Over-the-Air Federated Learning with Enhanced Privacy

Xiaochan Xue[a1], Moh Khalid Hasan[a1], Shucheng Yu[a1], Laxima Niure Kandel[b2], Min Song[a1]
[a]Department of Electrical and Computer Engineering, Stevens Institute of Technology, NJ 07030
[b]Electrical Engineering and Computer Science Dept, Embry-Riddle Aeronautical University, FL 32114
Email:[1]{xxue2, mhasan12, syu19, msong6}@stevens.edu, [2]{niurekal}@erau.edu

***Abstract*—Federated learning (FL) has emerged as a promising learning paradigm in which only local model parameters (gradients) are shared. Private user data never leaves the local devices thus preserving data privacy. However, recent research has shown that even when local data is never shared by a user, exchanging model parameters without protection can also leak private information. Moreover, in wireless systems, the frequent transmission of model parameters can cause tremendous bandwidth consumption and network congestion when the model is large. To address this problem, we propose a new FL framework with efficient over-the-air parameter aggregation and strong privacy protection of both user data and models. We achieve this by introducing pairwise cancellable random artificial noises (PCR-ANs) on end devices. As compared to existing over-the-air computation (AirComp) based FL schemes, our design provides stronger privacy protection. We analytically show the secrecy capacity and the convergence rate of the proposed wireless FL aggregation algorithm.**

***Index Terms*—Over-the-air computation (AirComp), wireless multiple-access channel, federated learning**

I. Introduction

In machine learning, especially deep learning, large-scale collection of sensitive data entails both high bandwidth consumption and privacy-related risks. To mitigate these limitations and leverage the power of proliferating edge devices, federated learning (FL) [1] has emerged as a promising new learning paradigm. In FL each edge device trains a local ML model using its private data and uploads only model parameters to a central server. The server then aggregates local models received from the distributed edge devices to obtain a global model that is expected to outperform the individual local models. While FL is promising as compared to centralized learning, the frequent transmission of model parameters can still cause significant bandwidth consumption and latency in wireless and mobile systems. Moreover, recent research has discovered vulnerabilities of FL under membership inference attacks [2, 3, 4, 5]. Specifically, it has been demonstrated that models implicitly memorize certain details about the underlying training data and can inadvertently reveal sensitive information to attackers. To strike a balance between efficiency and privacy in FL, existing research has resorted to various techniques, including *Secure Aggregation* (SA) [6] and *Differential Privacy* (DP) [7]. The former obfuscates parameters to the aggregator but needs pairwise key exchange which incurs non-trivial communication costs in edge computing environments. The latter on the other hand injects random noises into local training data so that it is computationally indistinguishable from that of other individuals. For FL, local differential privacy (LDP) which is a mode of DP, is more suitable because of its distributed nature and users can add noises to model parameters locally before disclosing them to the *untrusted model aggregator*. While LDP has the advantage of lower computational and communication overheads, it poses its own challenges. Specifically, an LDP model needs to introduce noises at a significantly higher level than what is required in a DP model. Moreover, since each user perturbs its parameters individually, the aggregated variance highly depends on the number of participating users [8].

Recently, the feasibility of over-the-air computation (AirComp) [9] coupled with LDP is being explored within the context of FL, to overcome communication bottlenecks and provide additional protection to local model privacy. The AirComp-based approach exploits the broadcast and the natural superposition property of wireless multiple access channels (MAC) for fast, free, and more efficient global model aggregation. The key idea is the simultaneous synchronized transmission of linear-analog modulated local gradients. With appropriate pre-channel coefficient equalization, superposed RF signals over the air can be demodulated as the additive result at the receiver without actually performing the addition operation. Together with local pre-processing, complex functions such as scalar products can be implemented via AirComp, which saves both local computation and latency for wireless devices. Despite of the challenges, existing research [10, 11, 12, 13] has demonstrated promising progresses both theoretically and through practical implementation.

Along this direction, this paper aims to explore the full potential of AirComp-based FL by providing enhanced privacy protection. Specifically, while protecting model privacy, existing research [10, 11, 12, 13, 14] mainly relies on *obfuscation via aggregation* (OVA) of parameters from multiple users with local adjustment of signal to noise ratio (SNR). Although this approach protects model privacy against the parameter aggregation server (PAS), it is vulnerable to stronger attacks in which an external attacker is equipped with a directional antenna to overhear RF signals from individual transmitters and bypass the aggregation. Moreover, the OVA approach requires a higher noise level for model privacy when the number of users is less, which adversely impacts the global model quality. To address this limitation, in this paper we introduce pairwise cancellable random artificial noises (PCR-ANs) to obfuscate individual private model parameters. By adjusting the PCR-AN level, our design is able to thwart external eavesdroppers

equipped with directional antennas. Because the PCR-ANs are pairwise cancellable, only residue noises remain in the aggregated model[1]. Our design can be considered as a novel integration of SA and DP at the physical layer. Analytical results provide both secrecy capacity and the FL convergence rate of our design. Our contributions can be summarized as follows:

- We introduce a new AirComp-based privacy-preserving FL scheme considering the presence of powerful eavesdroppers. The pairwise cancellable random artificial noise (PCR-AN) design leverages the properties of both secure aggregation and differential privacy and provides a better trade-off between privacy and model utility as compared to the state-of-the-art.
- We theoretically analyze the feasibility of the PCR-AN design and formulate the secrecy capacity of our proposed privacy-preserving FL scheme in the presence of powerful eavesdroppers. We analytically show the convergence rate of our proposed FL scheme.
- With the adjustable power parameters of artificial noises, our design is also able to preserve model privacy at the PAS based on the differential privacy constraints.

The rest of the paper is structured as follows. Section II describes the system and threat model for FL. Section III presents our design and elaborates on PCR-AN-aided privacy-preserving FL. Section IV presents an analytical privacy analysis, the secrecy capacity, and the convergence rate of our proposed scheme. Section V presents the simulation, and evaluation results, and Section VI concludes the paper.

II. System Model and Assumptions

A. System Model and Federated Learning

We consider a wireless federated learning system consisting of a parameter aggregation server (PAS) and multiple end users. The PAS is a single-antenna receiver and aggregates the distributed local model parameters from total K ($K = |\mathcal{K}|$) users, where $\mathcal{K} = \{1, 2, 3, ..., 2i\}, i \in \mathbb{Z}^+$. Each user participant k ($k \in \mathcal{K}$) is a spatially distributed single-antenna device and without loss of generality, it is assumed that all devices are identical to each other and within one-hop distance to PAS. Each user k has a private local data set $\mathcal{D}_k$ and we assume that all users have the same data size of $|\mathcal{D}_k|$. Data points are denoted as $\mathcal{D}_k = \{(\boldsymbol{u}_j^{(k)}, v_j^{(k)}) | j \in \mathcal{D}_k\}$, where $\boldsymbol{u}_j^{(k)} \in \mathbb{R}^d$ is the j-th data point and $v_j^{(k)}$ is the corresponding label for each data point. Each user individually trains an ML model using their private data $\mathcal{D}_k$ and then uploads a d-dimensional model parameter vector $\mathbf{w}$ wirelessly to the PAS. For efficiency, the participants use Gaussian multiple access channels (MAC) to simultaneously transmit their respective parameters. PAS receives aggregated parameters because of the over-the-air superposition of wireless signals. This process is called the over-the-air computation (AirComp) [9] which can implement complex functions if users are well synchronized and equalized. The global aggregated model is obtained by minimizing the loss function $F(\mathbf{w})$ as follows:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} F(\mathbf{w}) \triangleq \frac{1}{|\mathcal{D}|} \sum_{k=1}^{K} \sum_{j=1}^{\mathcal{D}_k} f_k((\boldsymbol{u}_j^{(k)}, v_j^{(k)}); \mathbf{w}) \quad (1)$$

where $\mathcal{D} = \bigcup_{k=1}^{K} \mathcal{D}_k$ denotes the entire dataset used for training, and $f_k(\bullet)$ is the loss function for user k. The minimization of $F(\mathbf{w})$ in eq. (1) is carried out iteratively through a gradient descent (GD) algorithm. At iteration t, the PAS broadcasts the global model parameter vector $\mathbf{w}_t$ and each user then updates its local gradient vector over the local dataset $\mathcal{D}_k$ as:

$$\boldsymbol{g}_k(\mathbf{w}_t) = \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{\mathcal{D}_k} \nabla f_k((\boldsymbol{u}_j^{(k)}, v_j^{(k)}); \mathbf{w}) \quad (2)$$

Next, the locally computed gradient is sent back to the PAS and the global model $\mathbf{w}_t$ is updated according to:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \frac{1}{K} \left(\sum_{k=1}^{K} \boldsymbol{g}_k(\mathbf{w}_t) \right) \quad (3)$$

$\mathbf{w}_{t+1}$ is the updated global model and η_t is the learning rate of the GD algorithm at iteration t. The PAS will broadcast $\mathbf{w}_{t+1}$ and the above process continues until convergence with total T iterations.

B. Threat Model

Our threat model considers honest-but-curious attackers, i.e., we assume the attacker passively eavesdrops on exchanged messages (e.g., gradients) between the client and the PAS. However, the attacker does not interfere with the training process. For instance, due to the broadcast nature of the wireless medium, the eavesdropper easily wiretaps the local parameter-modulated transmitted signal by pointing a directional antenna toward the transmitting victim device. After the adversary has the wiretapped model at its disposal, it can violate privacy by recovering the underlying sensitive data on which the model was trained by launching sophisticated model inversion attacks or may gain leaked private information when the wiretapped model is used for inference. We show that our design defends against such passive attackers and achieves both data and model privacy. More sophisticated active attackers will be explored in our future work.

III. Our Design

A. Preliminaries of AirComp for Ultrafast Aggregation

AirComp shows great promise to support ultrafast aggregation of local FL model parameters from distributed mobile users. The principle idea of AirComp is to exploit the analog-wave superposition property of wireless multiple access channels (MAC). As illustrated in Fig. 1, we consider a simplified baseline single-antenna AirComp system with non-zero receiver noise and unequal channel coefficients. Let s_k denote the analog modulated local model parameters symbols

[1]In a parallel work Liao et al. [15] also proposed a secure FL scheme based on pairwise cancellable noises.

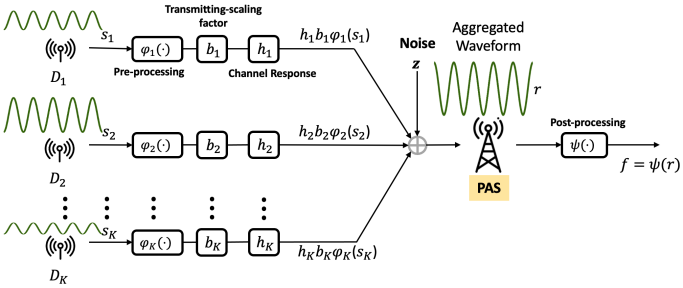

Figure 1: The data aggregation over the MAC via over-the-air computation.

calculated by client k. The aggregated function at the PAS then can be written as:

$$f = \psi(r) \tag{4}$$

$$r = \sum_{k=1}^{K} h_k b_k \varphi_k(s_k) + z_k \tag{5}$$

where in eq. (4), r is the received superimposed signal and $\psi(\bullet)$ is the post-processing function at the PAS. $\varphi_k(\bullet)$ is the pre-processing function at each transmitting device. The selection of pre-processing and post-processing functions depends on the desired function f. The variable h_k is the channel coefficient, b_k is the transmitter scaling factor to achieve channel inversion (CI) and z_k is the Additive White Gaussian Noise (AWGN) at user k. It is assumed that the channel is time-invariant and the transmitting mobile devices including the PAS have the channel state information (CSI) to achieve channel inversion.

B. Pairwise Cancellable Random Artificial Noise (PCR-AN)

We present a general gradient aggregation scheme for wireless FL based on AirComp, as shown in Fig. 2. Each user k synchronously transmits a linear combination of local gradients and pairwise cancellable random artificial noise (PCR-AN) over a wireless channel for total T training iterations. At each iteration t, all participating K users transmit their local computed gradient vector $\boldsymbol{s}_{k,t} := \boldsymbol{g}_k(\mathbf{w}_t) \in \mathbb{R}^d$ masked with PCR-AN to preserve modal privacy. More specifically, the transmitted signal of user k with added artificial noise $\boldsymbol{n}_{k,t}$ at iteration t is given as:

$$\boldsymbol{x}_{k,t} = b_k \varphi_k \left(\boldsymbol{s}_{k,t} + \boldsymbol{n}_{k,t} \right) + \boldsymbol{z}_{k,t} \tag{6}$$

The terms used in eq. (6) are explained below:

- $\boldsymbol{n}_{k,t} \in \mathbb{R}^d$ is the PCR-AN (Gaussian noise) with mean $\mu_{k,t}$, and variance $\sigma^2_{k,t}$ ($\boldsymbol{n}_{k,t} \sim \mathcal{N}(\mu_{k,t}, \sigma^2_{k,t})$) to mask the gradient vector, $\boldsymbol{s}_{k,t}$. Two pairwise devices secretly share the mean and variance value, then add artificial noise with opposite mean values to the gradients. For example, users a and b pre-share a secret (μ, σ^2) and then this secret will be used by user a to add noise of $\mathcal{N}(+\mu, \sigma^2_a)$ and noise of $\mathcal{N}(-\mu, \sigma^2_b)$ is added by user b.
- $\boldsymbol{z}_{k,t} \in \mathbb{R}^d$ is the additive zero-mean unit-variance Gaussian noise over the wireless channel ($\mathcal{N}(0, \sigma^2_z)$, $\sigma^2_z = 1$).
- $\varphi_k(\bullet)$ is the pre-processing at each user. Since, the desired function at the PAS in the context of FL is the arithmetic mean, $\varphi_k(\bullet) = 1$.

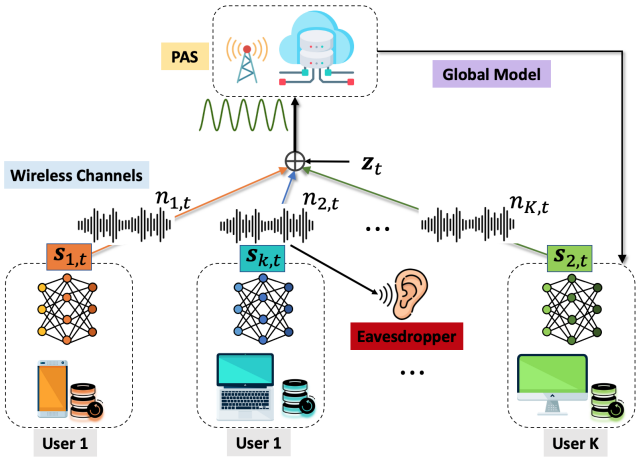

Figure 2: Federated learning with artificial noises based on AirComp in the presence of an eavesdropper.

- b_k is the Tx-scaling factor for each user to ensure the analog modulated waves add constructively in the air and a non-zero signal is received. Typically, the signal is multiplied by $e^{-j\phi_k}$ for local phase correction.

Also, in the above eq. (6), it is assumed that the gradient vectors have a bounded norm to bound the maximum changing rate, i.e., $\|\boldsymbol{s}_{k,t}\|_2 \leq L_s, \forall k$. Let $\alpha_k \in [0, 1]$ denote the coefficient of power dedicated to the gradient vector $\boldsymbol{s}_{k,t}$. The remaining power of $\beta_k \in [0, 1 - \alpha_k]$ ($\beta_k \geq \alpha_k$) is dedicated to the artificial noise to satisfy the maximum transmit power constraint needs of P_k. Using eq. (3) to (6), the received signal at the PAS can be written as:

$$\boldsymbol{r}_t = \sum_{k=1}^{K} |h_k| \left(\frac{\sqrt{\alpha_k P_k}}{L_s} \boldsymbol{s}_{k,t} + \sqrt{\beta_k P_k} \boldsymbol{n}_{k,t} \right) + \boldsymbol{z}_{k,t} \tag{7}$$

To represent eq. (7) in compact form, we introduce m as follows:

$$m := |h_k| \frac{\sqrt{\alpha_k P_k}}{L_s}, \forall k \tag{8}$$

Herein, m is a constant, and the upper bound of m can be computed by utilizing $\alpha_k \leq 1, \forall k$ in eq. (8). To maximize the power of aligned gradients, m is chosen as $m = \frac{\sqrt{\min\limits_q |h_q|^2 P_q}}{L_s}$ resulting in α_k as follows:

$$\alpha_k = \frac{\min\limits_q |h_q|^2 P_q}{|h_k|^2 P_k}$$

q is the user with worst effective SNR. Thus, above choice of α_k shows that the alignment of gradients is effectively limited by the user q with the worst effective SNR. Substituting m in eq. (7), we get the compact representation as follows:

$$\boldsymbol{r}_t = m \sum_{k=1}^{K} \boldsymbol{s}_{k,t} + \sum_{k=1}^{K} |h_k| \sqrt{\beta_k P_k} \boldsymbol{n}_{k,t} + \boldsymbol{z}_t \tag{9}$$

As seen in eq. (4), the PAS performs post-processing on received signal $\boldsymbol{r}_t$ and for the aggregation scheme, the post-

processing function is $\psi(\bullet) = \frac{1}{mK}$. Thus, the estimated function at PAS is as follows:

$$\begin{aligned}\hat{\boldsymbol{s}}_t &= \frac{1}{mK}(\boldsymbol{r}_t) \\ &= \underbrace{\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{s}_{k,t}}_{\nabla F(\mathrm{w}_t)} + \underbrace{\frac{1}{mK}\sum_{k=1}^{K}|h_k|\sqrt{\beta_k P_k}\boldsymbol{n}_{k,t} + \frac{1}{mK}\boldsymbol{z}_t}_{A_t} \end{aligned} \quad (10)$$

where $A_t + \frac{1}{mK}\boldsymbol{z}_t$ is the effective noise at the PAS. Since the pairwise devices add artificial noise of opposite mean, the summed artificial noise and channel noise will have a mean of 0 and variance of $\frac{1}{mK}\sum_{k=1}^{K}|h_k|\sqrt{\beta_k P_k}\boldsymbol{n}_{k,t} + \frac{1}{mK}\boldsymbol{z}_t$. Therefore, the PAS receives an unbiased estimate of the average gradient $\nabla F(\mathbf{w}_t)$.

IV. Analysis

In this section, we first evaluate the privacy protection provided when local wireless devices participating in the same learning task obfuscate local parameters through PCR-ANs. We show that the additive artificial noise protects individual users' privacy without interfering with the global model aggregation at PAS. Next, we discuss the secrecy capacity of our design in the presence of an eavesdropper who is listening to the user's communication with the PAS. Lastly, we prove the proposed FL scheme is convergent and show the optimization of convergence, which can also meet the differential privacy requirement to preserve privacy at PAS.

A. PCR-AN Aided Privacy

As mentioned in Section III, we allocate higher power to PCR-ANs such that SNR is low and the sensitive data is below the noise floor. This means any malicious device eavesdropping over the air can only acquire noise instead of sensitive data. However, in prior literature, low SNR would mean difficulty reconstructing the original data at PAS. Herein, we expatiate the feasibility of our design despite low SNR; we present a detailed analysis showing the added PCR-AN will not interfere with the reconstruction at PAS.

Let i represent the i-th pair of wireless devices $(+i, -i)$, where $+i \in \frac{+\mathcal{K}}{2}$ denotes the device adding a positive mean value of artificial noise, and $-i \in \frac{-\mathcal{K}}{2}$ denotes the device adding a negative mean value of artificial noise. Note, $\left(\frac{+\mathcal{K}}{2}\right) \cup \left(\frac{-\mathcal{K}}{2}\right) = \mathcal{K}$ and $\left(\frac{+\mathcal{K}}{2}\right) \cap \left(\frac{-\mathcal{K}}{2}\right) = 0$. The mean values of added artificial noises at user $+i$ and user $-i$ are also pairwise, i.e., user $+i$ and $-i$ adds $\boldsymbol{n}_{+i,t} = \mathcal{N}(\mu_{+i,t}, \sigma^2_{+i,t})$ and $\boldsymbol{n}_{-i,t} = \mathcal{N}(\mu_{-i,t}, \sigma^2_{-i,t})$, respectively. The PCR-ANs are randomly selected by users to mask the uploading gradient vector. Thus, summed PCR-ANs, denoted as A_t in eq. (10) can be written as:

$$A_t := \frac{1}{mK}\Big(\sum_{i=1}^{K/2}|h_i|\sqrt{\beta_i P_i}\boldsymbol{n}_{i,t} + \sum_{i=-1}^{-K/2}|h_i|\sqrt{\beta_i P_i}\boldsymbol{n}_{i,t}\Big) \quad (11)$$

$$A_t := \frac{1}{mK}\Big(\sum_{i=1}^{K/2}|h_i|\sqrt{\beta_i P_i}\Big)\underbrace{\Big(\sum_{i=1}^{K/2}\boldsymbol{n}_{i,t} + \sum_{i=-1}^{-K/2}\boldsymbol{n}_{i,t}\Big)}_{\text{Cancellable Noise (CN)}} \quad (12)$$

$$CN := \sum_{i=1}^{K/2}\mathcal{N}(\mu_{+i,t},\, \sigma^2_{+i,t}) + \mathcal{N}(\mu_{-i,t},\, \sigma^2_{-i,t}) \quad (13)$$

$$CN := \sum_{i=1}^{K/2}\mathcal{N}\left(0,\, (\sigma^2_{+i,t} + \sigma^2_{-i,t})\right) \quad (14)$$

Above eq. (13) to eq. (14) is based on the the property of PCR-ANs i.e., $(\mu_{+i,t} + \mu_{-i,t}) = 0$. Therefore, aggregated PCR-ANs at PAS will follow the distribution $\mathcal{N}(0, \sigma_A^2)$, where $\sigma_A^2 = \sum_{k=1}^{K/2}(\sigma^2_{+k,t} + \sigma^2_{-k,t})$. The aggregated variance σ_A^2 is bounded by the *Central Limit Theorem* (CLT). The uploading gradient for each user includes numerous parameters, which indicates the convergence in aggregated variance σ_A^2 from **Corollary 1**.

Corollary 1. *All added artificial noises are independent but not identically distributed. The μ_k and σ_k^2 for each user k satisfy the **Lyapunov's Condition**. Therefore, according to **Lyapunov's Central Limit Theorem**, the distribution of aggregated variances σ_A^2 of all artificial noises is convergent.*

The proof of Lyapunov's central limit theorem is out of the scope of this paper and interested readers in the proof and Lyapunov's condition are referred to [16, 17]. The high-power PCR-AN added to each user with different distributions will not interfere with the PAS to reconstruct the aggregation of locally trained model signals. The estimated function from eq. (10) can be written as:

$$\hat{\boldsymbol{s}}_t = \underbrace{\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{s}_{k,t}}_{\nabla F(\mathrm{w}_T)} + \underbrace{\frac{1}{mK}\left(\sum_{k=1}^{K}|h_k|\sqrt{\beta_k P_k}\boldsymbol{n}_{k,t} + \boldsymbol{z}_t\right)}_{\boldsymbol{z}_t'} \quad (15)$$

We denote $\frac{1}{mK}(\sum_{i=1}^{K/2}|h_i|\sqrt{\beta_i P_i})$ in eq. (12) as M, and $\boldsymbol{z}_t' \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{z}_t'})$ is the residual noise of aggregated PCR-ANs and channel noise at PAS, where $\sigma^2_{\boldsymbol{z}_t'} = M^2 \cdot {\sigma_A}^2 + {\sigma_z}^2$. As $\boldsymbol{z}_t'$ is zero mean, $\hat{\boldsymbol{s}}_t$ is an unbiased estimate of $\nabla F(\mathbf{w}_T)$.

B. Secrecy Capacity

To estimate the secrecy capacity, we select a two-user scenario. Herein, we consider two pairwise users a and b. User a transmits its signal with added PCR-AN with $\mathcal{N}(\mu_{a,t}, \sigma^2_{a,t})$, and user b transmits the parameters with added PCR-AN with $\mathcal{N}(\mu_{b,t}, \sigma^2_{b,t})$. The signals of both users are superposed in the air and the server receives the sum of the user signals. According to the above analysis, the residual noise of PCR-ANs after the aggregation at the server is $\sigma^2_{\boldsymbol{z}_t'}$. The signal-to-noise ratio of the received sum signal (SNR_s) at the PAS is given by

$$SNR_s = \frac{\frac{\sqrt{\alpha_a P_a}}{L_s}|h_a|^2}{\sigma^2_{\boldsymbol{z}_t'}} \quad (16)$$

The capacity at the server for user a can be represented as

$$\begin{aligned} C_s &= \log_2(1 + SNR_s) \\ &= \log_2\left(\frac{\sqrt{\alpha_a P_a}}{L_s}|h_a|^2 + \sigma^2_{\boldsymbol{z}_t'}\right) - \log_2\left(\sigma^2_{\boldsymbol{z}_t'}\right) \end{aligned} \quad (17)$$

We assume the eavesdropper wiretaps the data of user a. As the eavesdropper receives the PCR-AN with the actual signal from user a with variance σ_a^2, the SNR at the eavesdropper is given by,

$$SNR_{ev} = \frac{\frac{\sqrt{\alpha_a P_a}}{L_s}\left|h_a^{(e)}\right|^2}{\sigma_z^2 + \sigma_a^2} \tag{18}$$

where $\left|h_a^{(e)}\right|^2$ is the channel power gain corresponding to the channel coefficient $h_a^{(e)}$. The capacity at the eavesdropper is estimated as follows:

$$C_{ev} = \log_2\left(\frac{\sqrt{\alpha_a P_a}}{L_s}\left|h_a^{(e)}\right|^2 + \sigma_z^2 + \sigma_a^2\right) - \log_2\left(\sigma_z^2 + \sigma_a^2\right) \tag{19}$$

Now, we can estimate the secrecy capacity as follows.

$$C = \left[\log_2\left(\frac{\frac{\sqrt{\alpha_a P_a}}{L_s}|h_a|^2 + \sigma_{\boldsymbol{z}_t'}^2}{\frac{\sqrt{\alpha_a P_a}}{L_s}\left|h_a^{(e)}\right|^2 + \sigma_z^2 + \sigma_a^2}\right) - \log_2\left(\frac{\sigma_{\boldsymbol{z}_t'}^2}{\sigma_z^2 + \sigma_a^2}\right)\right]^+ \tag{20}$$

where $[x]^+ = \max\{x, 0\}$. The main objective is to enhance the secrecy capacity so that the privacy of user a is improved. It can be realized from eq. (20) that there is a direct influence of σ_a^2 on C, which means that more PCR-AN at the sender decreases the capacity at the eavesdropper, in other words, increases the privacy of user a. Also, it is evident in eq. (20) that if σ_A^2 increases, C decreases.

C. Convergence Rate of Private AirComp-based FL

Theorem 1. *Suppose the loss function F is λ-strongly convex and μ-smooth with respect to $\mathbf{w}^*$ over a convex set $\mathcal{W}$, and $\mathbb{E}[\|\hat{\boldsymbol{s}}_t\|^2] \leq G^2$. Then if we pick $\eta_t = 1/\lambda_t$, the convergence rate for iteration T is*

$$\mathbb{E}\left[F(\mathbf{w}_T) - F(\mathbf{w}^*)\right] \leq \frac{2\mu}{\lambda^2 T}\left(L_s^2 + \frac{d}{m^2K^2}\left[\sum_{k=1}^{K}|h_k|^2\beta_k P_k + \sigma_z^2\right]\right) \tag{21}$$

The detailed proof of Theorem 1 is given by [18] and [14]. The convergence rate can also be maximized by optimizing the artificial noise parameter β_k, which can also meet the differential privacy requirement to preserve the privacy at PAS in [14]. The β_k can be written as:

$$\beta_k = \frac{Z_k}{|h_k|^2 P_k}, \forall k \tag{22}$$

where $Z_k = \min\left[\lambda_k, (\Psi - \sum_{p=1}^{k-1} U_p)^+\right], \forall k$, $\Psi = \max_p \frac{\min_q |h_q|^2 P_q}{\epsilon_p} \log\frac{1.25}{\delta} - \sigma_z^2$, and $U_p = |h_p|^2\beta_p P_p$. The (ϵ, δ) is the local differential privacy level.

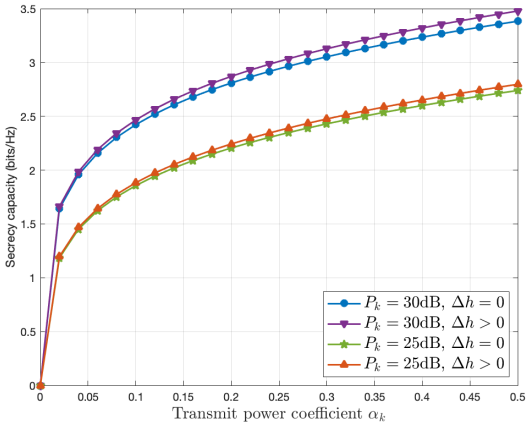

Figure 3: Secrecy capacity with respect to different transmitted signal's coefficient α_k.

V. Evaluation

In this section, we first provide simulation results of secrecy capacity to show the performance of our AirComp-based privacy-preserving FL model. The Rayleigh fading wireless channels for the simulation results are randomly generated over 10^6 realization samples in Matlab. The channel coefficients are drawn from $\mathcal{N}(0, 1)$, and the channel noise variance is set to $\sigma_z^2 = 1$. We set the variance σ_k^2 of the user k's PCR-AN to 25dB. The Lipschitz constant L_s is considered as 1. We define $\Delta h = |h_k|^2 - |h_k^{(e)}|^2$ as the capability of an eavesdropper to obtain the parameters from the victim compared with PAS. $\Delta h = 0$ means the eavesdropper has high capability as PAS, $\Delta h > 0$ means low capacity at eavesdropper, relatively. Based on the assumption of two coefficients α_k and β_k, we show the secrecy capacity with the respect to different transmit signal's coefficient $\alpha_k \in [0, 0.5]$ in Fig. 3. With the increase of transmit power coefficient α_k, the secrecy capacity increases for all scenarios. For both transmit power $P_k = 25$ dB and $P_k = 30$ dB, the secrecy capacity increases for the scenario that the eavesdropper has a worse channel gain ($\Delta h > 0$) than PAS.

We then consider the $\Delta h > 0$ scenario here, which is a general assumption in wireless communication. We set the

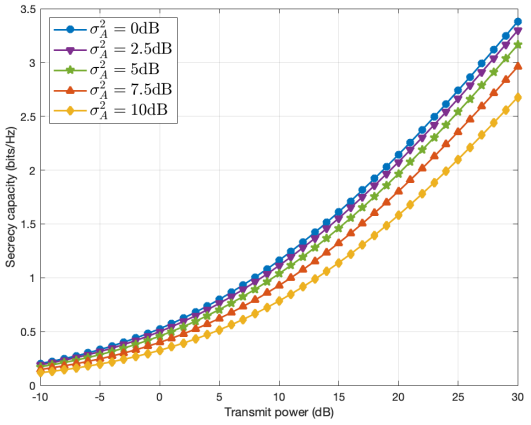

Figure 4: Secrecy capacity with respect to different transmit power.

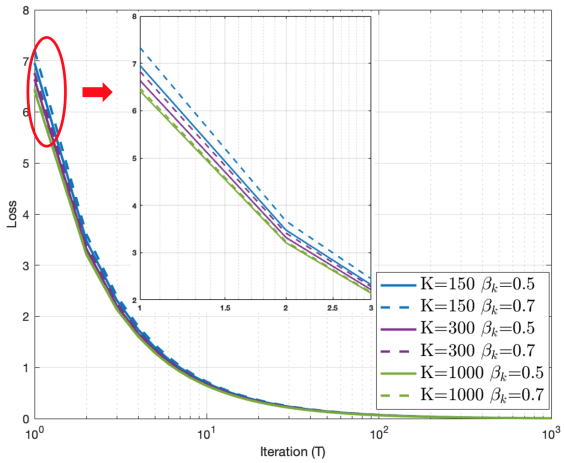

Figure 5: Convergence rate of private AirComp-based FL.

$\alpha_k = 0.5$ in the simulation. In Fig. 4, we show the impact of transmit signal power on the secrecy capacity of an individual user. For aggregated PCR-ANs variance $\sigma_A^2 = 0$ dB, which means the power of PCR-ANs is cancelled perfectly at PAS. With the convergence of aggregated PCR-ANs' variance σ_A^2, the secrecy capacity increase. The trend of secrecy capacity also increases with the increase of transmit power.

Fig. 5 shows the impact of the total number of users K and iteration T on the convergence rate based on eq. (21). For the GD algorithm, the regularization parameter λ is 10^{-3} and $T = 1000$ training iterations. We assume the transmit power $P_k = 30$ dB for each user k based on the analysis from Fig. 3 which can reach a higher secrecy capacity. We also assume the data points $d = 30$. From the enlarged detail for the beginning of the iteration, as we increase the number of users, the training loss decays with T. We also show the impact of PCR-AN's power coefficient β_k for each user k. The loss decreases with the decrease of PCR-AN's power. We compare different pair of coefficients of transmitting signal and PAC-AN, which is $\alpha_k = 0.3$, $\beta_k = 0.7$ and $\alpha_k = 0.5$, $\beta_k = 0.5$. From simulation results, the lower β_k performs a faster convergence rate. Therefore, we chose to set $\beta_k = 0.5$. This means only necessary PCR-AN power can help the FL model reach good convergence. We can easily figure out that the trend of training loss converges as the number of the iteration T increases.

VI. Conclusion

In this paper, we propose a new privacy-preserving FL framework with efficient over-the-air parameter aggregation and random pairwise cancellable artificial noises (PCR-ANs) to obfuscate individual private model parameters. We demonstrate the use of PCR-ANs by users provides strong privacy protection for both user data and models. By adjusting the PCR-AN power level, our design is able to thwart external eavesdroppers equipped with directional antennas. Also, because the PCR-ANs are pairwise cancellable, it does not cause a large error in the estimation of the global model at the aggregator. Some residual noise due to different variances remains in the aggregated model which aids in providing additional protection against malicious servers. Theoretical analysis of the secrecy capacity and convergence rate shows the feasibility of our design and the stronger privacy protection provided by the proposed FL.

Acknowledgment

This work was supported in part by the National Science Foundation under grants ECCS-1923739 and CNS-1817438.

References

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[2] Jamie Hayes, Luca Melis, George Danezis, and Emiliano De Cristofaro. Logan: Membership inference attacks against generative models. *arXiv preprint arXiv:1705.07663*, 2017.

[3] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pages 3–18. IEEE, 2017.

[4] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE symposium on security and privacy (SP)*, pages 691–706. IEEE, 2019.

[5] Wenqi Wei, Ling Liu, Margaret Loper, Ka-Ho Chow, Mehmet Emre Gursoy, Stacey Truex, and Yanzhao Wu. A framework for evaluating gradient leakage attacks in federated learning. *arXiv preprint arXiv:2004.10397*, 2020.

[6] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191, 2017.

[7] Cynthia Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

[8] Raef Bassily, Kobbi Nissim, Uri Stemmer, and Abhradeep Guha Thakurta. Practical locally private heavy hitters. *Advances in Neural Information Processing Systems*, 30, 2017.

[9] Bobak Nazer and Michael Gastpar. Computation over multiple-access channels. *IEEE Transactions on Information Theory*, 53(10):3498–3516, 2007.

[10] Mohammad Mohammadi Amiri and Deniz Gündüz. Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1432–1436, 2019. doi: 10.1109/ISIT.2019.8849334.

[11] Guangxu Zhu, Dongzhu Liu, Yuqing Du, Changsheng You, Jun Zhang, and Kaibin Huang. Toward an intelligent edge: Wireless communication meets machine learning. *IEEE Communications Magazine*, 58(1):19–25, 2020.

[12] Henrik Hellström, José Mairton B da Silva Jr, Viktoria Fodor, and Carlo Fischione. Wireless for machine learning. *arXiv preprint arXiv:2008.13492*, 2020.

[13] Guangxu Zhu, Yong Wang, and Kaibin Huang. Broadband analog aggregation for low-latency federated edge learning. *IEEE Transactions on Wireless Communications*, 19(1):491–506, 2020.

[14] Mohamed Seif, Ravi Tandon, and Ming Li. Wireless federated learning with local differential privacy. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2604–2609, 2020.

[15] Jialing Liao, Zheng Chen, and Erik G. Larsson. Over-the-air federated learning with privacy protection via correlated additive perturbations. In *2022 58th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8, 2022.

[16] Patrick. Billingsley. *Probability and measure / Patrick Billingsley.* Wiley series in probability and mathematical statistics. Probability and mathematical statistics. J. Wiley & Sons, New York, third edition. edition, 1995 - 1995. ISBN 0471007102.

[17] Alfredo Cuzzocrea, Edoardo Fadda, and Alessandro Baldo. Lyapunov central limit theorem: Theoretical properties and applications in big-data-populated smart city settings. ICCBDC '21, page 34–38, New York, NY, USA, 2021. Association for Computing Machinery.

[18] Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. *arXiv preprint arXiv:1109.5647*, 2011.