A Joint Communication and Learning Design for Secure Federated Learning with Differential Privacy

Licheng Lin*, Zhaohui Yang†, Qianqian Yang†, Mingzhe Chen*‡
*Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, 33146, USA,
†College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, 310027, China,
‡Frost Institute for Data Science and Computing, University of Miami, Coral Gables, FL, 33146, USA,
E-mails: {lxl1293, mingzhe.chen}@miami.edu, {yang_zhaohui, qianqianyang20}@zju.edu.cn

2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall) | 979-8-3315-1778-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/VTC2024-Fall63153.2024.10757930

Abstract—**In this paper, the problem of resource allocation for non-orthogonal multiple access (NOMA) enabled secure federated learning (FL) is investigated. In the considered model, a set of users participate in the FL training through transmitting their trained FL model parameters to the base stations (BSs) via NOMA techniques. To prevent data leakage, each user uses the differential privacy (DP) technique through adding Gaussian noise to its FL model parameters. The problem of minimizing overall privacy leakage of all FL participaring users is formulated as an optimization problem through jointly optimizing the connections between users and BSs, transmit power of the users, and the DP noise power. To solve the formulated non-convex optimization problem, a genetic algorithm is proposed to search for feasible solutions in which user connection matrix is taken as gene and the objective function value is taken as the fitness of solution. Simulation results show that the proposed genetic algorithm reduces privacy leakage by up to 73% compared to the conventional alternating optimization algorithm.**

Index Terms—**Federated Learning, Differential Privacy, Joint Communication and Learning Design**

I. Introduction

Centralized machine learning (ML) algorithms require to collect diverse data from edge devices for ML model training which may be hindered by privacy, security constraints [1]. To circumvent these limiting factors, federated learning (FL) [2] is proposed, which distributes the learning process across multiple distributed edge devices, thus improving data privacy of edge devices. In particular, instead of sharing raw data, edge devices in FL transmit only model parameters to a central server where ML model parameters are aggregated to form a global model. Since wireless networks have limited wireless and computing resources, it is necessary to study the optimization of these resource allocation to support the deployment of FL over wireless networks.

Currently, FL has been extensively studied in a number of current works as shown in [2]–[11]. In particular, in [3], a comprehensive overview of current FL methods, FL deployment challenges, and future directions was provided. The work in [2] designed an effective FL called FedAvg based on iterative model averaging and demonstrated its robustness experimentally. The authors in [4] introduced the FedProx framework to address data heterogeneity. In [5], a FL algorithm is designed by considering communication compression and decentralization. The authors in [6] studied how wireless transmission errors affect the FL training performance and optimized wireless resource (i.e., resource block and transmit power) allocation and the set of participating devices to reduce FL training loss. The work in [7] investigated the deployment of FL in an energy harvesting wireless network and designed a user scheduling algorithm to reduce FL training loss. The work in [8] studied a communication-efficient FL framework which enables edge devices to efficiently train and transmit model parameters, thus significantly improving FL performance and convergence speed. The authors in [9] addresses the critical issue of communication efficiency in wireless FL by proposing and analyzing quantization methods to reduce bandwidth usage while maintaining high accuracy. The work in [10] introduces a approach to FL that addresses heterogeneous quantization among clients, optimizing aggregation weights based on quantization precision to improve convergence and performance. The work in [11] presents a bandwidth allocation model for FL in wireless networks with multiple providers, addressing round length minimization under bandwidth and cost constraints. While exchanging ML model parameters in FL can improve data privacy of edge devices, current researches [12] have shown that ML model parameters transmitted by edge devices during FL training can still unintentionally leak sensitive information of participating devices. To solve this problem, differential privacy (DP) has been considered as a promising technique to prevent such privacy risks by carefully adding artificial noise into exchanged ML model parameters [13].

A number of recent works [14]–[18] have studied the use of DP to improve FL data privacy. In particular, in [14], a novel FL algorithm is proposed that applies differential privacy to preserve privacy, with precise calculation of the privacy costs within the differential privacy framework. The work in [15] studied the trade-off between convergence performance and degrees of privacy preservation, analyzed the improvement in convergence performance when the number of users increases and the existence of an optimal communication round in terms of convergence performance under a fixed privacy protection level. The authors in [16] showed that the superposition nature of the wireless channel provides bandwidth-efficient gradient aggregation and guarantees the data privacy of participating users. The authors in [17] introduced a FL framework based on multi-agent multi-armed bandit to reduce FL training delay

This work was supported by the U.S. National Science Foundation under Grants CNS-2350076 and CNS-2332834.

over wireless channels with differential privacy requirements. In [18] explores optimizing the trade-off between model performance and privacy preservation in federated edge learning systems using personalized differential privacy and proposed a method to jointly minimize global loss and privacy leakage. However, the above works on the use of DP for FL [14]–[18] ignored the joint optimization of user association and DP noise power allocation although the performance can be further enhanced. Meanwhile, none of these works [14]–[18]

The main contribution of this paper is a novel privacy preserving FL framework that uses non-orthogonal multiple access (NOMA) technique to improve FL parameter transmission efficiency and minimize user privacy leakage. Our key contributions include:

- We consider a NOMA based cellular network in which a set of users participate in the FL training through using the same time and frequency resource. To prevent data leakage, the users add Gaussian noise to the model parameters and then transmit them to their associated BSs. Our purpose is to jointly optimize the connections between the users and BSs, transmit power of users, and the DP noise power added to the FL model parameters so as to minimize the privacy leakage of all users.
- The problem of minimizing overall privacy leakage is formulated as an optimization problem with optimization variables including user association, transmission power, and DP noise power. To solve this problem, we propose a genetic algorithm to search for feasible solutions in which user connection matrix is taken as gene and the objective function value is taken as the fitness of solution.

Simulation results show that the proposed genetic algorithm reduces the total privacy leakage by up to 73% compared to the conventional alternating optimization algorithm.

The rest of this paper is organized as follows. The system model is described in Section II. The optimization problem is formulated in Section III. In Section IV, we introduce our genetic algorithm to solve the optimization problem. Simulation settings and results are introduced in Section V. Conclusions are drawn in Section VI.

II. System Model

We consider a cellular network in which a set $\mathcal{S}$ of S base stations (BSs), a set $\mathcal{U}$ of U users, and a parameter server jointly perform an FL algorithm, as shown in Fig. 1. The users applies a DP scheme to add artificial noise to their transmitted FL model parameters so as to preserve their data privacy. Next, we first introduce the FL training procedure. Then, we explain the models of FL parameter transmission.

A. Federated Learning System Model

The connection matrix between users and BSs is defined as

$$\boldsymbol{M} = \begin{pmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,U} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,U} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{S,1} & \mu_{S,2} & \cdots & \mu_{S,U} \end{pmatrix} \in \{0,1\}^{S\times U}, \quad (1)$$

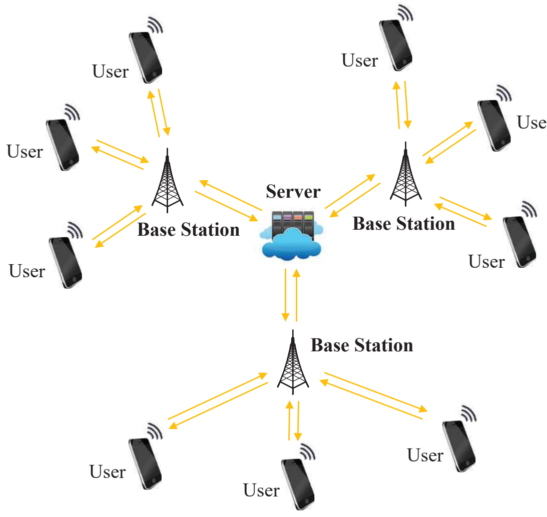

Fig. 1. DP-based FL system with multiple base stations

where $\mu_{s,i} = 1$ indicates that user $i \in \mathcal{U}$ is connected to BS $s \in \mathcal{S}$, otherwise, $\mu_{s,i} = 0$.

We assume that each user $i \in \mathcal{U}$ has its own private dataset, denoted as

$$\boldsymbol{X}_i = (\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)}, \cdots, \boldsymbol{x}_i^{(K_i)})^{\mathrm{T}} \in \mathbb{R}^{K_i \times m}, \quad (2)$$

where K_i represents the number of data samples in the dataset $\boldsymbol{X}_i$. Each $\boldsymbol{x}_i^{(k)}$ is an m-dimensional data sample vector, which can be expressed as

$$\boldsymbol{x}_i^{(k)} = (x_i^{(k)}(1), x_i^{(k)}(2), \cdots, x_i^{(k)}(m-1), y_i^{(k)}), \quad (3)$$

where the first $m-1$ elements represent the training input of user i's FL model and the last element $y_i^{(k)}$ represents the training output.

Given (1)-(3), the FL training procedure is summarized as the following five steps:

1) System Initialization. The server broadcasts the system parameters including the user connection matrix $(\mu_{s,i})_{s\in\mathcal{S},i\in\mathcal{U}}$, the transmission power used by each user $(p_i)_{i\in\mathcal{U}}$, and the standard deviation of the differential privacy noise added to user-uploaded models $(\sigma_i)_{i\in\mathcal{U}}$ to all BSs. Each BS will broadcast the received system parameters to the users that are closest to this BS. Based on the received connection matrix, the users establish the connections with BSs. Let $\mathcal{U}_s$ be the set of users that are associated with BS s. The parameters of the global FL model at each iteration t is $\boldsymbol{w}^{(t)} \in \mathbb{R}^d$. $\boldsymbol{w}^{(0)}$ is randomly initialized by the server.
2) The server transmits the initialized global model $\boldsymbol{w}^{(0)}$ to all BSs. Then, each BS s broadcasts this model to all its connected users $i \in \mathcal{U}_s$.
3) Local model update with DP strategy. Each user uses its private data set $\boldsymbol{X}_i$ and the model parameter vector $\boldsymbol{w}^{(t)}$ to calculate its local loss, denoted as $l(\boldsymbol{w}^{(t)}, \boldsymbol{x}_i^{(k)})$. Then,

users can calculate the gradient of the loss function with respect to the parameter vector $\boldsymbol{w}^{(t)}$ as follows:

$$q_i^{(t)}(\boldsymbol{X}_i) = \frac{1}{K_i}\sum_{k=1}^{K_i} \nabla l(\boldsymbol{w}^{(t)}, \boldsymbol{x}_i^{(k)}). \tag{4}$$

Following the operations in [12], users add Gaussian noise $\boldsymbol{n}_i^{(t)} \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \boldsymbol{I}_d)$ to the gradient of the loss function, as follows:

$$\mathfrak{M}_i^{(t)}(\boldsymbol{X}_i) = q_i^{(t)}(\boldsymbol{X}_i) + \boldsymbol{n}_i^{(t)}. \tag{5}$$

We assume that $\boldsymbol{n}_i^{(t)}$ is a independent valuable that is not related to the differential privacy noise in other rounds. Therefore, our considered differential privacy strategy satisfies the zero-concentrated differential privacy (zCDP) [19]. Let σ_i denote the standard deviation of the differential privacy noise added by user $i \in \mathcal{U}$. We also assume that the upper bound of the local loss function is L [13]. Hence, the privacy leakage of user i can be defined under the ρ-zCDP as follows:

$$\rho = 2\Big(\frac{L}{K_i \sigma_i}\Big)^2. \tag{6}$$

where ρ represents the measure of the privacy leakage for user i's private data during one round of FL iteration. After adding noise to the gradient, user i utilizes this gradient to locally update its model parameters, which is given by

$$\boldsymbol{w}_i^{(t+1)} = \boldsymbol{w}_i^{(t)} - \alpha \mathfrak{M}_i^{(t)}(\boldsymbol{X}_i), \tag{7}$$

where $\alpha > 0$ is the learning rate.

4) Model aggregations. Each user $i \in \mathcal{U}_s$ transmits its model $\boldsymbol{w}_i^{(t+1)}$ to its connected BS s. The BS then aggregates all the received models as follows:

$$\boldsymbol{w}_s^{(t+1)} = \frac{1}{\sum_{i\in\mathcal{U}} K_i \mu_{s,i}} \sum_{i\in\mathcal{U}} K_i \mu_{s,i} \boldsymbol{w}_i^{(t+1)}. \tag{8}$$

Finally, all BSs send their aggregated models to the server, which then aggregates them as

$$\boldsymbol{w}^{(t+1)} = \frac{1}{K_a}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i \mu_{s,i} \boldsymbol{w}_s^{(t+1)}, \tag{9}$$

where

$$K_a = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i \mu_{s,i}, \tag{10}$$

is the total number of data samples of all participating users. The model aggregated by the server will be used as the new global model for the next iteration and will be sent to all BSs at the next iteration.

5) Repeat steps 2)-4) for a pre-defined number of iterations (e.g., T iterations) until convergence.

The entire training process of the considered FL is summarized in Algorithm 1.

Algorithm 1 Multiple-BS DP-based FL

```
The server broadcasts $\mu_{s,i}, p_i, \sigma_i$ to all BSs and users.
Connections between users and BSs are determined according to the user connection index $\mu_{s,i}$.
Each BS $s$ determines its associated users, represented by a set $\mathcal{U}_s$.
for $t = 0 : T$ do
    The server transmits global model $\boldsymbol{w}^{(t)}$ to all BSs.
    for BS $s \in \mathcal{S}$ in parallel do
        BS $s$ transmits $\boldsymbol{w}^{(t)}$ to all its connected users.
        for user $i \in \mathcal{U}_s$ in parallel do
            User $i$ updates its local model according to (7).
            User $i$ sends $\boldsymbol{w}_i^{(t+1)}$ back to BS $s$.
        end for
        BS $s$ aggregates received models according to (8).
        BS $s$ sends $\boldsymbol{w}_s^{(t+1)}$ to the server.
    end for
    The server aggregates received models according to (9).
end for
```

According to [13], the power of the DP noise has a significant impact on the convergence of the considered FL algorithm since DP noise will introduce errors to the global models. To guarantee the FL convergence, the DP noise parameter σ_i must satisfy the following constraint:

$$\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i \sigma_i^2 \mu_{s,i} \le V_{\max} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i \mu_{s,i}, \tag{11}$$

where $V_{\max}$ is an upper bound on the total differential privacy noise error.

B. Communication System Model

Let p_i be the transmit power used by user $i \in \mathcal{U}$ for FL model parameter transmission, and $P_{\max}$ be the maximum transmit power of a user. We also assume that all users use the same communication frequency band. Let $I_s(\tilde{i})$ denote the interference signal power received by BS s from user $\tilde{i}$. $I_s(\tilde{i})$ can be expressed as

$$I_s(\tilde{i}) = \sum_{\tilde{s}\in\mathcal{S}} h_{s,\tilde{i}} \mu_{\tilde{s},\tilde{i}} p_{\tilde{i}}. \tag{12}$$

Hence, the total interference power received by BS s when receiving signals from user i can be written as $\sum_{\tilde{i}\in\mathcal{U}\setminus\{i\}} I_s(\tilde{i})$. Then, the channel capacity between user i and BS s is

$$c_{s,i}^{U} = \mu_{s,i} B \log(1 + \frac{p_i h_{s,i}}{\sum_{\tilde{i}\in\mathcal{U}\setminus\{i\}} I_s(\tilde{i}) + B N_0}), \tag{13}$$

where B is the channel bandwidth and N_0 is the power spectral density of thermal noise, $h_{s,i}$ is the channel gain, which can be obtained by calculating the path loss between user i and BS s as follows:

$$h_{s,i} = l^2 (\frac{c}{4\pi f})^2 (\frac{1}{d_{s,i}})^3, \tag{14}$$

where l follows a Rayleigh distribution with unit scale parameter, c is the speed of light, f is the center frequency, and $d_{s,i}$ is the distance between user i and BS s.

C. Non-Orthogonal Multiple Access Strategy

To enhance spectral efficiency, non-orthogonal multiple access (NOMA) technique is used for FL model parameter transmission between users and BSs. By utilizing superposition coding at the user (i.e., transmitter) and successive interference cancellation (SIC) at the BS (i.e., receiver), the users that are associated with the same BS can transmit FL model parameters concurrently. Through sharing their bandwidth, each user can obtain a larger channel capacity. Hence, the channel capacity between user i and BS s can be obtained as follows:

$$c^{U}_{s,i}=\mu_{s,i}\sum_{j\in\mathcal{U}}\mu_{s,j}B\log\left(1+\frac{p_ih_{s,i}}{\sum_{\tilde{\imath}\in\mathcal{U}\setminus\{i\}}I_s(\tilde{\imath})+C_{s,i}}\right), \tag{15}$$

where $I_s(\tilde{\imath}) = \sum_{\tilde{s}\in\mathcal{S}\setminus\{s\}}\mu_{\tilde{s},i}p_ih_{s,\tilde{\imath}}$ is the interference from the users that are associated with other BSs, and $C_{s,i} = BN_0 + I_{\text{NOMA}}(s,i)$ with $I_{\text{NOMA}}(s,i) = \sum_{\{k\in\mathcal{U}|h_{s,k}<h_{s,i}\}}\mu_{s,k}p_kh_{s,k}$ being the interference caused by other users connected to BS s. Here, the signals from the users that are associated with one BS are decoded in a descending order of their channel gains. When decoding a signal with a stronger channel gain, the signals with weaker channel gains are treated as interference.

III. Problem Formulation

Given the defined system model, next, we introduce an optimization problem that aims to minimize the total privacy leakage of all users via jointly optimizing the user-BS connection matrix $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$, user transmit power $(p_i)_{i\in\mathcal{U}}$, and the standard deviation of differential privacy noise adopted by each user $(\sigma_i)_{i\in\mathcal{U}}$. The studied problem is formulated as

$$\min_{\sigma_i,\mu_{s,i},p_i}\ \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}}\left(\frac{1}{K_i\sigma_i}\right)^2\mu_{s,i}, \tag{16}$$

$$\text{s.t.}\quad \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}}K_i\sigma_i^2\mu_{s,i}\leqslant V_{\max}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}}K_i\mu_{s,i}, \tag{16a}$$

$$K_i\sigma_i\geqslant N_{\min}\mu_{s,i},\ \forall s\in\mathcal{S},i\in\mathcal{U}, \tag{16b}$$

$$0\leqslant p_i\leqslant P_{\max},\ \forall s\in\mathcal{S},i\in\mathcal{U}, \tag{16c}$$

$$c^{U}_{s,i}\geqslant R_{\min}\mu_{s,i},\ \forall s\in\mathcal{S},i\in\mathcal{U}, \tag{16d}$$

$$\mu_{s,i}\in\{0,1\},\ \forall s\in\mathcal{S},i\in\mathcal{U}, \tag{16e}$$

$$\sum_{s\in S}\mu_{s,i}\leqslant 1,\ \forall i\in\mathcal{U}, \tag{16f}$$

$$\sum_{i\in\mathcal{U}}\mu_{s,i}\leqslant N_1,\ \forall s\in\mathcal{S}, \tag{16g}$$

$$\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}}\mu_{s,i}\geqslant N_2, \tag{16h}$$

where $N_{\min}$ is a lower bound of DP noise for all users, N_1 represents the maximum user connection capacity of a BS, N_2 is the minimum number of users that FL requires to converge. Constraint (16a) limits the overall differential privacy noise error to a predetermined value $V_{\max}$ so as to ensure the FL convergence [13]. Constraint (16b) sets a lower bound $N_{\min}$ on the differential privacy noise added by each user so as to prevent the privacy leakage of each user. Constraint (16c) indicates the maximum transmit power of each user. Constraint (16d) is the minimum channel capacity $R_{\min}$ required by each user to send FL model parameters. Here, $R_{\min}$ also affects the optimization of the connection matrix since a user will connect to the BS that can provide a data rate that is larger than $R_{\min}$. Constraints (16e) and (16f) imply that each user can only connect to one BS at a given time slot. Constraint (16g) limits the number of users that each BS can connect to. Constraint (16h) shows the minimum number of participating users that are required for FL training. Since problem (16) is a non-convex mixed integer optimization problem, which is generally hard to solve, we propose a genetic algorithm to solve this problem.

IV. Solution

In this section, we introduce the use of a genetic algorithm to solve the optimization problem in (16) since both $(p_i)_{i\in\mathcal{U}}$ and $(\sigma_i)_{i\in\mathcal{U}}$ depend on $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$, but $(p_i)_{i\in\mathcal{U}}$ and $(\sigma_i)_{i\in\mathcal{U}}$ are mutually independent. In particular, we will use the value of the objective function in (16) as the fitness criterion for a heuristic search of $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$. Next, we first explain the process of obtaining the optimal transmit power and DP noise under a given $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$. Then we introduce the genetic algorithm for searching $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$.

A. Transmit Power Optimization

Given the user connection index $\mu_{s,i}$, each user i's transmit power that satisfies the constraint (16d) is

$$\sum_{j\in\mathcal{U}}\mu_{s,j}B\log(1+\frac{p_ih_{s,i}}{\sum_{\tilde{\imath}\in\mathcal{U}\setminus\{i\}}I_s(\tilde{\imath})+BN_0})\geqslant R_{\min} \\ ,\forall(s,i)\in\{(s,i)|\mu_{s,i}=1\}, \tag{17}$$

which can be rewritten as

$$p_i\ \geqslant\frac{\left(\sum_{\tilde{\imath}\in\mathcal{U}\setminus\{i\}}I_s(\tilde{\imath})+BN_0+I_{\text{NOMA}}(s,i)\right)}{h_{s,i}} \\ \times\left(\exp\left(\frac{R_{\min}}{\sum_{j\in\mathcal{U}}\mu_{s,j}B}\right)-1\right),\forall(s,i)\in\{(s,i)|\mu_{s,i}=1\}. \tag{18}$$

From (18), we see that the transmit power of each user i depends on the transmit power of other users. To solve this problem, we can directly use the algorithm in [20]. To use the algorithm in [20], we assume

$$I_i\left(\boldsymbol{p}\right)=\left(\sum_{\tilde{\imath}\in\mathcal{U}\setminus\{i\}}I_s(\tilde{\imath})(\boldsymbol{p}(t))+BN_0+I_{\text{NOMA}}(s,i)(\boldsymbol{p}(t))\right) \\ \times\left(\exp\left(\frac{R_{\min}}{\sum_{j\in\mathcal{U}}\mu_{s,j}B}\right)-1\right)\Big/h_{s,i}. \tag{19}$$

Given $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$ and (19), the optimal transmit power can be iteratively calculated as follows:

$$p_i(t+1) = \begin{cases} I_i\left(\boldsymbol{p}(t+1)\right), & \forall (s,i) \in \{(s,i)|\mu_{s,i}=1\}. \\ 0 & , \forall (s,i) \in \{(s,i)|\mu_{s,i}=0\}. \end{cases} \tag{20}$$

When $|p_i(t+1) - p_i(t)| \leq \epsilon, \forall i \in \mathcal{U}$, with ϵ being a preset threshold, the optimal transmit power p_i^* of each user is obtained. If p_i^* cannot meet constraint (16c), we will discard this $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$ in our genetic algorithm.

B. DP Noise Optimization

Given the user connection index $\mu_{s,i}$, the problem in (16) can be simplified as

$$\min_{\sigma_i} \quad \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} \left(\frac{1}{K_i\sigma_i}\right)^2 \mu_{s,i}, \tag{21}$$

$$\text{s.t.} \quad \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i\sigma_i^2\mu_{s,i} \geqslant V_{\max}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{U}} K_i\mu_{s,i}, \tag{21a}$$

$$K_i\sigma_i \geqslant N_{\min}\mu_{s,i}, \quad \forall s \in \mathcal{S}, i \in \mathcal{U}. \tag{21b}$$

From (21), we see that the optimization problem (21) is a convex optimization problem that can be solved using MATLAB's fmincon function.

C. User-BS Connection Optimization

Given the relationship among $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$, $(p_i)_{i\in\mathcal{U}}$, and $(\sigma_i)_{i\in\mathcal{U}}$, next, we introduce the use of a genetic algorithm to solve problem (16). Specifically, the values of $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$ are taken as the genes. For a given $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$, the optimal values of $(p_i)_{i\in\mathcal{U}}$ and $(\sigma_i)_{i\in\mathcal{U}}$ are determined by solving the problem in (17) and the problem in (21). The fitness of each $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$ is the objective function value calculated based on $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}, (p_i)_{i\in\mathcal{U}}, (\sigma_i)_{i\in\mathcal{U}}$. A roulette wheel selection scheme [21] is used to evolve the next generation. The full procedure of our genetic algorithm for solving problem (16) is summaried in Algorithm 2.

V. Numerical Results

In this section, we evaluate the effectiveness of the proposed algorithms. The parameters used in simulations are summarized in Table I. For comparison purposes, we consider an alternating algorithm that iteratively optimizes the parameters in a sequential manner. In simulations, each generation consists of 400 individuals (i.e., 400 $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}$). Among these individuals, 200 individuals are randomly selected from the last generation, 100 individuals are produced by crossover, and 100 individuals are produced by mutation. The solutions and objective function values corresponding to the optimal $(\mu_{s,i})_{i\in\mathcal{U},s\in\mathcal{S}}, (p_i)_{i\in\mathcal{U}}, (\sigma_i)_{i\in\mathcal{U}}$ obtained by the genetic algorithm with 5/10/20 generations are labeled by GA5/GA10/GA20, respectively.

Figure 2 shows how the total privacy leakage changes as the maximum transmit power of each user varies. It can be seen that when N_2 increases from 2 to 5, the genetic algorithm with 5 generations (GA5) is superior to the alternating algorithm. This is because the genetic algorithm can find a globally optimal solution instead of finding a local optimal solution as done by the iterating algorithm. Fig. 2 also shows that when N_2 increases from 2 to 5, the total privacy leakage increases from 5×10^{-7} to 1.5×10^{-6}. This is due to the fact that the increased participating users will result in a higher overall privacy leakage.

Figure 3 shows the distribution of the user's DP noise when the considered algorithms converge. It can be seen that both the alternating algorithm and the genetic algorithm require several users to use large DP noise power. This is because a large DP noise can reduce the privacy leakage of a user. Fig. 3 also shows that the genetic algorithm with 20 generations (GA20) require users to use larger DP noise power compared to that of GA5 or GA10. This is because the algorithm with more generations can explore regions of the solution space more, thus increasing the probability of finding a better solution.

Algorithm 2 Genetic Algorithm for Solving Optimization Problem (16)

```
Randomly initialize the first generation of $\mu_{s,i}$ (i.e., individual).
Set the number of generations $T$ for the genetic algorithm.
for t = 1, 2, ..., $T$ do
    Selection:
    Calculate fitness function for each individual.
    Select individuals for next generation population based on their fitness.
    Crossover:
    Randomly choose pairs of parent individuals.
    Perform crossover on pairs to create new offspring.
    Append offspring individuals to the population.
    Mutation:
    Randomly choose individuals from the population.
    Make random changes (mutations) on chosen individuals to generate the offspring.
    Append offspring individuals to the population.
end for
```

TABLE I
Fixed System Parameters in Simulation

System Parameters	Value
Number of BSs (S)	2
Number of users (U)	10
Upper bound of DP noise error ($V_{\max}$)	12
Uplink bandwidth (B)	180 KHz
Uplink center frequency (f)	2450 MHz
Thermal noise power spectral density (N_0)	-174 dBm/Hz
Scale parameter of channel's rayleigh distribution (l)	1
Minimum DP noise at each user ($N_{\min}$)	100
Minimum communication rate ($R_{\min}$)	100 bps
Maximum number of connections at each BS (N_1)	3

VI. Conclusion

In this paper, we have studied the problem of resource allocation for NOMA enabled secure FL. We have considered an

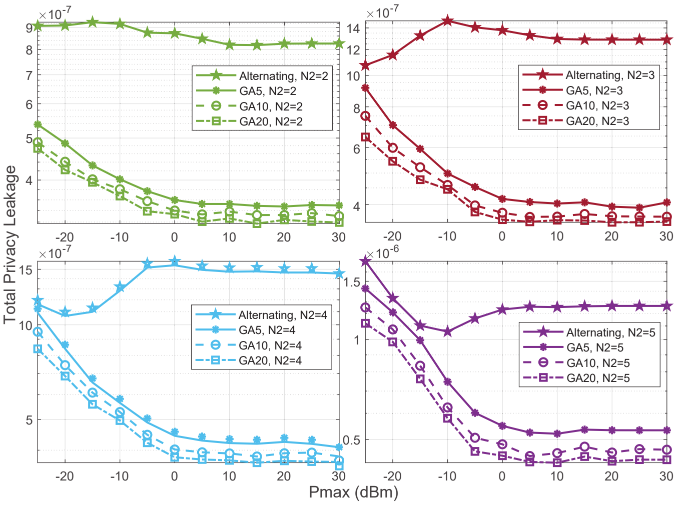

Fig. 2. The total privacy leakage changes as the maximum transmit power of each user varies in diverse N_2 conditions.

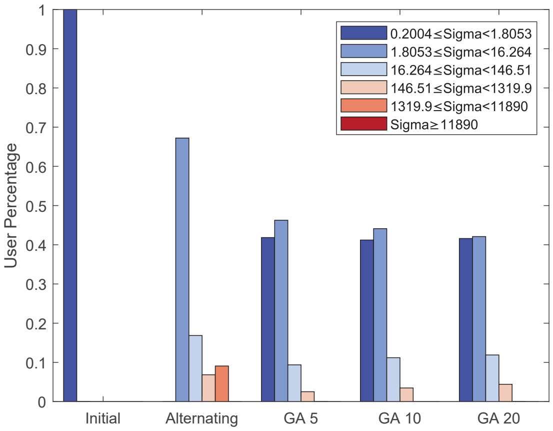

Fig. 3. The final distribution of user's DP noise varies as the optimization algorithm changes.

FL system in which a set of users participate in the FL training through transmitting their trained FL model parameters to the BSs via NOMA techniques. To prevent data leakage, each user uses the DP technique through adding Gaussian noise to its FL model parameters. We have formulated the problem of minimizing overall privacy leakage of all FL participaring users as an optimization problem through jointly optimizing the connections between users and BSs, transmit power of the users, and the DP noise power. To solve the formulated non-convex optimization problem, we have proposed a genetic algorithm to search for feasible solutions in which user connection matrix is taken as gene and the objective function value is taken as the fitness of solution. Simulation results have demonstrated the superiority of the proposed genetic algorithm in terms of privacy leakage protection.

References

[1] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, April 2017.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, May 2020.

[4] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems*, Austin, TX, USA, March 2020.

[5] H. Tang, S. Gan, C. Zhang, T. Zhang, and J. Liu, "Communication compression for decentralized training," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, 2018 Dec.

[6] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, Oct. 2020.

[7] R. Hamdi, M. Chen, A. B. Said, M. Qaraqe, and H. V. Poor, "Federated learning over energy harvesting wireless networks," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 92–103, June 2021.

[8] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, "Communication-efficient federated learning," *Proceedings of the National Academy of Sciences*, vol. 118, no. 17, p. e2024789118, April 2021.

[9] S. Zheng, C. Shen, and X. Chen, "Design and analysis of uplink and downlink communications for federated learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2150–2167, July 2021.

[10] S. Chen, C. Shen, L. Zhang, and Y. Tang, "Dynamic aggregation for heterogeneous quantization in federated learning," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6804–6819, Oct. 2021.

[11] J. Xue, Z. Qu, J. Xu, Y. Liu, and Z. Lu, "Bandwidth allocation for federated learning with wireless providers and cost constraints," *IEEE Transactions on Mobile Computing*, vol. 23, no. 6, pp. 7470–7482, June 2024.

[12] R. Hu, Y. Guo, and Y. Gong, "Concentrated differentially private and utility preserving federated learning," *arXiv preprint arXiv:2003.13761*, 2020.

[13] N. Tavangaran, M. Chen, Z. Yang, J. M. B. D. Silva Jr, and H. V. Poor, "On differential privacy for federated learning in wireless systems with multiple base stations," *arXiv preprint arXiv:2208.11848*, 2022.

[14] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, Oct. 2016, pp. 308–318.

[15] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, April 2020.

[16] M. Seif, R. Tandon, and M. Li, "Wireless federated learning with local differential privacy," in *2020 IEEE International Symposium on Information Theory (ISIT)*. Los Angeles, CA, USA: IEEE, June 2020, pp. 2604–2609.

[17] K. Wei, J. Li, C. Ma, M. Ding, C. Chen, S. Jin, Z. Han, and H. V. Poor, "Low-latency federated learning over wireless channels with differential privacy," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 290–307, Nov. 2021.

[18] T. Liu, B. Di, and L. Song, "Privacy-preserving federated edge learning: Modeling and optimization," *IEEE Communications Letters*, vol. 26, no. 7, pp. 1489–1493, April 2022.

[19] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography*. Beijing, China: Springer, Oct. 2016, pp. 635–658.

[20] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[21] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT press, 1992.