Article

Resource Allocation for Intelligent Reflecting Surfaces Assisted Federated Learning System with Imperfect CSI

Wei Huang [1], Zhiren Han [1], Li Zhao [2,*] , Hongbo Xu [2], Zhongnian Li [2] and Ze Wang [2]

1 Wuhan Maritime Communication Research Institute (WMCRI), Wuhan 430079, China; wolf185269571@163.com (W.H.); hanzr@csic722.com (Z.H.)
2 Department of Electronic and Information Engineering, Central China Normal University, Wuhan 430079, China; xuhb@ccnu.edu.cn (H.X.); zhongnian.li@mails.ccnu.edu.cn (Z.L.); 17771968067@163.com (Z.W.)
* Correspondence: zhli@mails.ccnu.edu.cn

Abstract: Due to its ability to significantly improve the wireless communication efficiency, the intelligent reflective surface (IRS) has aroused widespread research interest. However, it is a challenge to obtain perfect channel state information (CSI) for IRS-related channels due to the lack of the ability to send, receive, and process signals at IRS. Since most of the existing channel estimation methods are developed to obtain cascaded base station (BS)-IRS-user devices (UDs) channel, this paper studies the problem of computation and communication resource allocation of the IRS-assisted federated learning (FL) system based on the imperfect CSI. Specifically, we take the statistical CSI error model into consideration and formulate the training time minimization problem subject to the rate outage probability constraints. In order to solve this issue, the semi-definite relaxation (SDR) and the constrained concave convex procedure (CCCP) are invoked to transform it into a convex problem. Subsequently, a low-complexity algorithm is proposed to minimize the delay of the FL system. Numerical results show that the proposed algorithm effectively reduces the training time of the FL system base on imperfect CSI.

Keywords: federated learning; intelligent reflector surfaces; imperfect channel state information; outage probability

1. Introduction

In recent years, with the advancement of 5G technology, the internet of things (IoT) has seen rapid development. In the world of "internet of everything", many sensors are deployed in the environment to collect information in real time and generate a large amount of data. The traditional machine learning (ML) framework stores data on a central node, and its all functions are implemented on this node. However, such a centralized learning framework has high latency and poses a great challenge to privacy protection. Fortunately, due to the significant increase in data processing capabilities of mobile devices, the concept of edge learning has been introduced to solve these problems, which processes data at the edge rather than at the cloud center. Federated learning (FL) is one of the most promising edge learning frameworks, where user devices (UDs) only send local models calculated by local resource to the base station (BS) without sharing local data [1,2].

In FL, the time and energy consumed in model training are the important aspects of its performance, which depends on firmly the allocated computation resources for local model training and communication resources for model transmission between BS and UDs [3]. In addition, due to the limited computing power and energy reserve of UDs, the problem of joint computation and communication resource allocation is crucial for the deployment of FL in wireless networks. The authors proposed a resource management algorithm for the multi-access edge computing enabled FL system in [4,5]. In [6,7], the resource scheduling problem was studied under imperfect channel state information (CSI).

Due to the development of the metamaterial technology, an emerging paradigm called intelligent reflector surfaces (IRS) was introduced into the field of wireless communications. IRS can significantly improve the wireless communication efficiency by adjusting the phase shift of the incident signal using passive reflective elements, which are low-cost and low energy [8,9]. Thus, this attracts people to study the beneficial effects of IRS in FL [10–12]. However, these studies are based on the assumption that perfect CSI can be obtained. Unfortunately, due to the lack of the ability to receive, send and process signals for IRS, it is a challenge to estimate IRS-related channels, including the UDs-IRS channel and the IRS-BS channel, which thus have attracted widespread research attention. At present, channel estimation of IRS-related channels can be divided into two methods. One is to directly estimate IRS-related channels [13,14]. In [13], some active channel components can be equipped on the IRS to enable it to receive, send and process signals. The authors proposed a channel estimation protocol to estimate the channel by adjusting the reflective elements ON/OFF in [14]. However, these methods will bring additional cost, such as hardware cost and power consumption cost, which will cause an unaffordable burden on the IRS. The other is to estimate the cascade UD-IRS-BS channel. The advantage of this method is that the cascaded channel can be estimated without additional hardware and power cost. Therefore, there are a lot of works dedicated to IRS cascaded channel estimation [15–18]. In [15,16], the transmission protocol was proposed to estimate the cascaded channel. The author studied the robust transmission design base on imperfect CSI in [17,18].

Against the above background, this paper studies the computation and communication resource allocation problem in IRS-assisted FL systems based on imperfect CSI. Specifically, we aim to design the resource allocation algorithm to minimize the delay of FL. The main contributions of this paper can be summarized as follows:

- In this paper, we study the FL system assisted by IRS under the imperfect CSI, and outage probability is introduced to characterize the impact of imperfect CSI. We aim to minimize the delay of FL by jointly designing the computation and communication resource allocation;
- Due to the non-convex form of the objective function, the coupling effect of multiple variables and the binary constraint brought by the user selection vector, the optimization problem is difficult to solve. By introducing additional variables and invoking the semi-definite relaxation (SDR) and the constrained concave convex procedure (CCCP) techniques, we convert it into a convex optimization problem. Then, a low-complexity algorithm is proposed to optimize the computation and communication resources allocation;
- The simulation results demonstrate that the effectiveness of proposed algorithm in reducing the training time of the FL system base on imperfect CSI.

The rest of this paper is organized as follows. The system model and channel model are described in Section 2. In Section 3, we formulate the latency minimization problem. The iterative algorithm to find the local optimal solution of the latency minimization problem is given in Section 4. In Section 5, we discuss the simulation results. Finally, our conclusions are drawn in Section 6.

Notations: In the paper, scalars, vectors and matrices are denote by lowercase letters, lowercase bold letters and uppercase bold letters, respectively. $C^{M \times N}$ denotes the space of $M \times N$ complex matrix. j denotes the imaginary unit. $\boldsymbol{I}$ denotes the identity matrix with appropriate dimensions. $(\cdot)^H$ and $Tr(\cdot)$ represent the Hermitian transpose and the trace, respectively. $F_X(\cdot)$ denotes the cumulative distribution function (CDF) of the variable X. Finally, $CN(0, \sigma^2)$ denotes the circularly symmetric complex Gaussian (CSCG) associated with zero-mean and variance σ^2. The other parameters used in the paper are listed in Table 1. The acronyms used in the paper are listed in Table 2.

Table 1. Key symbols.

Symbol	Definition	Symbol	Definition
K	Number of UDs	N	Number of IRS elements
$\mathcal{K}$	Set of UDs	$\mathcal{N}$	Set of IRS elements
τ_k^l, e_k^l	The time and energy consumed on local training of the k-th UD	τ_k^u, e_k^u	The time and energy consumed on the model transmission of the k-th UD
L	Number of local iteration	α_k	Effective capacitance coefficient of the k-th UD's computing chipset
c_k	Number of CPU cycles for the k-th UD to process one sample data	D_k	Training data set of the k-th UD
f_k	CPU frequency variable of the k-th UD for local training	f_{min}, f_{max}	The minimum and maximum computation capacity of UDs
p_k	The transmission power variable of the k-th UD for the model transmission	p_{max}	The max transmission power of UDs
b	Bandwidth of each UD	$h_{d,k}, h_{r,k}, g_k$	Channel from the k-th UD to BS, the k-th UD to IRS and from IRS to BS
σ^2	The noise power spectral density	$\varepsilon_{d,k}, \varepsilon_{a,k}$	Estimation error of direct channel and cascaded channel
R_k	The transmission rate of the k-th UD	ν_k	The transmission target rate of the k-th UD

Table 2. Acronyms.

Acronym	Meaning	Acronym	Meaning
ML	Machine learning	FL	Federated learning
UD	User device	BS	Base station
CSI	Channel state information	IRS	Intelligent reflector surfaces
SDR	Semi-definite relaxation	CCCP	Constrained concave convex procedure
CCDF	Complementary cumulative distribution function	CSCG	Circularly symmetric complex Gaussian
MINLP	Mixed integer non-linear programming	QCQP	Quadratically constrained quadratic program

2. System Model and Channel Model

In this section, we introduce the IRS-assisted FL system model, its CSI error model and outage probability as follows.

2.1. System Model

An FL system with IRS is considered as shown in Figure 1, where an IRS, composed of N reflective elements, deployed in the system to assist the communication between a single-antenna BS and K single-antenna UDs. UDs and reflective elements are indexed by $\mathcal{K} = \{1, \ldots, K\}$ and $\mathcal{N} = \{1, \ldots, N\}$. In the FL system, after receiving the global model from the BS, UDs train it to obtain the local models using the local dataset. Then, the local models are sent to BS to be aggregated into a new global model. This process is repeated until the accuracy of the learning model is reached. Specifically, the process can be divided

into three steps, which are "local model computation", "local model transmission", and "global model aggregation and broadcast" [19,20]. The details are as follows:
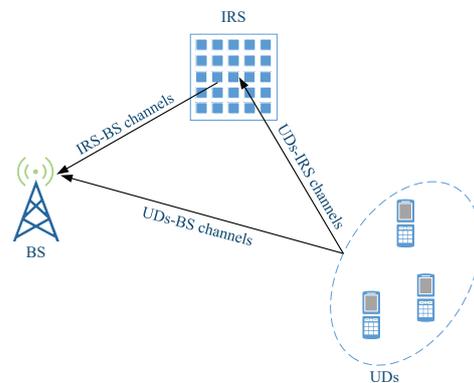

Figure 1. An FL system model with IRS.

2.1.1. Local Model Computation

At this step, UDs use local sample subsets to update the global model, which is received from BS. Assuming that local sample subsets are enough in each UD, from which D_k are taken out for local training. Let c_k denotes the number of CPU cycles required to train one sample data, and f_k denotes CPU frequency. The time and energy consumed on local training of the k-th UD can be expressed as:

$$\tau_k^l = L\frac{c_k|D_k|}{f_k},\tag{1}$$

$$e_k^l = L\frac{\alpha_k}{2}f_k^2 c_k D_k,\tag{2}$$

where L is the number of local iteration, and $\alpha_k/2$ denotes the effective capacitance coefficient of the k-th UD's computing chipset [21].

2.1.2. Local Model Transmission

After finishing local model computation, the local models are sent to BS via multiple channel access, such as the frequency division multiple access (FDMA) or the time division multiple access (TDMA) [19]. In this paper, we adopt FDMA for communication between BS and UDs. The total band width is equally divided into K sub-channels, each of which has a bandwidth b. Let p_k denotes the k-th UD's transmit power, and the transmission rate of the k-th UD can be written as

$$R_k = b\log(1 + \frac{|h_k|^2 p_k}{\sigma^2}),\tag{3}$$

where $h_k = h_{d,k} + \boldsymbol{g}_k\Theta\boldsymbol{h}_{r,k}$ is the composite channels spanning from the k-th UD to BS. The channel from the k-th UD to IRS and from IRS to BS are, respectively, modeled by $\boldsymbol{h}_{r,k} \in C^{N\times 1}$ and $\boldsymbol{g}_k \in C^{1\times N}$, and the direct channel from the k-th UD to BS is modeled by $h_{d,k} \in C$. σ^2 denotes the noise power spectral density. The phase shift effect of IRS is defined as a diagonal matrix, that is $\Theta = diag\{e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_n}\}$, where $\theta_n \in [0, 2\pi]$ for $n \in \mathcal{N}$. Accordingly, the delay and energy consumed in this step can be written as

$$\tau_k^u = \frac{d}{R_k},\tag{4}$$

$$e_k^u = \tau_k^u p_k,\tag{5}$$

where d denotes the size of local model.

2.1.3. Global Model Aggregation and Broadcast

At this step, BS receives the local models from all UDs and aggregate them to obtain the global model. Then, the global model is broadcast to each UD for the next iteration. Due to the powerful computing capability and transmission power of the BS, the delay in this step is neglected compared with other steps. Furthermore, because of the stable energy supply in BS, the energy consumed is also ignored [20].

2.2. Imperfect Channel Model

In the IRS-assisted FL system, there are two kinds of channels, that are direct UDs-BS channel and cascaded UDs-IRS-BS channel. For cascaded channel, it is a challenge to directly obtain the IRS-related channels due to the passive features of the IRS. Thus, we focus on channel error models in the cascaded UDs-IRS-BS channel.

The composite channel from the k-th UD to BS can be rewritten as

$$h_k = h_{d,k} + \boldsymbol{g}_k \Theta \boldsymbol{h}_{r,k} = h_{d,k} + \boldsymbol{\phi}^H \boldsymbol{a}_k, \tag{6}$$

where $\boldsymbol{\phi} = [\phi_1, \phi_1, \ldots, \phi_n]^T$, $\boldsymbol{a}_k = diag(\boldsymbol{g}_k)\boldsymbol{h}_{rk}$, $\phi_n = e^{j\theta_n}$. Assuming that both direct UDs-BS channels and cascaded UDs-IRS-BS channels are imperfect, which can be represented as

$$h_{d,k} = \hat{h}_{d,k} + \varepsilon_{d,k}, \tag{7}$$

$$\boldsymbol{a}_k = \hat{\boldsymbol{a}}_k + \boldsymbol{\varepsilon}_{a,k}, \tag{8}$$

where $\hat{h}_{d,k}$ and $\hat{\boldsymbol{a}}_k$ stand for the direct UDs-BS estimated channel and cascaded UDs-IRS-BS estimated channel, respectively. $\varepsilon_{d,k}$ and $\boldsymbol{\varepsilon}_{a,k}$ are estimated error followed the CSCG distribution, i.e., $\varepsilon_{d,k} \sim CN(0, \sigma_{d,k}^2)$ and $\boldsymbol{\varepsilon}_{a,k} \sim CN(0, \sigma_{a,k}^2 \boldsymbol{I})$, where $\sigma_{d,k}^2$ and $\sigma_{a,k}^2$, respectively, represent variance of the direct channel estimation and cascaded channels estimation. Thus, the composite channel from k-th UD to BS can be further rewritten as

$$\begin{aligned} h_k &= h_{d,k} + \boldsymbol{\phi}^H \boldsymbol{a}_k \\ &= \hat{h}_{d,k} + \varepsilon_{d,k} + \boldsymbol{\phi}^H (\hat{\boldsymbol{a}}_k + \boldsymbol{\varepsilon}_{a,k}) \\ &= \hat{h}_k + \varepsilon_{h,k} \sim CN(\hat{h}_k, \sigma_{h,k}^2), \end{aligned} \tag{9}$$

where $\hat{h}_k = \hat{h}_{d,k} + \boldsymbol{\phi}^H \hat{\boldsymbol{a}}_k$ and $\varepsilon_{h,k} = \varepsilon_{d,k} + \boldsymbol{\phi}^H \boldsymbol{\varepsilon}_{a,k}$. Due to the lack of correlation of estimation errors for the different channel, we have

$$\begin{aligned} \sigma_{h,k}^2 &= E((\varepsilon_{d,k} + \boldsymbol{\phi}^H \boldsymbol{\varepsilon}_{a,k})(\varepsilon_{d,k} + \boldsymbol{\phi}^H \boldsymbol{\varepsilon}_{a,k})^H) \\ &= E(\varepsilon_{d,k} \varepsilon_{d,k}^H + \boldsymbol{\phi}^H \boldsymbol{\varepsilon}_{a,k} \boldsymbol{\varepsilon}_{a,k}^H \boldsymbol{\phi}) \\ &= \sigma_{d,k}^2 + N\sigma_{a,k}^2, \end{aligned} \tag{10}$$

2.3. Outage Probability

Outage probability is an important performance measurement, typically used to characterize the events that the instantaneous data rate is less than the target rate under imperfect CSI [22]. Defining the transmission target rate of the k-th UD as ν_k, and the outage probability can be written as

$$\begin{aligned} \Pr[R_k < \nu_k] &= \Pr[\frac{|h_k|^2 p_k}{\sigma^2} < 2^{\nu_k/b} - 1] \\ &= \Pr[\frac{|h_k|^2}{\sigma_{h,k}^2} < \frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}] \\ &= F_{X_k}(\frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}), \end{aligned} \tag{11}$$

where $\gamma_k = 2^{v_k/b} - 1$ and $X_k = \frac{|h_k|^2}{\sigma_{h,k}^2}$. It is obvious that X_k is a random variable that obeys non-central chi-squared distribution with degrees of freedom 2, and its non-central parameter is $\lambda_k = \frac{|\hat{h}_k|^2}{\sigma_{h,k}^2/2}$, i.e., $X_k \sim \chi^2(\lambda_k)$. Due to that the generalized Marcum Q-function $Q_n(a, b)$ is the complementary cumulative distribution function (CCDF) or reliability function of the normalized non-central chi-squared distribution with $2n$ degrees of freedom, the outage probability can be rewritten as [23]

$$
\Pr[\chi^2(\lambda_k) < \frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}] = F(\frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2})
$$
$$
= 1 - Q(\sqrt{\lambda_k}, \sqrt{\frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}}), \tag{12}
$$

where $Q_n(\sqrt{a}, \sqrt{b}) = \int_b^\infty \frac{1}{2}(\frac{x}{a})^{\frac{n}{2}-\frac{1}{2}} e^{-\frac{x+a}{2}} I_{n-1}(\sqrt{ax}) dx$, I_n denotes modified Bessel function of order n. In order to obtain the closed expression of the outage probability, we approximate the non-central chi-square distribution with the central chi-square distribution, which can written as [22]

$$
\Pr[\chi^2(\lambda_k) < \frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}] \approx \Pr[\chi^2(0) < \frac{\gamma_k \sigma^2 / p_k \sigma_{h,k}^2}{1 + \lambda_k/2}]
$$
$$
= 1 - Q(0, \sqrt{\frac{\gamma_k \sigma^2 / p_k \sigma_{h,k}^2}{1 + \lambda_k/2}}), \tag{13}
$$

where $Q_n(0, \sqrt{b}) = e^{-\frac{b}{2}} \sum_{m=0}^{n-1} \frac{(b/2)^m}{m!}$ when n is a positive integer. Thus, the closed expression of the outage probability can be approximately written as

$$
\Pr[\chi^2(\lambda_k) < \frac{\gamma_k \sigma^2}{p_k \sigma_{h,k}^2}] \approx 1 - \exp(-\frac{\gamma_k \sigma^2 / p_k \sigma_{h,k}^2}{2(1 + \lambda_k^2/2)}), \tag{14}
$$

3. Problem Formulation

Due to differences in the calculation and communication capabilities of UDs, we aim for minimizing FL delay under imperfect CSI by jointly optimizing the user selection vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]^T$, UDs' CPU frequency $\boldsymbol{f} = [f_1, f_2, \dots, f_K]^T$, transmit power $\boldsymbol{p} = [p_1, p_2, \dots, p_K]^T$, the IRS phase shifts vector $\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_N]$, interrupt transmission rate $\boldsymbol{v} = [v_1, v_2, \dots, v_K]^T$. Assuming that each UD uploads its local model after finishing local model computation. Moreover, the BS starts aggregation until receiving the local model from all UDs. The total delay of FL under one communication round can be expressed as $t^{lu} = \max_{\mathcal{K}} \mu_k(\tau_k^l + \tau_k^u)$. Thus, the delay minimization problem is formulated as

$$P0: \min_{\mu,\nu,p,f,\phi} \max_{k \in \mathcal{K}} \mu_k (L\frac{c_k D_k}{f_k} + \frac{d}{\nu_k}), \tag{15}$$

$$s.t. \quad L\frac{\alpha_k}{2} f_k^2 c_k D_k + \frac{dp_k}{\nu_k} \le e_{k\max}, \quad \forall k \in \mathcal{K}, \tag{16}$$

$$\Pr[R_k < \nu_k] \le \delta_0, \forall k, \tag{17}$$

$$|\phi_n|^2 = 1, \quad \forall n \in \mathcal{N}, \tag{18}$$

$$f_{\min} \le f_k \le f_{\max}, \quad \forall k \in \mathcal{K}, \tag{19}$$

$$0 \le p_k \le p_{\max}, \quad \forall k \in \mathcal{K}, \tag{20}$$

$$\sum_{k \in \mathcal{K}} \mu_k = \mu, \tag{21}$$

$$\mu_k \in \{0,1\}, \quad \forall k \in \mathcal{K}, \tag{22}$$

where Equation (16) denotes the energy constraint and $e_{k\max}$ is the maximum energy consumption of the k-th UD in FL; Equation (17) is the constraint of outage probability and δ_0 is limit of the outage probability; Equation (18) represents the range of the phase shift; Equations (19) and (20) denote the range of CPU frequency and transmit power, where $f_{\max}$ and $f_{\min}$, respectively, represent the upper and lower bounds of f_k range, and $p_{\max}$ is the maximum transmit power; Finally, Equations (21) and (22) denote the constraints of user selection vector, where μ represents the total number of UDs participating in this model training. $\mu_k = \{0,1\}, \forall k \in \mathcal{K}$, where $\mu_k = 1$ indicates that k-th UD participate in this model training, and $\mu_k = 0$, otherwise.

4. Resource Allocation Design for Delay Minimization

Problem $P0$ is non-convex due to the binary constraints, the coupling effect between variables and the min-max form of the objective function. Additionally, it is a mixed integer non-linear programming (MINLP) which is NP-hard. Thus, it is a challenge to find its solution. In order to solve the optimization problem, a low-complexity iterative algorithm is proposed.

In the problem $P0$, the objective function, Equations (16)–(18) and (22) are non-convex. Next, we solve them in turn. First, for the Equation (22), it is a binary constraint, which can be equivalent to

$$0 \le \mu_k \le 1, \forall k \in \mathcal{K}, \tag{23}$$

$$\mu_k(\mu_k - 1) \ge 0, \forall k \in \mathcal{K}. \tag{24}$$

Note that optimization variable u_k is a continuous value between 0 and 1. However, Equation (24) is a non-convex. In order to deal with the Equation (24), we reformulate problem $P0$ as

$$P1: \min_{\mu,\nu,p,f,\phi} \max_{k \in \mathcal{K}} \mu_k (L\frac{c_k D_k}{f_k} + \frac{d}{\nu_k}) - \eta \sum_{k \in \mathcal{K}} \mu_k(\mu_k - 1), \tag{25}$$

$$s.t. \quad (16) - (21), (23)$$

where η is usually a constant and represents a penalty factor for penalizing the objective function that any μ_k is not equal to 0 or 1 [24]. In order to solve the min–max form of

the objective function, the variable T is introduced. Therefore, the problem $P1$ can be reformulated as

$$P2 : \min_{\mu,\nu,p,f,\phi,T} T, \tag{26}$$

$$s.t. \ \mu_k(L\frac{c_k D_k}{f_k} + \frac{z}{\nu_k}) - \eta \sum_{k \in \mathcal{K}} \mu_k(\mu_k - 1) \leq T, \tag{27}$$

$$(16) - (21), (23)$$

Note that although we introduce the variable T to solve the min–max form of the objective function, it also brings a new non-convex Equation (27). Due to non-convex Equations (16)–(18) and (27), the problem $P2$ is still difficult to solve.

For Equation (16), its non-convexity comes from the form x/y of the second term. By replacing the variable p_k with a new variable $q_k = \sqrt{p_k}$, the Equation (16) can be rewritten as

$$L\frac{\alpha_k}{2}f_k{}^2 c_k D_k + \frac{dq_k^2}{\nu_k} \leq e_{k\,\max}. \tag{28}$$

It is obvious that the Equation (28) is convex. Thus, the optimization problem $P2$ can be rewritten as

$$P3 : \min_{\mu,\nu,q,f,\phi,T} T, \tag{29}$$

$$s.t. \ \sigma_{h,k}^2(-\frac{1}{\log(1-\delta_0)}\frac{\hat{\gamma}_k\sigma^2}{q_k^2\sigma_{h,k}^2} - 2) \leq \left|\hat{h}_k\right|^2, \tag{30}$$

$$(18) - (21), (23), (27), (28)$$

Note that the problem $P3$ is a non-convex quadratically constrained quadratic program (QCQP). By introducing variables κ and Λ, the problem $P3$ can be rewritten as

$$P4 : \min_{\mu,\nu,q,f,\Lambda,T} T, \tag{31}$$

$$s.t. \ \sigma_{h,k}^2(-\frac{1}{\log(1-\delta_0)}\frac{\hat{\gamma}_k\sigma^2}{q_k^2\sigma_{h,k}^2} - 2) \leq \hat{h}_{d,k}^H\hat{h}_{d,k} + Tr(\hat{\Xi}_k\Lambda), \tag{32}$$

$$\Lambda_{n,n} = 1, \ \forall n \in \{1, 2, \ldots, N+1\}, \tag{33}$$

$$\text{Rank}(\Lambda) = 1, \tag{34}$$

$$(19) - (21), (23), (27), (28)$$

where $\bar{\phi} = \begin{bmatrix} \phi \\ \kappa \end{bmatrix}$, $\Lambda = \bar{\phi}\bar{\phi}^H$, $\hat{\Xi}_k = \begin{bmatrix} \hat{a}_k\hat{a}_k^H & \hat{a}_k\hat{h}_{dk}^H \\ \hat{h}_{dk}\hat{a}_k^H & 0 \end{bmatrix}$, and $\text{Rank}(\Lambda)$ represents the rank of matrix Λ. Due to non-convex Equations (27), (32) and (34), problem $P4$ still cannot be solved directly. For the Equation (34), SDR can be invoked to relax it. Furthermore, by introducing the variable ζ_k, the Equation (32) can be rewritten as

$$\sigma_{h,k}^2(\frac{1}{\zeta_k} - 2) \leq h_{d,k}^H h_{d,k} + Tr(\Xi_k\Lambda), \ \forall k \in \mathcal{K}, \tag{35}$$

$$-\frac{1}{\log(1-\delta_0)}\frac{\hat{\gamma}_k\sigma^2}{\sigma_{h,k}^2} \leq \frac{q_k^2}{\zeta_k}, \ \forall k \in \mathcal{K}. \tag{36}$$

It is obvious that the Equation (32) is converted into Equation (35) by introducing variables, which is convex. However, it also brings a non-convex Equation (36). Next, the CCCP is invoked to solve this issue, which main idea is that when the original problem

is difficult to optimize, the algorithm does not directly seek its local optimal solution, but looks for its approximation problem to solve iteratively [25].

For a given point $(\boldsymbol{q}^{(n)}, \boldsymbol{\zeta}^{(n)})$ at the n-th iteration, the function $f_k(q_k, \zeta_k) = q_k^2/\zeta_k$ has a lower bound, that is

$$\tilde{f}_k(q_k, \zeta_k) = 2\beta_k^{(n)} q_k - (\beta_k^{(n)})^2 \zeta_k, \tag{37}$$

where $\beta_k^{(n)} = q_k^{(n)}/\zeta_k^{(n)}$. Thus, Equation (36) can be rewritten as

$$-\frac{1}{\log(1-\delta_0)} \frac{\hat{\gamma}_k \sigma^2}{\sigma_{h,k}^2} \le 2\beta_k^{(n)} q_k - (\beta_k^{(n)})^2 \zeta_k, \ \forall k \in \mathcal{K}. \tag{38}$$

Now, Equation (38) is convex. For the convenience of derivations, we rewrite Equation (27) as

$$g_{1,k}(\mu_k, f_k, \nu_k) + g_{2,k}(\mu_k, f_k, \nu_k) + g_{3,k}(\mu_k) \le T, \forall k \in \mathcal{K}, \tag{39}$$

where

$$g_{1,k}(\mu_k, f_k, \nu_k) = \frac{1}{4}\left(\mu_k + (L\frac{c_k D_k}{f_k} + \frac{z}{\nu_k})\right)^2, \tag{40}$$

$$g_{2,k}(\mu_k, f_k, \nu_k) = -\frac{1}{4}\left(\mu_k - (L\frac{c_k D_k}{f_k} + \frac{z}{\nu_k})\right)^2, \tag{41}$$

$$g_{3,k}(\mu_k) = -\eta \sum_{k \in \mathcal{K}} \mu_k(\mu_k - 1). \tag{42}$$

Note that $g_{1,k}(\mu_k, f_k, \nu_k)$ is convex, while $g_{2,k}(\mu_k, f_k, \nu_k)$, and $g_{3,k}(\mu_k)$ are non-convex. For a given point $(\boldsymbol{\mu}^{(n)}, \boldsymbol{f}^{(n)}, \boldsymbol{\nu}^{(n)})$ at the n-th iteration, both $g_{2,k}(\mu_k, f_k, \nu_k)$ and $g_{3,k}(\mu_k)$ have an upper bound, which are given as

$$
\begin{aligned}
\tilde{g}_{2,k}(\mu_k, f_k, \nu_k) = {} & -\frac{1}{4}\left(\mu_k^{(n)} - (L\frac{c_k D_k}{f_k^{(n)}} + \frac{z}{\nu_k^{(n)}})\right)^2 + \frac{1}{2}\mu_k^{(n)}\left(\frac{z\nu_k}{(\nu_k^{(n)})^2} + \frac{z}{\nu_k} - \frac{2z}{\nu_k^{(n)}}\right) \\
& -\frac{1}{2}\left(\mu_k^{(n)} - (L\frac{c_k D_k}{f_k^{(n)}} + \frac{z}{\nu_k^{(n)}})\right)\left((\mu_k - \mu_k^{(n)}) + \frac{Lc_k D_k}{(f_k^{(n)})^2}(f_k - f_k^{(n)}) + \frac{z}{(\nu_k^{(n)})^2}(\nu_k - \nu_k^{(n)})\right) \\
& +\frac{1}{2}\mu_k^{(n)}\left(\frac{Lc_k D_k f_k}{(f_k^{(n)})^2} + \frac{Lc_k D_k}{f_k} - \frac{2Lc_k D_k}{f_k^{(n)}}\right).
\end{aligned}
\tag{43}
$$

$$\tilde{g}_{3,k}(\mu_k) = -\eta \sum_{k \in \mathcal{K}} \left(\mu_k^{(n)}(\mu_k^{(n)} - 1) + (2\mu_k^{(n)} - 1)(\mu_k - \mu_k^{(n)})\right). \tag{44}$$

As such, Equation (27) can be approximated by

$$g_{1,k}(\mu_k, f_k, \nu_k) + \tilde{g}_{2,k}(\mu_k, f_k, \nu_k) + \tilde{g}_{3,k}(\mu_k) \le T, \forall k \in \mathcal{K}, \tag{45}$$

Now, Equation (45) is also convex. For a given point $(\boldsymbol{q}^{(n)}, \boldsymbol{\zeta}^{(n)}, \boldsymbol{\mu}^{(n)}, \boldsymbol{f}^{(n)}, \boldsymbol{\nu}^{(n)})$ at the $(n+1)$-th iteration, the problem $P4$ can be reformulated as

$$P5: \min_{\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{q}, \boldsymbol{f}, \boldsymbol{\Lambda}, T} T, \tag{46}$$

$$s.t. \ (19) - (21), (23), (28), (33), (35), (38), (45)$$

It is clear that the problem $P5$ is a convex semi-definite program (SDP). Finally, we solve it directly by using standard optimization packages, such as CVX [26]. However,

note that *P*5 is a relaxed problem, which usually cannot lead to a rank-one solution. Thus, Gaussian randomization is invoked to construct a rank-one solution [27]. For the optimal μ_k obtained by solving problem *P*5, it is close to 0 or 1 due to the punishment item. Rounding operation is run to obtain 0 or 1. The algorithm to find a feasible solution for problem *P*5 is summarized in Algorithm 1.

Algorithm 1: Local optimal iterative algorithm.

---

1. Initialization
Initialize the a feasible initial point $(q^{(0)}, \zeta^{(0)}, \mu^{(0)}, f^{(0)}, \nu^{(0)})$.
Initialize the maximum number of iterations n_{max} and penalty factor η
Initialize iteration index $n = 1$ and accuracy index $\xi^{(0)} = 1$.
2. Joint optimization of q, ζ, μ, f, ν and Λ
repeat
- Calculate $\beta^{(0)}$ by $\beta_k = q_k^{(n)} / \zeta_k^{(n)}$, $\forall k \in \mathcal{K}$
- Calculate $(q^{(n)}, \zeta^{(n)}, \mu^{(n)}, f^{(n)}, \nu^{(n)}, \Lambda^{(n)})$ by solving problem
  *P*5 for a given point $(q^{(n-1)}, \zeta^{(n-1)}, \mu^{(n-1)}, f^{(n-1)}, \nu^{(n-1)})$
- Calculate accuracy $\xi^{(n)} = \frac{\left| t^{lu(n)} - t^{lu(n-1)} \right|}{t^{lu(n)}}$
- $n = n + 1$
until $\xi^{(t)} \leq \xi$ **or** $n = n_{max}$
3. Output optimal μ^*, p^*, f^*, ν^* and θ^*
find θ^* from $\Lambda^{(n)}$ by SVD and Gaussian randomization.
$\mu^* = \mu^{(n)}$, $p^* = \left| q^{(n)} \right|^2$, $f^* = f^{(n)}$, $\nu^* = \nu^{(n)}$

Based on the Algorithm 1 description, the complexity analysis of the proposed algorithm is performed. The problem *P*5 is a SDR problem, which can be solved by using the interior point method. Therefore, we analyzed the complexity of the proposed algorithm based on this method. At each iteration, the worst case complexity of solving problem *P*5 is on the order of $(N+1)^6 + K(N+1)^2 + 4K+1$, and the complexity of finding variable β is negligible. Therefore, the overall complexity of the Algorithm 1 can be evaluated by $O(T_{CCCP}((N+1)^6 + K(N+1)^2 + 4K+1))$, where T_{CCCP} represents the required iterations number to achieve the target accuracy ξ [28,29].

5. Numerical Results

In this section, we evaluate the performance of proposed scheme in the IRS-assisted FL systems. In the simulation scenario, the BS and IRS are located at (300 m, 0 m) and (30 m, 30 m), respectively. In total, 20 UDs are randomly distributed in a circular area centered at (0 m, 0 m) with radius of 20 m. For channel models, both the large-scale fading and small-scale fading are considered. The large-scale fading model is expressed as $PL = PL_0 - 10\alpha_{PL}\log_{10}(x)$, where α_{PL} and x are the path loss exponent and the link distance in meters, respectively. PL_0 denotes the path loss at the reference distance of 1 m, which is set as -30 dB. The pass loth exponent of the direct UDs-BS channel and the IRS-related channel are given as $\alpha_{PL,BU} = 3.75$ and $\alpha_{PL,IRS} = 2.2$, respectively. The small-scale fading of the direct UDs-BS channel and the IRS-related channel are considered as Rayleigh fading and Rician fading, respectively. Assuming that all UEs have the same energy consumption limit, i.e., $e_{k,max} = E$. For the statistical CSI error model, the variance of $\varepsilon_{d,k}$ and $\varepsilon_{a,k}$ are, respectively, defined as $\delta_d^2 \left\| \hat{h}_{d,k} \right\|_2^2$ and $\delta_a^2 \left\| \hat{h}_{a,k} \right\|_2^2$, where $\delta_d = 0.02$ and $\delta_a = 0.01$ denote the normalized CSI error. In order to guarantee the communication quality, the outage probability is set to $\delta_0 = 0.05$. In the simulation, we set $L = 5$, $c = 12{,}000$ cycle/sample, $f_k = [4, 6] \times 10^8$ cycle/s, $\alpha = 10^{-27}$, $b = 200$ KHZ, $\sigma^2 = -174$ dBm/Hz and $d = 25{,}000$ nats.

In order to describe the advantages of our algorithm, the FL algorithm without IRS is compared, where the IRS-related channels is set to 0, and edge computation and communication resource allocation is optimized.

Figure 2 describes the delay versus the maximum transmit power with different energy consumption limits. It is shown that the delay of the FL algorithm decreases as the maximum power increases. This is because that the low transmit power leads to a narrow feasible set of the optimization problem and the feasible set increases as the maximum power increases. Compared with "FL algorithm without IRS", the delay decrease in "FL algorithm with IRS" is up to 32%. In addition, we also found that the scheme with more relaxed energy consumption limits consumes less time.
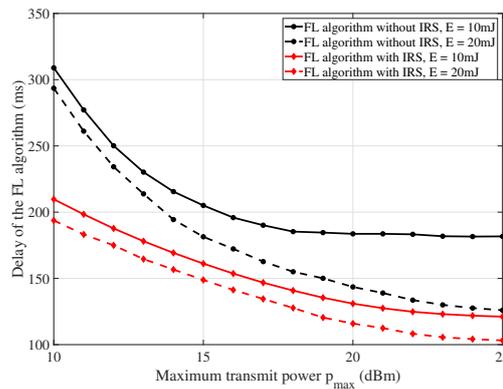
Figure 2. The delay versus the maximum transmit power $p_{\max}$, where $N = 30$ and $\delta_0 = 0.05$.

Figure 3 presents the relationship between the delay and the outage probability. It is obvious that the delay of the FL algorithm decreases as the outage probability increases. This is because that the quality of service is sacrificed in order to increase the communication rate. As the communication rate increases, the energy consumption becomes less and less. Therefore, the delay gap between the algorithm with $E = 10$ mJ and $E = 20$ mJ becomes less.
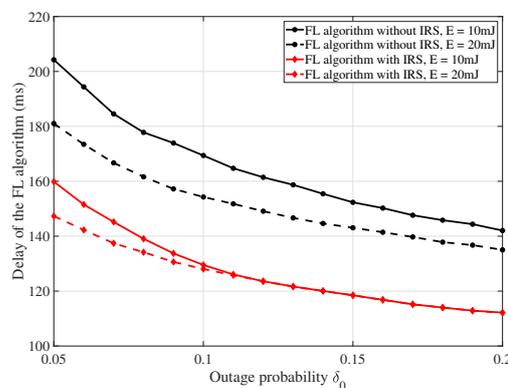

Figure 3. The delay versus the outage probability δ_0, where $N = 30$ and $p_{\max} = 15$.

Figure 4 presents the relationship between the delay and the number of the IRS elements with different energy consumption limits. It is obvious that there is a performance gap between the "FL algorithm with IRS" and "FL algorithm without IRS", and it increases with the number of the IRS elements increase. This is because that the IRS lead to a reflection-based beamforming gain, which increases as the number of IRS elements increases. Compared with "FL algorithm without IRS", the time reduction by "FL algorithm with IRS" is up to 48% with $E = 10$ mJ and $N = 100$.
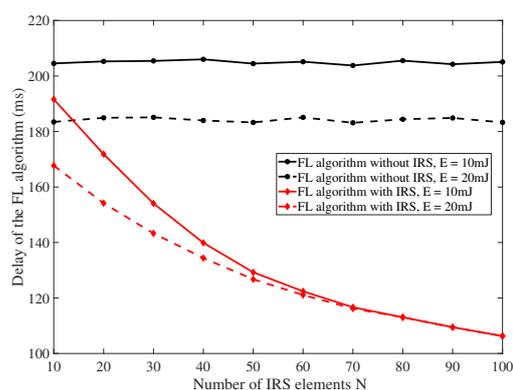
Figure 4. The delay versus the number of IRS elements N, where $\delta_0 = 0.05$ and $p_{\max} = 15$.

6. Conclusions

In this paper, we investigate the delay minimum problem in FL systems with IRS under imperfect CSI, and outage probability is introduced to characterize the impact of imperfect CSI. By using central chi-square distribution approximations, we transform the outage probability into tractable form. Then, the problem of minimizing delay was formulated in the IRS assisted FL framework, and it is a MINLP problem. Next, a low-complexity algorithm was developed base on the CCCP and SDR to solve the intractable problem. In various simulation environments, the benefits of the proposed algorithm are evaluated. Numerical results show that the effectiveness of proposed algorithm in reducing the training time of the FL system base on imperfect CSI.

References

1. Gündüz, D.; Kurka, D.B.; Jankowski, M.; Amiri, M.M.; Ozfatura, E.; Sreekumar, S. Communicate to Learn at the Edge. *IEEE Commun. Mag.* **2020**, *58*, 14–19. [CrossRef]
2. Nguyen, D.C.; Cheng, P.; Ding, M.; Lopez-Perez, D.; Pathirana, P.N.; Li, J.; Seneviratne, A.; Li, Y.; Poor, H.V. Enabling AI in Future Wireless Networks: A Data Life Cycle Perspective. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 553–595. [CrossRef]
3. Nguyen, M.N.H.; Tran, N.H.; Tun, Y.K.; Han, Z.; Hong, C.S. Toward Multiple Federated Learning Services Resource Sharing in Mobile Edge Networks. *IEEE Trans. Mob. Comput.* **2021**,
4. Zaw, C.W.; Hong, C.S. Loss and Energy Tradeoff in Multi-access Edge Computing Enabled Federated Learning. In Proceedings of the 2021 International Conference on Information Networking (ICOIN), Jeju Island, Korea, 13–16 January 2021; pp. 597–602. [CrossRef]
5. Zaw, C.W.; Pandey, S.R.; Kim, K.; Hong, C.S. Energy-Aware Resource Management for Federated Learning in Multi-Access Edge Computing Systems. *IEEE Access* **2021**, *9*, 34938–34950. [CrossRef]
6. Pase, F.; Giordani, M.; Zorzi, M. On the Convergence Time of Federated Learning Over Wireless Networks Under Imperfect CSI. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–7. [CrossRef]
7. Wadu, M.M.; Samarakoon, S.; Bennis, M. Federated Learning under Channel Uncertainty: Joint Client Scheduling and Resource Allocation. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea, 25–28 May 2020; pp. 1–6. [CrossRef]
8. Wu, Q.; Zhang, R. Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network. *IEEE Commun. Mag.* **2020**, *58*, 106–112. [CrossRef]

9. Chu, Z.; Xiao, P.; Shojafar, M.; Mi, D.; Mao, J.; Hao, W. Intelligent Reflecting Surface Assisted Mobile Edge Computing for Internet of Things. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 619–623. [CrossRef]

10. Zheng, J.; Ni, W.; Tian, H.; Wang, Y. QoS-Constrained Federated Learning Empowered by Intelligent Reflecting Surface. In Proceedings of the 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, 13–16 September 2021; pp. 947–952. [CrossRef]

11. Liu, H.; Yuan, X.; Zhang, Y.J.A. Reconfigurable Intelligent Surface Enabled Federated Learning: A Unified Communication-Learning Design Approach. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 7595–7609. [CrossRef]

12. Ni, W.; Liu, Y.; Yang, Z.; Tian, H. Over-the-Air Federated Learning and Non-Orthogonal Multiple Access Unified by Reconfigurable Intelligent Surface. In Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, BC, Canada, 10–13 May 2021; pp. 1–6. [CrossRef]

13. Taha, A.; Alrabeiah, M.; Alkhateeb, A. Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning. *IEEE Access* **2021**, *9*, 44304–44321. [CrossRef]

14. Mishra, D.; Johansson, H. Channel Estimation and Low-complexity Beamforming Design for Passive Intelligent Surface Assisted MISO Wireless Energy Transfer. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 4659–4663. [CrossRef]

15. Nadeem, Q.U.A.; Alwazani, H.; Kammoun, A.; Chaaban, A.; Debbah, M.; Alouini, M.S. Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design. *IEEE Open J. Commun. Soc.* **2020**, *1*, 661–680. [CrossRef]

16. Zheng, B.; Zhang, R. Intelligent Reflecting Surface-Enhanced OFDM: Channel Estimation and Reflection Optimization. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 518–522. [CrossRef]

17. Hong, S.; Pan, C.; Ren, H.; Wang, K.; Chai, K.K.; Nallanathan, A. Robust Transmission Design for Intelligent Reflecting Surface-Aided Secure Communication Systems With Imperfect Cascaded CSI. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 2487–2501. [CrossRef]

18. Zhou, G.; Pan, C.; Ren, H.; Wang, K.; Nallanathan, A. A Framework of Robust Transmission Design for IRS-Aided MISO Communications With Imperfect Cascaded Channels. *IEEE Trans. Signal Process.* **2020**, *68*, 5092–5106. [CrossRef]

19. Ren, J.; Yu, G.; Ding, G. Accelerating DNN Training in Wireless Federated Edge Learning Systems. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 219–232. [CrossRef]

20. Yang, Z.; Chen, M.; Saad, W.; Hong, C.S.; Shikh-Bahaei, M. Energy Efficient Federated Learning Over Wireless Communication Networks. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 1935–1949. [CrossRef]

21. Luo, S.; Chen, X.; Wu, Q.; Zhou, Z.; Yu, S. HFEL: Joint Edge Association and Resource Allocation for Cost-Efficient Hierarchical Federated Edge Learning. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 6535–6548. [CrossRef]

22. Fang, F.; Wang, K.; Ding, Z.; Leung, V.C.M. Energy-Efficient Resource Allocation for NOMA-MEC Networks With Imperfect CSI. *IEEE Trans. Commun.* **2021**, *69*, 3436–3449. [CrossRef]

23. Sun, Y.; Baricz, Á.; Zhou, S. On the Monotonicity, Log-Concavity, and Tight Bounds of the Generalized Marcum and Nuttall *Q*-Functions. *IEEE Trans. Inf. Theory* **2010**, *56*, 1166–1186. [CrossRef]

24. Sun, Y.; Ng, D.W.K.; Ding, Z.; Schober, R. Optimal Joint Power and Subcarrier Allocation for MC-NOMA Systems. In Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; pp. 1–6. [CrossRef]

25. Cheng, Y.; Pesavento, M. Joint Optimization of Source Power Allocation and Distributed Relay Beamforming in Multiuser Peer-to-Peer Relay Networks. *IEEE Trans. Signal Process.* **2012**, *60*, 2962–2973. [CrossRef]

26. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, Version 2.1. 2014. Available online: http://cvxr.com/cvx (accessed on 3 November 2021).

27. Wu, Q.; Zhang, R. Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5394–5409. [CrossRef]

28. Christopoulos, D.; Chatzinotas, S.; Ottersten, B. Weighted Fair Multicast Multigroup Beamforming Under Per-antenna Power Constraints. *IEEE Trans. Signal Process.* **2014**, *62*, 5132–5142. [CrossRef]

29. Nemirovskii, A.S.; Nesterov, Y.E. *Interior-Point Polynomial Algorithms in Convex Programming*; SIAM: Philadelphia, PA, USA, 1994.