Latency Minimization in Covert Communication-Enabled Federated Learning Network

Nguyen Thi Thanh Van, Nguyen Cong Luong, Huy T. Nguyen, Feng Shaohan, Dusit Niyato, *Fellow, IEEE*, and Dong In Kim, *Fellow, IEEE*

***Abstract*—Federated Learning (FL) as a promising technique is able to address the privacy issues in machine learning. However, due to the broadcast nature of wireless channel, one of the key challenges of FL is its vulnerability to wireless security threats. Thus, in this paper, we consider the model update security in FL. In particular, we propose to adopt a covert communication technique with which a friendly jammer transmits jamming signals to prevent a warden from detecting local model update transmissions of mobile devices in FL. The use of jamming signals reduces the transmission rate of the devices. Thus, we formulate an optimization problem that jointly determines the jamming power, local model transmission power, and local training accuracy to minimize the FL latency, given a security performance requirement. The problem is non-convex, and we propose an alternating descent algorithm to solve it. Extensive simulations are conducted and the results demonstrate the effectiveness and network performance improvement of the proposed algorithm.**

***Index Terms*—Covert communication, federated learning, secure model update, latency minimization.**

I. Introduction

Federated Learning (FL) has recently proposed as a promising technique in machine learning. In the FL, the mobile devices use their local data to train cooperatively a machine learning model required by the server. Then, the mobile devices transmit the local model updates to the server via wireless channels. Since the mobile devices transmit model updates instead of the raw data to the server, FL addresses the privacy issues and network bottleneck.

However, one of the key challenges of FL is its vulnerability to wireless security threats. In particular, due to the broadcast nature of wireless channels, attackers can easily detect the model update transmissions between the server and mobile devices. Then, they can eavesdrop the trained models for certain malicious purposes, or they can infer sensitive information, such as gender, occupation, and location of the mobile devices' owners [1]. More seriously, the attackers can launch jamming attacks to disrupt the model update transmissions between the mobile devices and the server. Nevertheless, such an attacker has not well been investigated in FL. In particular, there are some works, e.g., [2] and [3], proposed to secure the model updates. Specifically, the authors in [2] address the model poisoning attack in which an attacker attempts to directly poison the global model. The authors in [3] combat a backdoor attack in which a backdoor attacker stealthily manipulates the global model so that the attacker-chosen inputs result in wrong predictions. Recently, few works have been proposed to prevent an attacker from disclosing the model updates transmitted from the mobile devices. For example, the authors in [4] adopt a homomorphic encryption technique to encode the local updates from the mobile devices. However, this technique usually requires multi-round communications and high power consumption at the mobile devices.

Recently, covert communication (CC) has been proposed as an effective solution to the wireless security. The key idea of CC is to hide the existence of data communications from a warden by transmitting artificial noise signals along with the data transmission signals. Note that CC is different from the physical layer security (PLS) which aims to prevent the warden from decoding the transmission data. As presented in [5], CC is able to achieve a covert rate, i.e., data transmission rate without being detected, higher than the secrecy rate obtained by PLS. Moreover, CC may not require channel state information (CSI) of the warden, which is hard to be acquired. As a result, CC is a promising solution for several applications such as military and Internet of Things (IoTs).

For the above reasons, in this paper, we propose to leverage CC for the model updates in FL. To the best of our knowledge, this is the first work that adopts the CC in FL. In particular, we consider an FL network which includes a base station (BS), i.e., the model owner or server, mobile devices as workers, and a warden, namely Willie. We aim to secure the local model transmissions from the mobile devices to the BS, and a friendly jammer is deployed that transmits jamming signals to prevent the warden from detecting the transmissions by the mobile devices. However, this jamming reduces the signal-to-interference-plus-noise ratio (SINR) at the BS and increases the FL latency of the system. Thus, we formulate an optimization problem that jointly determines the jamming power of the friendly jammer, and the transmission power and local training accuracy of the devices to minimize the maximum of the FL latency, constrained by the security performance. The optimization problem is non-convex. To solve the problem, develop an alternating descent algorithm by resorting them to the successive convex approximation (SCA). The simulation results show the effectiveness of the proposed algorithm.

The main contributions of the paper are sthe followings:

Manuscript received May 30, 2021; revised August 12, 2021 and October 12, 2021; accepted October 14, 2021. Date of publication October 19, 2021; date of current version December 17, 2021. This work was supported in part by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant 102.02-2019.305, in part by Programme DesCartes, in part by National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) Programme, Alibaba Group through Alibaba Innovative Research (AIR) Program and Alibaba-NTU Singapore Joint Research Institute (JRI), in part by the National Research Foundation, Singapore, under the AI Singapore Programme (AISG) (AISG2-RP-2020-019), WASP/NTU under Grant M4082187 (4080) and Singapore Ministry of Education (MOE) Tier 1 (RG16/20), and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant 2021R1A2C2007638 and the MSIT under Grant IITP-2020-0-01821 supervised by the IITP. The review of this article was coordinated by Dr. Ying-Dar Lin. *(Corresponding author: Nguyen Cong Luong.)*

Nguyen Thi Thanh Van is with the Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi 12116, Vietnam (e-mail: van.nguyenthithanh@phenikaa-uni.edu.vn).

Nguyen Cong Luong is with the Faculty of Computer Science, Phenikaa University, Hanoi 12116, Vietnam, and also with the Research and Technology Institute (PRATI), A&A Phoenix Group JSC, Hanoi 11313, Vietnam (e-mail: luong.nguyencong@phenikaa-uni.edu.vn).

Huy T. Nguyen and Feng Shaohan are with the Institute for Infocomm Research, Singapore 138632, Singapore (e-mail: huynguyencse@gmail.com; Feng_Shaohan@i2r.a-star.edu.sg).

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

Dong In Kim is with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea (e-mail: dikim@skku.ac.kr).

Digital Object Identifier 10.1109/TVT.2021.3121004

TABLE I
LIST OF FREQUENCY SYMBOLS USED IN THIS PAPER

Notation	Description	Notation	Description
T_i	FL latency of device i	N	Number of mobile devices
$\mathbb{P}_{\psi_1}$	Probabillity of data transmission to the BS	S	Size of the global/local model (bits)
p_i	Transmit power of device i	ι	Global accuracy
p_j	Transmit power of jammer	η	Local accuracy
B	Total bandwidth	D_i	Number of data samples of device i
f_i	Computation capacity of device i	C_i	Number of CPU cycles of device i

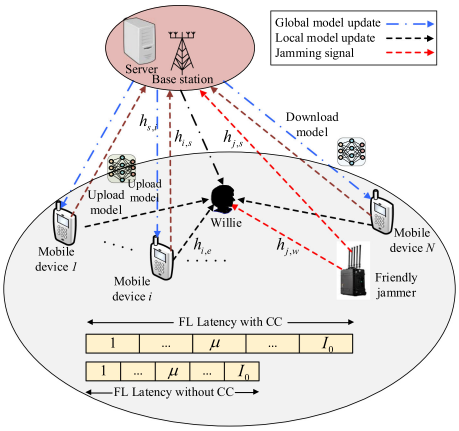

Fig. 1. A covert communication-enabled FL network.

- We propose a covert communication-enabled federated learning (CCFL) system. In the CCFL system, a friendly jammer is deployed to transmit jamming signals so as to prevent a warden from detecting the local model transmissions between the mobile devices and the BS.
- We formulate an optimization problem for the CCFL system. The problem is to optimize the jamming power of the friendly jammer as well as the local accuracy and the transmit power of the mobile devices to minimize the FL latency subject to the covert communication requirement. To solve this problem, we develop an alternating descent algorithm by resorting to the successive convex approximation (SCA).
- We conduct experiments to show the convergence of the proposed algorithm. We evaluate and discuss the performance obtained by the proposed algorithm.

The rest of the paper is organized as follows. In Section II, we present the covert communication-enabled FL system and formulate the FL latency minimization problem. In Section III, we develop an alternating descent algorithm to solve the problem. The simulation results and discussions are presented in Section IV, and the conclusions are given in Section V.

II. SYSTEM MODEL

This section presents the system model, the covert communication-enabled FL system and formulate the FL latency minimization problem. Typical notations used in this paper are summarized in Table I.

We consider an FL network as shown in Fig. 1. The network includes a set $\mathcal{N}$ of N mobile devices, one server at a BS and a friendly jammer. The mobile devices as FL workers cooperatively train a deep neural network (DNN) model required by the server. There is a warden (Willie) in the network that tries to detect the model parameter transmissions from the devices to the BS. We consider the information security for the mobile devices that may be more vulnerable to the adversary. To hide the model update communications transmitted from the devices to the BS, the jammer transmits a jamming signal while the devices transmit the local model updates. It is important that the BS as a central controller decides the jamming power of the jammer and the model update transmission power of the devices. The BS can exchange the decision information via control signals at the beginning of each training iteration. Let $d_{i,\mathrm{s}}$, $d_{j,\mathrm{s}}$, $d_{i,\mathrm{w}}$, and $d_{j,\mathrm{w}}$ denote the distance between device i and the BS, the distance between the jammer and the BS, the distance between device i and Willie, and the distance between the jammer and Willie, respectively. Also, we denote the channel coefficients between device i and the BS and between the device and Willie as $h_{i,s}$ and $h_{i,\mathrm{w}}$, respectively. Additionally, we denote the channel coefficients between the jammer and the BS and that between the jammer and Willie as $h_{j,s}$ and $h_{j,\mathrm{w}}$, respectively. The channel coefficients are assumed to be circularly symmetric complex Gaussian with zero mean and unit variance.

To achieve a high spectrum efficiency, the orthogonal frequency-division multiple access (OFDMA) technique is adopted for the local model transmissions from the devices to the BS. It is important that with the OFDMA technique, the covert communication fits well into the FL environments, compared with the PLS. The reason is that in the FL environments, the number of devices may be very large, and the amount of bandwidth assigned to each device is low. As a result, the noise power at Willie is low that decreases the secrecy rate of PLS, while the covert communication produces a fixed security performance as shown in (9) in this paper.

For FL, let $\boldsymbol{w}$ denote the vector including global model parameters. Each device i has a local dataset $\mathcal{D}_i$ with D_i data samples. Each data sample k consists of an input $\boldsymbol{x}_{ik}$ and its corresponding output $\boldsymbol{y}_{ik}$. We introduce the loss function $g_i(\boldsymbol{w}, \boldsymbol{x}_{ik}, \boldsymbol{y}_{ik})$ that captures the FL performance. The total loss function of device i is given by [7]

$$G_i(\boldsymbol{w}) = \frac{1}{D_i} \sum_{k=1}^{D_i} g_i(\boldsymbol{w}, \boldsymbol{x}_{ik}, \boldsymbol{y}_{ik}). \tag{1}$$

In the FL algorithm, we denote I_0 as the number of global iterations to achieve a global accuracy ι for the global model, and I_i as the number of local iterations at device i to achieve a local accuracy η for the local model. We assume that $G_i(\boldsymbol{w})$ is L-Lipschitz continuous and γ-strongly convex, i.e., $\gamma \boldsymbol{I} \preceq \nabla^2 G_i(\boldsymbol{w}) \preceq L\boldsymbol{I}, \forall i \in \mathcal{N}$, where $\boldsymbol{I}$ presents an identity matrix, and the values of γ and L are determined by the loss function. The assumption is reasonable since the FL loss functions is typically linear or logistic functions. According to Theorem 1 in [6], we use $\frac{a}{1-\eta}$ to approximate the number I_0 of global iterations, where $a = \frac{2L^2}{\gamma^2 \xi} \ln \frac{1}{\iota}$ with ξ being a constant value, i.e., $0 < \xi \leq \frac{\gamma}{L}$.

After the mobile devices train the models, they transmit the trained models, i.e., the local models, to the server. The server generates a new global model by taking average of the local models. As such, the number of elements in the local model can be the same as that in the global model. Therefore, the data size of the local model is equal to that of the global model. Here, the data size of the local model is defined by the number of bits. Let S denote the data size of the local model that is measured in bits that is mapped into n symbols defined as $\mathbf{q}_i = [q_i^1, \ldots, q_i^n]$. For the covert communication, the jammer transmits a jamming signal $\mathbf{q}_j = [q_j^1, \ldots, q_j^n]$.

A. Covert Communication

For simplicity, we omit index μ that indicates the μ-th iteration of the FL training process. The signal received at receiver m (m is s for the BS and w for Willie) is [5]

$$\boldsymbol{z}_{i,m} = \begin{cases} \frac{\sqrt{p_j}h_{j,m}\mathbf{q}_j}{d_{j,m}^{\alpha/2}} + \boldsymbol{\omega}_{i,m}, & \text{if } \Psi_{i,0}, \\ \frac{\sqrt{p_j}h_{j,m}\mathbf{q}_j}{d_{j,m}^{\alpha/2}} + \frac{\sqrt{p_i}h_{i,m}\mathbf{q}_i}{d_{i,m}^{\alpha/2}} + \boldsymbol{\omega}_{i,m}, & \text{if } \Psi_{i,1}, \end{cases} \tag{2}$$

where $\Psi_{i,0}$ specifies the case that device i does not transmit any message to the BS, $\Psi_{i,1}$ refers to the case that device i transmits a message to the BS, p_j and p_i respectively are the jamming power and local update transmission power, α denotes the path-loss exponent, respectively, and $\boldsymbol{\omega}_{i,m} \sim \mathcal{CN}(0, \sigma_{i,m}^2 \boldsymbol{I}_n)$ is the Gaussian noise at receiver m. $\boldsymbol{I}_n$ presents an $n \times n$ identity matrix. Each symbol of the received signal at Willie from device i follows $\sim \mathcal{CN}(0, \sigma_{i,\mathrm{w}}^2 + \theta)$, in which the probability density function (PDF) for Willie's observation of θ for $\theta > 0$ is given by [5]. The SINR at the BS corresponding to device i is [5]

$$\zeta_i = \begin{cases} 0, & \text{if } \Psi_{i,0}, \\ \frac{p_i|h_{i,s}|^2}{\varrho(d_{j,s}^{\alpha}\sigma_{i,s}^2 + p_j|h_{j,s}|^2)}, & \text{if } \Psi_{i,1}, \end{cases} \tag{3}$$

where $\varrho = (\frac{d_{i,s}}{d_{j,s}})^{\alpha}$.

B. Federated Learning Latency

The FL process in each iteration consists of three steps: local training at each device, local model update transmissions of the device, and result aggregation and broadcast at the BS.

1) Local Training Latency: Each device trains the local model by using the gradient method with the step size δ. Let f_i be the computation capacity, i.e., the number of CPU cycles per second, of device i. Then, given the local accuracy η, the number of local iterations required for the local computation at each device is approximately $v \log_2(1/\eta)$, where $v = \frac{2}{(2-L\delta)\delta\gamma}$ [6]. Thus, the computation time at device i required for the data processing at each iteration is [7]

$$\tau_i = \frac{vC_iD_i\log_2(1/\eta)}{f_i} = \frac{A_i\log_2(1/\eta)}{f_i}, \forall i \in \mathcal{N}, \tag{4}$$

where C_i (cycles/bit) is the number of CPU cycles required for computing one data sample at device i and $A_i = vC_iD_i$.

2) Model Upload Latency: The data rate achieved by device i for transmitting its local model to the BS is given by [5]

$$r_i = \mathbb{P}_{\psi_{i,1}} b_i \log_2\left(1 + \frac{p_i|h_{i,s}|^2}{\varrho(p_j|h_{j,s}|^2 + d_{j,s}^{\alpha}\sigma_{i,s}^2)}\right), \tag{5}$$

where $\mathbb{P}_{\psi_{i,1}}$ refers to the probability of data transmission to the BS, b_i is the bandwidth assigned to device i, and $\sigma_{i,s}^2 = b_i\sigma_0^2$ where σ_0^2 is the one-sided power spectral density level of the noise at the BS. Due to the limited bandwidth, we have $\sum_{i=1}^{N} b_i \le B$, where B is the total bandwidth. Then, the time required for transmitting the local model from device i to the BS is given by $t_i = S/r_i$, where S is the data size of the local model. As such, t_i (and hence the FL latency) is inversely proportional to the probability of local model transmission to the BS, i.e., $\mathbb{P}_{\psi_{i,1}}$.

3) Aggregation and Broadcast Latency: In this step, the BS aggregates the global models and then broadcasts the global model to all devices in the downlink. Due to the high power budget of the BS, the downlink latency is ignored. Therefore, the FL latency of device i at each iteration is $\tau_i + t_i$, and the FL latency of the device over I_0 iterations is

$$T_i = I_0(\tau_i + t_i) = \frac{a}{1-\eta}(\tau_i + t_i). \tag{6}$$

The total training latency or the FL latency is the maximum latency among the devices, that is $\max_{i\in\mathcal{N}} T_i$.

C. False Alarm and Miss Detection Probabilities

When Willie decides $\Psi_{i,1}$ while $\Psi_{i,0}$ is true, a false alarm (FA) with probability $\mathbb{P}_{i,FA}$ occurs, while if Willie decides $\Psi_{i,0}$ while $\Psi_{i,1}$ is true, a miss detection (MD) with probability $\mathbb{P}_{i,MD}$ happens. The security requirement of the BS is expressed by the condition [7]

$$\mathbb{P}_{i,FA} + \mathbb{P}_{i,MD} \ge 1 - \epsilon, \forall i \in \mathcal{N}, \tag{7}$$

where $\epsilon > 0$ is a covertness requirement.

As illustrated in [5], the error detection probability at Willie is given by

$$\mathbb{P}_{i,FA} + \mathbb{P}_{i,MD} = \begin{cases} 1 + e^{-\frac{(\vartheta_i - \sigma_{i,\mathrm{w}}^2)}{\psi_{i,0}}} - e^{-\frac{(\vartheta_i - \sigma_{i,\mathrm{w}}^2)}{\psi_{i,1}}}, & \text{for } \vartheta_i - \sigma_{i,\mathrm{w}}^2 \ge 0, \\ 1, & \text{for } \vartheta_i - \sigma_{i,\mathrm{w}}^2 < 0, \end{cases} \tag{8}$$

where $\psi_{i,0} = \frac{p_j}{d_{j,\mathrm{w}}^{\alpha}}$ and $\psi_{i,1} = \frac{p_j}{d_{j,\mathrm{w}}^{\alpha}} + \frac{p_i}{d_{i,\mathrm{w}}^{\alpha}}$. For the minimal detection error, Willie must to determine a power threshold ϑ_i to have its true decision. The optimal power threshold corresponding to device i is determined by $\min_{\vartheta_i} \mathbb{P}_{i,FA} + \mathbb{P}_{i,MD}$. According to [5], and since $\vartheta_i^* \ge \sigma_{i,\mathrm{w}}^2$ is always true, the optimal value of ϑ_i^* is given by $\vartheta_i^* = (\frac{\psi_{i,0}\psi_{i,1}}{\psi_{i,0}-\psi_{i,1}})\ln(\frac{\psi_{i,0}}{\psi_{i,1}}) + \sigma_{i,\mathrm{w}}^2$. Therefore, by substituting ϑ_i^* into (8) we have

$$\mathbb{P}_{i,FA}(\vartheta_i^*) + \mathbb{P}_{i,MD}(\vartheta_i^*) = 1 + e^{-\left(\frac{\psi_{i,1}}{\psi_{i,0}-\psi_{i,1}}\right)\ln\left(\frac{\psi_{i,0}}{\psi_{i,1}}\right)} - e^{-\left(\frac{\psi_{i,0}}{\psi_{i,0}-\psi_{i,1}}\right)\ln\left(\frac{\psi_{i,0}}{\psi_{i,1}}\right)}. \tag{9}$$

This is the minimal error detection probability at Willie which also presents the covertness of the devices in the FL system.

D. Problem Formulation

In general, we can increase the jamming power to prevent Willie from detecting the model update transmissions from the devices to the BS. However, this reduces the SINR at the BS and increases the FL latency. Otherwise, the local accuracy η in this paper is defined as the mean square error between the target value and the predict value. Thus, we can increase the local accuracy η to reduce the number of local iterations, thereby reducing the computation time at the devices. However, this requires more global iterations to achieve the global accuracy ι that may increase the FL latency. Therefore, the problem is to determine the friendly jamming power, the model upload transmission power of the devices, and the local accuracy to minimize the FL latency while guaranteeing the CC constraint. Hence, we have the following

optimization problem

$$\min_{\eta,\mathbf{p}_i=[p_i,p_j]} \quad \max_{i\in\mathcal{N}} T_i(\eta,\mathbf{p}_i), \tag{10a}$$

$$\text{s.t.} \quad p_i \le p_i^{\max}, \forall i \in \mathcal{N}, \tag{10b}$$

$$p_j \le p_j^{\max}, \tag{10c}$$

$$\mathbb{P}_{i,FA}(\vartheta_i^*) + \mathbb{P}_{i,MD}(\vartheta_i^*) \ge 1-\epsilon, \forall i \in \mathcal{N}, \tag{10d}$$

$$0 \le \eta \le 1, \tag{10e}$$

where $p_i^{\max}$ is the maximum transmit power of device i, and $p_j^{\max}$ is the maximum transmit power of the jammer. In particular, the constraint in (10b) specifies the power constraint of device i, while that in (10d) is the CC requirement, the left-hand-side (LHS) of which is given in (9). In addition, (10e) is the local accuracy constraint.

III. Alternating Descent Algorithm

It is observed that the optimization problem (10) is nonconvex because the objective function and the constraint in (10d) are nonconvex. In this section, we develop an alternating descent algorithm to solve it.

First, let us make the following change of variable $\frac{1}{\eta} = 1 + \frac{1}{\rho}$, which satisfies the linear constraint $\rho > 0$. Thus, the FL latency of device i is reformulated by

$$T_i(\rho,\mathbf{p}_i) = a(1+\rho)(\tau_i^\rho + t_i), \tag{11}$$

where $\tau_i^\rho = \frac{A_i}{f_i}\log_2\frac{1+\rho}{\rho}$. Accordingly, the problem (10) is rewritten as follows:

$$\min_{\rho,\mathbf{p}_i=[p_i,p_j]} \quad \max_{i\in\mathcal{N}} T_i(\rho,\mathbf{p}_i), \tag{12a}$$

$$\text{s.t} \quad \text{(10b), (10c), (10d),}$$

$$\rho > 0. \tag{12b}$$

It can be observed from (12) that the constraints in (10b), (10c), and (10d) are only related to the power, i.e., the jamming power p_j and transmit power p_i, and the remaining constraint in (12b) is only related to the local learning accuracy, i.e., ρ. Therefore, we can divide problem (12) into two sub-problems that are alternatively optimized at each iteration. Let $(\rho^{(\kappa)}, \mathbf{p}_i^{(\kappa)})$ be a feasible point for (12) that is found from the $(\kappa-1)$-th iteration. In iteration κ, we fix $\rho = \rho^{(\kappa)}$ and determine $\mathbf{p}_i^{(\kappa+1)}$, then we fix $\mathbf{p}_i = \mathbf{p}_i^{(\kappa+1)}$ to determine $\rho^{(\kappa+1)}$.

A. Sub-Optimization Problem 1

Given a fixed $\rho^{(\kappa)}$, we have the following sub-problem

$$\min_{\mathbf{p}_i=[p_i,p_j]} \quad \max_{i\in\mathcal{N}} T_i(\mathbf{p}_i),$$

$$\text{s.t} \quad \text{(10b), (10c) and (10d).} \tag{13}$$

First, we consider the objective in problem (13). The data rate achieved by device i for transmitting its local model update to the BS can be written as

$$r_i(\mathbf{p}_i) = \mathbb{P}_{\psi_{i,1}} b_i (\Xi_i(\mathbf{p}_i) - \Upsilon_i(p_j) - \log_2 \varrho), \tag{14}$$

where $\Xi_i(\mathbf{p}_i) = \log_2(p_i|h_{i,s}|^2 + \varrho(p_j|h_{j,s}|^2 + d_{j,s}^\alpha \sigma_{i,s}^2))$ and $\Upsilon_i(p_j) = \log_2(p_j|h_{j,s}|^2 + d_{j,s}^\alpha\sigma_{i,s}^2)$ are concave [8].

As the function $\Upsilon_i(p_j)$ is concave, its gradient, i.e, $\nabla\Upsilon_i(p_j^{(\kappa)})$ is its super-gradient [8], so

$$\Upsilon_i(p_j) \le \Upsilon_i(p_j^{(\kappa)}) + \nabla\Upsilon_i(p_j^{(\kappa)})(p_j - p_j^{(\kappa)}), \tag{15}$$

where $\nabla\Upsilon_i(p_j^{(\kappa)}) = \frac{|h_{j,s}|^2}{(p_j^{(\kappa)}|h_{j,s}|^2 + d_{j,s}^\alpha\sigma_{i,s}^2)\ln 2}$. Therefore

$$r_i(\mathbf{p}_i) \ge \mathbb{P}_{\psi_{i,1}} b_i \Big[\Xi_i(\mathbf{p}_i) - \Upsilon_i(p_j^{(\kappa)}) - \nabla\Upsilon_i(p_j^{(\kappa)})(p_j - p_j^{(\kappa)}) - \log_2\varrho\Big] \triangleq r_i^{(\kappa)}(\mathbf{p}_i). \tag{16}$$

Using (16), the objective in (13) is approximated by the following function

$$T_i(\mathbf{p}_i) \le a(1+\rho^{(\kappa)})(\tau_i(\rho^{(\kappa)}) + t_i^{(\kappa)}(\mathbf{p}_i)) \triangleq T_i^{(\kappa)}(\mathbf{p}_i), \tag{17}$$

where $t_i^{(\kappa)}(\mathbf{p}_i) \triangleq \frac{S}{r_i^{(\kappa)}(\mathbf{p}_i)}$ over the trust region $r_i^{(\kappa)}(\mathbf{p}_i) \ge 0$.

Let $h(r_i^{(\kappa)}(\mathbf{p}_i)) = \frac{S}{r_i^{(\kappa)}(\mathbf{p}_i)}$, then h is a decreasing and convex function. From (16), $r_i^{(\kappa)}(\mathbf{p}_i)$ is a positive and concave function. Therefore, according to [9], $t_i^{(\kappa)}(\mathbf{p}_i)$ is a convex function, and $T_i^{(\kappa)}(\mathbf{p}_i)$ is convex.

Now, we consider the constraint in (10d). According to (9), and after some mathematical transformations, the constraint in (10d) can be rewritten as

$$p_i\ln p_i + \beta p_j \ln \beta p_j - (p_i+\beta p_j)\ln(p_i+\beta p_j) \le p_i \ln\epsilon, \tag{18}$$

where $\beta = (\frac{d_{i,\mathrm{w}}}{d_{j,\mathrm{w}}})^\alpha$. Let the LHS in (18) be $\Omega_i(\mathbf{p}_i)$, then $\Omega_i(\mathbf{p}_i) = u_i(\mathbf{p}_i) + v_i(\mathbf{p}_i)$, where $u_i(\mathbf{p}_i) = p_i\ln p_i + \beta p_j \ln\beta p_j$ is convex, and $v_i(\mathbf{p}_i) = -(p_i+\beta p_j)\ln(p_i+\beta p_j)$ is concave [8]. Similar to (15), we have

$$\Omega_i(\mathbf{p}_i) \le u_i(\mathbf{p}_i) + v_i(\mathbf{p}_i^{(\kappa)}) + \nabla^T v_i(\mathbf{p}_i^{(\kappa)})(\mathbf{p}_i - \mathbf{p}_i^{(\kappa)}) \triangleq \Omega_i^{(\kappa)}(\mathbf{p}_i) \le p_i\ln\epsilon, \tag{19}$$

where $\nabla^T v_i(\mathbf{p}_i) = [-\ln(p_i+\beta p_j) - 1, -\beta\ln(p_i+\beta p_j) - \beta]$. It is observed that $\Omega_i^{(\kappa)}(\mathbf{p}_i)$ is a convex function. Hence, the nonconvex constraint in (18) is innerly approximated by the convex constraint in (19). From (17) and (19), the sub-optimization problem 1 can now be equivalently expressed by

$$\min_{\mathbf{p}_i=[p_i,p_j]} \quad \max_{i\in\mathcal{N}} T_i^{(\kappa)}(\mathbf{p}_i),$$

$$\text{s.t} \quad \text{(10b), (10c) and (19).} \tag{20}$$

B. Sub-Optimization Problem 2

Given a fixed $\mathbf{p}_i^{(\kappa)}$, we have the following sub-problem

$$\min_{\rho} \quad \max_{i\in\mathcal{N}} T_i(\rho),$$

$$\text{s.t} \quad \text{(12b).} \tag{21}$$

The objective in (21) can be rewritten as $T_i(\rho) = \frac{aA_i}{f_i}f(\rho) + at_i(\mathbf{p}_i^{(\kappa)})(1+\rho)$, where $f(\rho) = (1+\rho)\log_2\frac{1+\rho}{\rho}$. By taking the second derivative of $f(\rho)$ with respect to ρ, we can easily prove that $f(\rho)$ is a convex function. Thus, the convexity of objective function in (21) and the linear constraint (12b) make sub-optimization problem (21) being convex.

Taking any feasible point $(\rho^{(0)}$ for(12b), it follows from (20) that initialized by a feasible point $(\rho^{(0)}, \mathbf{p}_i^{(\kappa)})$ for the convex constraint (19), we iterate

$$\min_{\mathbf{p}_i=[p_i,p_j]} \quad \max_{i\in\mathcal{N}} (\Omega_i^{(\kappa)}(\mathbf{p}_i) - p_i\ln\epsilon),$$

$$\text{s.t} \quad \text{(10b) and (10c).} \tag{22}$$

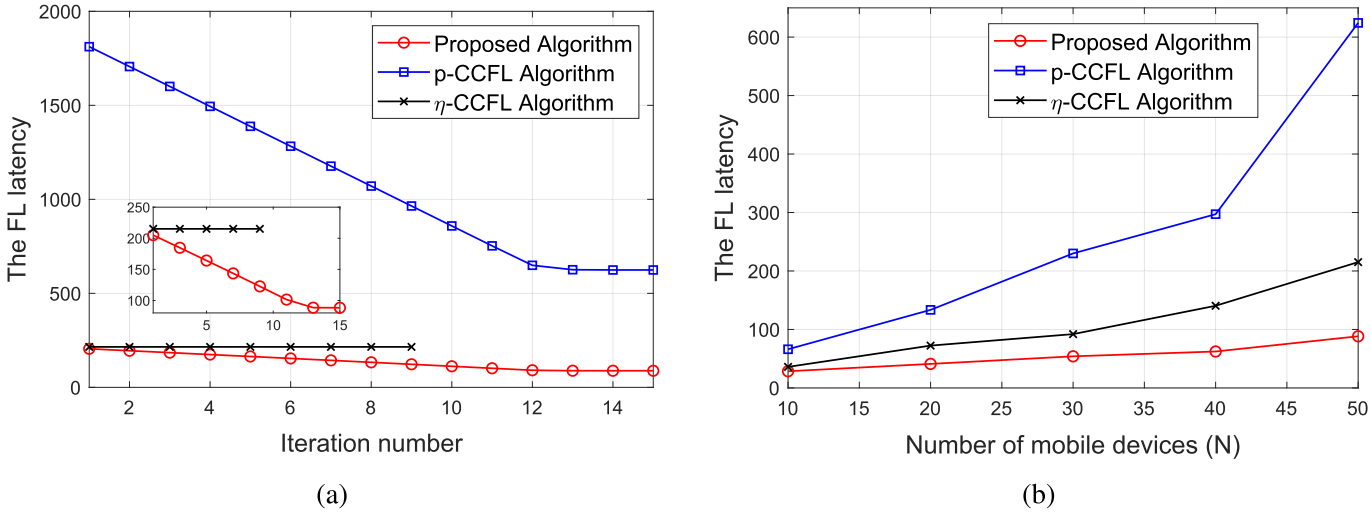

Fig. 2. The FL latency versus (a) iteration number and (b) number of mobile devices.

Algorithm 1: Alternating Descent Algorithm.

1: Initialize: Take any feasible point $\rho^{(0)}$ for (12b), iterate (22) for a feasible point $(\rho^{(0)}, \mathbf{p}_i^{(\kappa)})$ for (20). Set $\kappa = 0$.
2: **repeat**
3: Solve the problem (20) for $\rho = \rho^{(\kappa)}$ to obtain $\mathbf{p}_i^{(\kappa+1)} = \mathbf{p}_i^*$;
4: Solve the problem (21) for $\mathbf{p}_i = \mathbf{p}_i^{(\kappa+1)}$ to obtain $\rho^{(\kappa+1)} = \rho^*$;
5: $\kappa \leftarrow \kappa + 1$;
6: **until** Convergence.

TABLE II
SIMULATION PARAMETERS

Parameters	*Value*	Parameters	Value
$\mathbb{P}_{\psi_1}$	0.7	S	28.1 kbits
ϵ	0.1	α	2.5
$p_i^{\max}$, $p_j^{\max}$	10 dBm	$f_i^{\max}$	2 GHz
ι	10^{-3}	δ, ξ	1/10
B	20 MHz	D_i	500

for $\kappa = 0, 1, \ldots$ until the value of the objective in (22) reaches less than or equal to 0, making $(\rho^{(0)}, \mathbf{p}_i^{(\kappa)})$ feasible for (20). Therefore, it qualifies as an initial point for the above proposed alternating decent process. Algorithm 1 outlines the steps to solve the min-max latency optimization problem (12).

IV. PERFORMANCE EVALUATION

We consider a network with $N = 50$ devices, a friendly jammer, and a warden that are distributed randomly in a square area of size of 500 m $\times$ 500 m. The BS is at the center of area and at an altitude of 25 m above the ground-level. We consider a building or densely populated area, and thus we use the log-distance path loss model, that is $128.1 + 37.6 \log_{10} d$ (d is in km) and the standard deviation of shadowing fading is 8 dB. In addition, C_i is uniformly distributed in $[1, 3] \times 10^4$ cycles/sample. Table II lists the other simulation parameters, the values of which are similar to those in [6] and [5].

To evaluate the proposed algorithm, we introduce p-CCFL and η-CCFL algorithms as baseline schemes. With the p-CCFL, we optimize $\mathbf{p}_i$ and fix the local accuracy η. Here, we set $\eta = 10/11$ such that the constrain in (18d) is satisfied. With the η-CCFL, we optimize η while fixing $\mathbf{p}_i$, which are the feasible points to sastify the constraints in (18b) and (18c).

First, we discuss the convergence of the algorithms and the FL latency obtained by the algorithms. As shown in Fig. 2(a), all the three algorithms are able to converge quickly to stable values, i.e., with few iterations. Moreover, the FL latency value obtained by the proposed algorithm is much lower than those obtained by the baseline algorithms. Especially, shown in Fig. 2(b), the performance gap increases with the increase the number of mobile devices N, that demonstrates the effectiveness and scalability of the proposed algorithm. Note that as N increases, the FL latency of the three algorithms increases. The reason is that when N increases, the total bandwidth B is divided to more devices that reduces the transmission rate of the devices and increases the FL latency.

Next, we discuss how the security performance and the FL latency change as the amming power p_j varies. In this case, the proposed algorithm is actually the η-CCFL algorithm that only optimizes η. As shown in Fig. 3(a), as $p_j/p_j^{\max}$ increases, the security performance, i.e, the covert probability, increases. The reason is that the higher jamming power causes a higher interference Willie and makes it more difficult to detect the update transmissions of the devices. However, the higher jamming power also reduces the SINR of the received signals at the BS. Thus, as shown in Fig. 3(a), the FL latency increases with the increase of $p_j/p_j^{\max}$. It is also seen from Fig. 3(a) that given a value of $p_j/p_j^{\max}$, both the security performance and FL latency increase as the update transmission power, i.e., p_i, of the devices decreases. The reason is that the increase of p_i increases the ratio p_j/p_i and reduces the update transmission rate of the devices.

Finally, we discuss the impact of the threshold ϵ on the security performance and the FL latency. As shown in Fig. 3(b), as the threshold ϵ increases, the FL latency decreases. However, the security performance decreases in the proposed algorithm and p-CCFL algorithms. This is due to the fact that as the security requirement decreases, i.e, ϵ increases, the training process is faster. It is clear that the proposed algorithm achieves better performance than the baseline algorithms. With the η-CCFL algorithm, the latency and the security performance does not

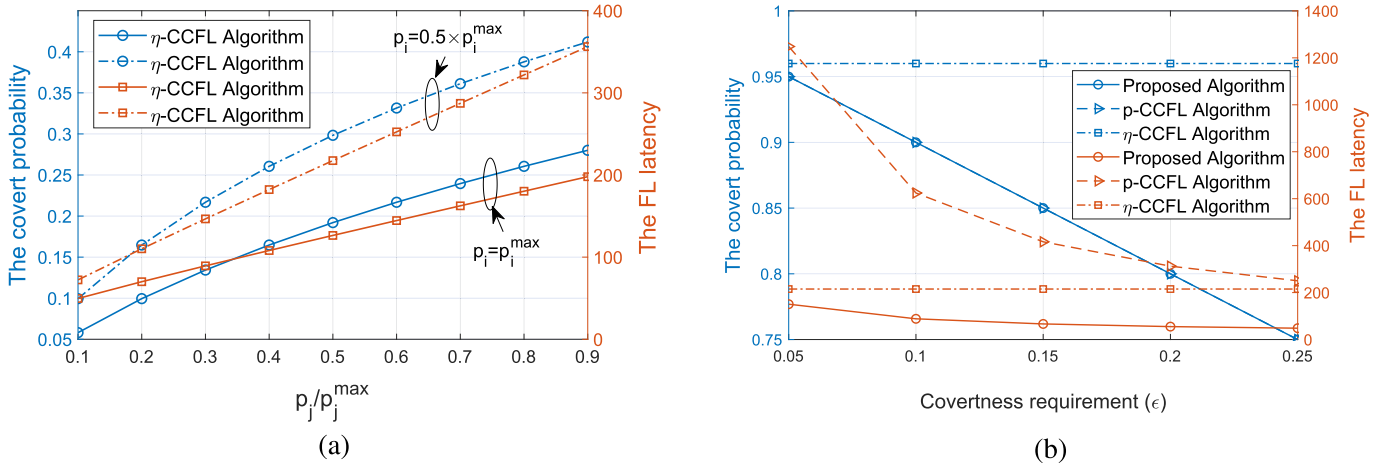

Fig. 3. Security performance and the FL latency versus (a) jamming power and (b) security threshold ϵ.

change. The reason is that the η-CCFL algorithm aims to optimize η that is not affected by the change of ϵ.

V. Conclusion

In this paper, we have investigated the FL latency and communication security in the FL network. In particular, we have proposed to adopt the covert communication in FL. Providing the covert communication results in the increase of FL latency. Thus, we have formulated an optimization problem that minimizes the FL latency constrained to a security requirement. To solve the problem, we have proposed the alternating descent algorithm. The simulation results have shown the effectiveness of the proposed model and algorithm in terms of both FL latency and communication security. As the future work, the dynamic prices set by the jammer can be investigated. In this case, the Stackelberg game can be used to model the interactions between the jammer and the BS. Furthermore, a general scenario with multiple wardens can be considered.

References

[1] S. Truex, L. Liu, M. E. Gursoy, L. Yu, and W. Wei, "Demystifying membership inference attacks in machine learning as a service," *IEEE Trans. Services Comput.*, vol. 14, no. 6, pp. 2073–2089, Nov./Dec. 2021.

[2] A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Analyzing federated learning through an adversarial lens," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 634–643.

[3] K. Liu, B. Dolan-Gavitt, and S. Garg, "Fine-pruning: Defending against backdooring attacks on deep neural networks," in *Proc. Int. Symp. Res. Attacks, Intrusions, Defenses*, Springer, 2018, pp. 273–294.

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

[5] M. Forouzesh, P. Azmi, N. Mokari, and K. K. Wong, "Covert communications versus physical layer security," 1803, *arXiv:1803.06608*.

[6] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021, doi: 10.1109/TWC.2020.3037554.

[7] K. Shahzad, X. Zhou, and S. Yan, "Covert wireless communication in presence of a multi-antenna adversary and delay constraints," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12 432–12436, Dec. 2019.

[8] H. Tuy, *Convex Analysis and Global Optimization*. Berlin, Germany: Springer, 1998.

[9] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.