\frac{\sum_{n\in\mathcal{J}^t}\left(h_{n,B}^t\right)^2 P_n+\sigma_B^2}{d}$. By replacing a with n, one completes the proof of Lemma 1.

Appendix B
Proof of Lemma 2

Firstly, since the elements in $\boldsymbol{g}_n^t$ are uniformly distributed in $[a,b]$, the $\boldsymbol{g}_{ave}^t$ follows the same distribution in $[a,b]$. For analysis, we define $\tilde{\mathcal{E}}^t : (\boldsymbol{g}_n^t)_{n\in\mathcal{K}^t} \to \tilde{\boldsymbol{z}}^t \in \tilde{\mathcal{Z}}$ where $\tilde{\boldsymbol{z}}^t = \sum_{n\in\mathcal{K}^t}\gamma^t\boldsymbol{g}_n^t + \boldsymbol{r}_{E,Tot}^t$. Assume that the variance of $\tilde{\boldsymbol{z}}^t$ is σ. Following Lemma 3 and Lemma 4 in [28], the minimum MSE estimator $e\left(\tilde{\boldsymbol{z}}^t\right)$ for estimating $\boldsymbol{g}_{ave}^t$ from the observations $\tilde{\boldsymbol{z}}^t$ satisfies $\mathbb{E}\left[\left(\boldsymbol{g}_{ave}^t - e\left(\tilde{\boldsymbol{z}}^t\right)\right)^2\right] = \sigma\Xi\left(\frac{b-a}{\sqrt{\sigma}}\right)$. The lowest-variance unbiased estimator is given by

$$e\left(\tilde{\boldsymbol{z}}^t\right) = \boldsymbol{g}_{ave}^t + \frac{1}{|\mathcal{K}^t|\,\gamma^t}\boldsymbol{r}_{E,Tot}^t, \tag{43}$$

with the variance $(\varpi^t)^2 = \frac{\sum_{n\in\mathcal{J}^t}\left(h_{n,E}^t\right)^2 P_n+\sigma_E^2}{d(|\mathcal{K}^t|\gamma^t)^2}$. In reality, the minimal value of ϖ^t is achieved when $\gamma^t = \max_{n\in\mathcal{K}^t}\left\{\frac{h_{n,B}^t\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\right\}$ due to the peak transmit power constraint. It thus follows from (43) and Definition 2 that $\tilde{\mathcal{E}}^t$ guarantees $\left(\tilde{\mathcal{E}}^t, (\varpi^t)^2\,\Xi\left(\frac{b-a}{\varpi^t}\right)\right)$. On the other hand, one has $\mathbb{E}\left[\left\|e\left(\tilde{\boldsymbol{z}}^t\right) - \boldsymbol{g}_{ave}^t\right\|^2\right] = \frac{1}{\left(|\mathcal{K}|^t\gamma^t\right)^2}\mathbb{E}\left[\left\|\boldsymbol{r}_{E,Tot}^t\right\|^2\right]$. Similarly, we also have

$$\begin{aligned}\mathbb{E}\left[\left\|e\left(\boldsymbol{z}^t\right) - \boldsymbol{g}_{ave}^t\right\|^2\right] &\overset{(a)}{=} \frac{1}{\left(|\mathcal{K}|^t\gamma^t\right)^2}\mathbb{E}\left[\left\|\boldsymbol{r}_{E,Tot}^t\right\|^2\right] \\ &+ \frac{1}{|\mathcal{K}^t|^2}\mathbb{E}\left[\left\|\sum_{n\in\mathcal{K}^t}\left(\frac{h_{n,E}^t\sqrt{\eta_n P_n}}{\gamma^t} - 1\right)\boldsymbol{g}_n^t\right\|^2\right],\end{aligned} \tag{44}$$

where (a) comes from $\mathbb{E}\left[\boldsymbol{r}_{E,Tot}^t\right] = 0$. Obviously, $\mathbb{E}\left[\left\|e\left(\tilde{\boldsymbol{z}}^t\right) - \boldsymbol{g}_{ave}^t\right\|^2\right]$ is smaller than $\mathbb{E}\left[\left\|e\left(\boldsymbol{z}^t\right) - \boldsymbol{g}_{ave}^t\right\|^2\right]$, therefore, $e\left(\tilde{\boldsymbol{z}}^t\right)$ is a closer estimate of $\boldsymbol{g}_{ave}^t$. Then, $e\left(\boldsymbol{z}^t\right)$ has a larger variance and can achieve at least $\left(\tilde{\mathcal{E}}^t, (\varpi^t)^2\,\Xi\left(\frac{b-a}{\varpi^t}\right)\right)$-MSE-security. Alternatively, from the communication point of view, one can also get that $\tilde{\boldsymbol{z}}^t$ could have a better recovery of gradient than $\boldsymbol{z}^t$ because of a higher SNR as $\gamma^t = \max_{n\in\mathcal{K}^t}\left\{\frac{h_{n,B}^t\sqrt{\eta_n P_n}}{\|\boldsymbol{g}_n^t\|_2}\right\}$. Therefore, if $\tilde{\boldsymbol{z}}^t$ can guarantee at least $\left(\tilde{\mathcal{E}}^t, (\varpi^t)^2\,\Xi\left(\frac{b-a}{\varpi^t}\right)\right)$-MSE-security, then so can $\boldsymbol{z}^t$. Then, we complete the proof of Lemma 2.

Appendix C
Proof of Lemma 3

We first derive the upper bound of the expected loss function $\mathbb{E}\left[L\left(\boldsymbol{m}^{t+1}\right)\right]$ as follows:

$$\begin{aligned}&\mathbb{E}\left[L\left(\boldsymbol{m}^{t+1}\right)\right] \\ \leq&\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right] + \mathbb{E}\left[\left\langle\nabla L\left(\boldsymbol{m}^t\right), \boldsymbol{m}^{t+1} - \boldsymbol{m}^t\right\rangle\right] \\ &+ \frac{\theta}{2}\mathbb{E}\left[\left\|\boldsymbol{m}^{t+1} - \boldsymbol{m}^t\right\|_2^2\right] \\ \leq&\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right] - \tau\mathbb{E}\left[\left\langle\nabla L\left(\boldsymbol{m}^t\right), \nabla L\left(\boldsymbol{m}^t\right)\right\rangle\right] \\ &- \tau\mathbb{E}\left[\left\langle\nabla L\left(\boldsymbol{m}^t\right), \Delta\boldsymbol{g}_{err}^t\right\rangle\right] \\ &+ \frac{\tau^2\theta}{2}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right) + \Delta\boldsymbol{g}_{err}^t\right\|_2^2\right] \\ =&\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right] - \tau\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right] + \frac{\tau^2\theta}{2}\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\end{aligned}$$

$$+\frac{\tau^2\theta}{2}\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{err}^t\right\|_2^2\right]-\tau\left(1-\tau\theta\right)\mathbb{E}\left[\left\langle\nabla L\left(\boldsymbol{m}^t\right),\Delta \boldsymbol{g}_{err}^t\right\rangle\right]$$
$$\overset{(a)}{\leq}\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right]-\frac{\tau}{2}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right]+\frac{\tau}{2}\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{err}^t\right\|_2^2\right]$$
$$\overset{(b)}{\leq}\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right]-\frac{\tau}{2}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right]$$
$$+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{com}^t\right\|_2^2\right]+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{ds}^t\right\|_2^2\right], \tag{45}$$

where (a) stems from that

$$-\mathbb{E}\left[\left\langle\nabla L\left(\boldsymbol{m}^t\right),\Delta \boldsymbol{g}_{err}^t\right\rangle\right]$$
$$\leq\frac{\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right]}{2}+\frac{\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{err}^t\right\|_2^2\right]}{2}, \tag{46}$$

and (b) is from

$$\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{err}^t\right\|_2^2\right]\leq 2\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{com}^t\right\|_2^2\right]+2\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{ds}^t\right\|_2^2\right]. \tag{47}$$

By applying the fact that $\mathbb{E}[\boldsymbol{e}_n^t]=\mathbb{E}[\boldsymbol{r}_B^t]=0$, the communication MSE is bounded by

$$\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{com}^t\right\|_2^2\right]$$
$$\leq\mathbb{E}\Big[\Big\|\frac{1}{|\mathcal{K}^t|}\sum_{n\in\mathcal{K}^t}\Big(\frac{h_{n,B}^t\sqrt{\eta_nP_n}}{\varphi^t\left\|\boldsymbol{g}_n^t\right\|_2}-1\Big)\boldsymbol{g}_n^t\Big\|_2^2\Big]$$
$$+\mathbb{E}\Big[\Big\|\frac{1}{|\mathcal{K}^t|\,\varphi^t}\Big(\sum_{n\in\mathcal{J}^t}h_{n,B}^t\sqrt{P_n}\boldsymbol{e}_n^t+\boldsymbol{r}_B^t\Big)\Big\|_2^2\Big]$$
$$\overset{(a)}{\leq}\frac{1}{|\mathcal{K}^t|}\sum_{n\in\mathcal{K}^t}\Big(\frac{h_{n,B}^t\sqrt{\eta_nP_n}}{\varphi^t\left\|\boldsymbol{g}_n^t\right\|_2}-1\Big)^2\mathbb{E}\left[\left\|\boldsymbol{g}_n^t\right\|_2^2\right]$$
$$+\frac{1}{\left(|\mathcal{K}^t|\,\varphi^t\right)^2}\Big(\sum_{n\in\mathcal{J}^t}\left(h_{n,B}^t\right)^2P_n\mathbb{E}\left[\left\|\boldsymbol{e}_n^t\right\|_2^2\right]+\mathbb{E}\left[\left\|\boldsymbol{r}_B^t\right\|_2^2\right]\Big)$$
$$\leq\frac{1}{|\mathcal{K}^t|}\sum_{n\in\mathcal{K}^t}\Big(\frac{h_{n,B}^t\sqrt{\eta_nP_n}}{\varphi^t}-\left\|\boldsymbol{g}_n^t\right\|_2\Big)^2$$
$$+\frac{1}{\left(|\mathcal{K}^t|\,\varphi^t\right)^2}\Big(\sum_{n\in\mathcal{J}^t}\left(h_{n,B}^t\right)^2P_n+\sigma_B^2\Big), \tag{48}$$

where (a) comes from Jensen's inequality. The bound of the device scheduling MSE is bounded by

$$\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{ds}^t\right\|_2^2\right]$$
$$\overset{(a)}{\leq}\Big(\Big(\frac{1}{|\mathcal{K}^t|}-\frac{1}{N}\Big)\sum_{n\in\mathcal{K}^t}\left\|\boldsymbol{g}_n^t\right\|_2+\frac{1}{N}\sum_{n\in\mathcal{N}/\mathcal{K}^t}\left\|\boldsymbol{g}_n^t\right\|_2\Big)^2$$
$$\overset{(b)}{\leq}4\Big(1-\frac{|\mathcal{K}^t|}{N}\Big)^2G^2, \tag{49}$$

where (a) comes from that $\|\boldsymbol{a}+\boldsymbol{b}\|_2^2\leq\left(\|\boldsymbol{a}\|_2+\|\boldsymbol{b}\|_2\right)^2$ and (b) is from the Assumption 1.

Appendix D
Proof of Theorem 1

Under Assumption 3, we could derive a useful result, i.e., $\|\nabla L\left(\boldsymbol{\iota}\right)\|_2^2\geqslant 2\rho\left[L\left(\boldsymbol{\iota}\right)-L\left(\boldsymbol{\iota}^*\right)\right]$. Then, we have

$$\mathbb{E}\left[L\left(\boldsymbol{m}^{t+1}\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{m}^*\right)\right]$$
$$\leq\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{m}^*\right)\right]-\frac{\tau}{2}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right]$$
$$+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{com}^t\right\|_2^2\right]+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{ds}^t\right\|_2^2\right]$$
$$\leq\left(1-\rho\tau\right)\left(\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{m}^*\right)\right]\right)+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{com}^t\right\|_2^2\right]$$
$$+\tau\mathbb{E}\left[\left\|\Delta \boldsymbol{g}_{ds}^t\right\|_2^2\right]$$
$$\leq\left(1-\rho\tau\right)^{t+1}\left(\mathbb{E}\left[L\left(\boldsymbol{m}^0\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{m}^*\right)\right]\right)$$
$$+\tau\Lambda\sum_{i=0}^{t}\left(1-\rho\tau\right)^i$$
$$\leq\left(1-\rho\tau\right)^{t+1}\mathbb{E}\left[L\left(\boldsymbol{m}^0\right)-L\left(\boldsymbol{m}^*\right)\right]$$
$$+\frac{1-\left(1-\rho\tau\right)^{t+1}}{\rho}\max_{t\in T}\left\{\Lambda\left(\left\{\eta_n\right\}_{n\in\mathcal{K}^t},\varphi^t,\mathcal{K}^t,\mathcal{J}^t\right)\right\}. \tag{50}$$

By replacing $t+1$ as T, we finish the proof of Theorem 1.

Appendix E
Proof of Theorem 2

Accroding to Lemma 3, we have

$$\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right]\leq\frac{2\left[\mathbb{E}\left[L\left(\boldsymbol{m}^t\right)\right]-\mathbb{E}\left[L\left(\boldsymbol{m}^{t+1}\right)\right]\right)}{\tau}+\tau\Lambda. \tag{51}$$

By summing t from 0 to T, we complete the proof of Theorem 2 as follows:

$$\frac{1}{T}\sum_{t=0}^{T}\mathbb{E}\left[\left\|\nabla L\left(\boldsymbol{m}^t\right)\right\|_2^2\right] \tag{52}$$
$$\leq\frac{2\mathbb{E}\left[L\left(\boldsymbol{m}^0\right)-L\left(\boldsymbol{m}^T\right)\right]}{\tau T}+2\Lambda$$
$$\overset{(a)}{\leq}\frac{2\mathbb{E}\left[L\left(\boldsymbol{m}^0\right)-L\left(\boldsymbol{m}^*\right)\right]}{\tau T}+2\Lambda, \tag{53}$$

where (a) comes from the fact that $L\left(\boldsymbol{m}^*\right)\leq L\left(\boldsymbol{m}^T\right)$.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Eng.*, vol. 8, pp. 33–41, 2022.

[3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2017, pp. 1273–1282.

[4] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021.

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

[6] 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, 2019.

[7] M. Goldenbaum, H. Boche, and S. Stańczak, "Harnessing interference for analog function computation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4893–4906, 2013.

[8] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Security Privacy (SP)*, 2017, pp. 3–18.

[9] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *Proc. IEEE Symp. Security Privacy (SP)*, 2019, pp. 691–706.
[10] L. Zhu and S. Han, "Deep leakage from gradients," in *Federated learning*. Springer, 2020, pp. 17–31.
[11] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Security*, 2015, pp. 1322–1333.
[12] B. Zhao, K. R. Mopuri, and H. Bilen, "IDLG: Improved deep leakage from gradients," *arXiv preprint arXiv:2001.02610*, 2020.
[13] A. Sharma and N. Marchang, "A review on client-server attacks and defenses in federated learning," *Computers & Security*, p. 103801, 2024.
[14] J. Liao, Z. Chen, and E. G. Larsson, "Over-the-air federated learning with privacy protection via correlated additive perturbations," in *Proc. IEEE Annu. Allerton Conf. Commun. Control Comput.* IEEE, 2022, pp. 1–8.
[15] S. Park and W. Choi, "On the differential privacy in federated learning based on over-the-air computation," *IEEE Trans. Wireless Commun.*, 2023.
[16] H. Liu, J. Yan, and Y.-J. A. Zhang, "Differentially private over-the-air federated learning over mimo fading channels," *IEEE Trans. Wireless Commun.*, 2024.
[17] M. Seif, R. Tandon, and M. Li, "Wireless federated learning with local differential privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 2604–2609.
[18] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Differentially private AirComp federated learning with power adaptation harnessing receiver noise," in *Proc. IEEE Global Commun. Conf*, 2020, pp. 1–6.
[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, 2020.
[20] S. Chen, D. Yu, Y. Zou, J. Yu, and X. Cheng, "Decentralized wireless federated learning with differential privacy," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 6273–6282, 2022.
[21] M. Kim, A. L. Swindlehurst, and D. Park, "Beamforming vector design and device selection in over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7464–7477, 2023.
[22] A. Bereyhi, A. Vagollari, S. Asaad, R. R. Müller, W. Gerstacker, and H. V. Poor, "Device scheduling in over-the-air federated learning via matching pursuit," *IEEE Trans. Signal Process.*, vol. 71, pp. 2188–2203, 2023.
[23] Y. Zhang, M. Duan, D. Liu, L. Li, A. Ren, X. Chen, Y. Tan, and C. Wang, "Csafl: A clustered semi-asynchronous federated learning framework," in *Proc. Int. Joint Conf. Neural Netw.* IEEE, 2021, pp. 1–10.
[24] Y. Zhang, D. Liu, M. Duan, L. Li, X. Chen, A. Ren, Y. Tan, and C. Wang, "Fedmds: An efficient model discrepancy-aware semi-asynchronous clustered federated learning framework," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 3, pp. 1007–1019, 2023.
[25] C. Xie, S. Koyejo, and I. Gupta, "Asynchronous federated optimization," arXiv:1903.03934, 2019.
[26] X. Yuan, X. Ma, L. Zhang, Y. Fang, and D. Wu, "Beyond class-level privacy leakage: Breaking record-level privacy in federated learning," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2555–2565, 2021.
[27] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.
[28] Y. Shao, D. Gündüz, and S. C. Liew, "Bayesian over-the-air computation," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 589–606, 2022.
[29] N. Yan, K. Wang, C. Pan, and K. K. Chai, "Device scheduling for over-the-air federated learning with differential privacy," in *Proc. IEEE Int. Conf. Commun. (ICC)*. IEEE, 2023, pp. 51–56.
[30] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized power control design for over-the-air federated edge learning," *IEEE J. Sel. Areas. Commun.*, vol. 40, no. 1, pp. 342–358, 2021.
[31] N. Yan, K. Wang, C. Pan, and K. K. Chai, "Private federated learning with misaligned power allocation via over-the-air computation," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 1994–1998, 2022.
[32] S. P. Kasiviswanathan, M. Rudelson, and A. Smith, "The power of linear reconstruction attacks," in *Proc. ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2013, pp. 1415–1433.
[33] N. T. T. Van, N. C. Luong, H. T. Nguyen, F. Shaohan, D. Niyato, and D. I. Kim, "Latency minimization in covert communication-enabled federated learning network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13 447–13 452, 2021.
[34] Y.-A. Xie, J. Kang, D. Niyato, N. T. T. Van, N. C. Luong, Z. Liu, and H. Yu, "Securing federated learning: A covert communication-based approach," *IEEE Netw.*, 2022.
[35] J. Yao and N. Ansari, "Secure federated learning by power control for internet of drones," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 4, pp. 1021–1031, 2021.
[36] A. Madi, O. Stan, A. Mayoue, A. Grivet-Sébert, C. Gouy-Pailler, and R. Sirdey, "A secure federated learning framework using homomorphic encryption and verifiable computing," in *Proc. Reconciling Data Anal.,Automat., Privacy, Secur., Big Data Challenge (RDAAPS)*. IEEE, 2021, pp. 1–8.
[37] J. Park and H. Lim, "Privacy-preserving federated learning using homomorphic encryption," *Appl. Sci.*, vol. 12, no. 2, p. 734, 2022.
[38] M. Frey, I. Bjelaković, and S. Stańczak, "Towards secure over-the-air computation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 700–705.
[39] U. Maurer and S. Wolf, "Secret key agreement over a nonauthenticated channel—parts I-III: Definitions and bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 822–831, 2003.
[40] A. Mukherjee and A. L. Swindlehurst, "Equilibrium outcomes of dynamic games in mimo channels with active eavesdroppers," in *Proc. IEEE Int. Conf. Commun. (ICC)*. IEEE, 2010, pp. 1–5.
[41] ——, "Jamming games in the MIMO wiretap channel with an active eavesdropper," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 82–91, 2012.
[42] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.
[43] X. Wei and C. Shen, "Federated learning over noisy channels: Convergence analysis and design examples," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 1253–1268, 2022.
[44] 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, 2020.
[45] Y. Shi, Y. Liu, K. Wei, L. Shen, X. Wang, and D. Tao, "Make landscape flatter in differentially private federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 24 552–24 562.
[46] L. Lyu, X. Xu, Q. Wang, and H. Yu, "Collaborative fairness in federated learning," in *Federated Learning: Privacy and Incentive*. Springer, 2020, pp. 189–204.
[47] P. Sun, H. Che, Z. Wang, Y. Wang, T. Wang, L. Wu, and H. Shao, "Pain-FL: Personalized privacy-preserving incentive for federated learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3805–3820, 2021.
[48] J. Zhang, S. Guo, Z. Qu, D. Zeng, Y. Zhan, Q. Liu, and R. Akerkar, "Adaptive federated learning on Non-IID data with resource constraint," *IEEE Trans. Comput.*, vol. 71, no. 7, pp. 1655–1667, 2021.
[49] Y. Liu, X. Zhang, Y. Zhao, Y. He, S. Yu, and K. Zhu, "Chronos: Accelerating federated learning with resource aware training volume tuning at network edges," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3889–3903, 2022.
[50] R. Chen, L. Li, K. Xue, C. Zhang, M. Pan, and Y. Fang, "Energy efficient federated learning over heterogeneous mobile devices via joint design of weight quantization and wireless transmission," *IEEE Trans. Mobile Comput.*, vol. 22, no. 12, pp. 7451–7465, 2023.
[51] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-iid data," in *Proc. Int. Conf. Learn. Represent (ICLR)*, 2020.
[52] M. M. Amiri, D. Gündüz, S. R. Kulkarni, and H. V. Poor, "Convergence of update aware device scheduling for federated learning at the wireless edge," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3643–3658, 2021.
[53] W. Mei and R. Zhang, "Joint base station and IRS deployment for enhancing network coverage: A graph-based modeling and optimization approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8200–8213, 2023.