Low-Latency Federated Learning Over Wireless Channels With Differential Privacy

Kang Wei, *Graduate Student Member, IEEE*, Jun Li, *Senior Member, IEEE*, Chuan Ma, *Member, IEEE*,
Ming Ding, *Senior Member, IEEE*, Cailian Chen, *Member, IEEE*, Shi Jin, *Senior Member, IEEE*,
Zhu Han, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

Abstract—In federated learning (FL), model training is distributed over clients and local models are aggregated by a central server. The performance of uploaded models in such situations can vary widely due to imbalanced data distributions, potential demands on privacy protections, and quality of transmissions. In this paper, we aim to minimize FL training delay over wireless channels, constrained by overall training performance as well as each client's differential privacy (DP) requirement. We solve this problem in a multi-agent multi-armed bandit (MAMAB) framework to deal with the situation where there are multiple clients confronting different unknown transmission environments, e.g., channel fading and interference. Specifically, we first transform long-term constraints on both training performance and each client's DP into a virtual queue based on the Lyapunov drift technique. Then, we convert the MAMAB to a max-min bipartite matching problem at each communication round, by estimating rewards with the upper confidence bound (UCB) approach. More importantly, we propose two efficient solutions to this matching problem, i.e., a modified Hungarian algorithm and greedy matching with a better alternative (GMBA), of which the former can achieve the optimal solution with high complexity while the latter approaches a better trade-off by enabling verified low-complexity with little performance loss. In addition, we develop an upper bound on the expected regret of this MAMAB based FL framework, which shows a linear growth over the logarithm of communication rounds, justifying its theoretical feasibility. Extensive experimental results are conducted to validate the effectiveness of our proposed algorithms, and the impacts of various parameters on the FL performance over wireless edge networks are also discussed.

Index Terms—Federated learning, differential privacy, multi-agent multi-armed bandit, max-min bipartite matching.

Kang Wei and Jun Li are with the School of Electrical and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: kang.wei@njust.edu.cn; jun.li@njust.edu.cn).

Chuan Ma is with the School of Electrical and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China (e-mail: chuan.ma@njust.edu.cn).

Ming Ding is with the Data61, CSIRO, Sydney, NSW 2015, Australia (e-mail: ming.ding@data61.csiro.au).

Cailian Chen is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China (e-mail: cailianchen@sjtu.edu.cn).

Shi Jin is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: jinshi@seu.edu.cn).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: hanzhan22@gmail.com).

H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2021.3126052.

Digital Object Identifier 10.1109/JSAC.2021.3126052

I. Introduction

WITH the development of the Internet-of-Things (IoT), data from intelligent devices is proliferating at unprecedented rates [1]–[3]. Meanwhile, machine learning (ML), which relies heavily on such data, is revolutionizing many aspects of our lives [4]. However, conventional centralized ML [5] offers little scalability for efficiently processing data. To tackle this challenge, several distributed ML architectures have been proposed [6]. Moreover, data privacy and confidentiality are of increasing concern as exchanged data often contain clients' sensitive information in distributed ML settings. In this light, federated learning (FL) has been proposed, which allows decoupling of data provision at clients and machine learning model aggregation at a central server [7]–[9]. In FL, all clients with the same data structure collaboratively learn a shared model with the help of a central server. Owing to the local training, FL does not require clients to upload their private data, thereby effectively reducing transmission overhead as well as helping preserve clients' privacy. As such, FL is applicable to a variety of scenarios where data are either high-cost or sensitive to be transmitted to the server, e.g., health-care records, private images, personally identifiable information, etc. [10].

Since many communication rounds are often required by FL to reach a desired model accuracy, especially when the number of participating clients in the training process is comparably large [5], the latency caused by unreliable wireless transmissions and unequal local computations at clients can be a bottleneck in wireless FL systems. Despite the computational

efficiency, the long communication distance between smart devices and the remote cloud inevitably introduces a high transmission latency, resulting in an unsatisfactory user quality of experience (QoE), especially for real-time delay-sensitive applications [11]–[13]. The waveform-superposition property of the wireless medium has been exploited for wireless FL systems [14] to overcome the communication bottleneck. However, in many FL systems, since each client will upload its local model immediately after finishing local training, i.e., asynchronous FL, it may be difficult to utilize the super-position property of the wireless channel. Therefore, proper resource scheduling management (e.g., spectrum) plays a crucial role in improving wireless network efficiency. Especially in the envisioned sixth generation (6G) networks [15], [16], fully utilizing spectrum is expected to guarantee system performance in terms of throughput, delay, and so on.

From the perspective of resource allocation, many recent studies have focused on efficient local computation and communications between clients and edge servers in FL-supported networks [17]–[20]. In order to characterize the performance of FL in wireless networks, an analytical model [17] in terms of FL convergence rate has been developed to evaluate the effectiveness of three different client scheduling policies, i.e., random scheduling, round-robin, and proportional fair. The work in [18] has formulated FL over a wireless network as an optimization problem and a control algorithm has been developed to minimize the loss function based on the convergence bound of distributed gradient descent. Via constructing the connection between wireless resource allocation and FL learning performance, the work in [19] and [21] adjusted the user selection and power allocation to minimize the FL loss function. By involving the fairness constraint for each client associated with the sizes of the local datasets [22]–[24], the scheduling policy becomes more efficient in terms of learning accuracy, but the sizes of the local datasets cannot represent the data distribution of the whole.

Beyond resource allocation, as a well-established mechanism for privacy-preservation of local models, differential privacy (DP) has been proposed for FL systems. Some works on DP in FL focus on the impact of DP mechanisms on learning performance under the assumption that the edge server is semi-honest and that communication is reliable (without noise and interference) and unconstrained, aiming to achieve a better tradeoff between privacy and performance [25]–[27]. Other works [28], [29] focus on uncoded transmission of gradients using either orthogonal or non-orthogonal protocols, and analytically demonstrate that for these transmission schemes, privacy may be obtained "for free" in the sense that enforcing a DP constraint causes no performance loss with respect to a non-private design as long as the signal-to-noise ratio (SNR) is sufficiently low. Unfortunately, it is not realistic or in general to apply the uncoded transmission in modern communication systems.

The above works do not address imbalances in clients' characteristics, such as computing and data resources, and in this paper we consider such issues in the design of a channel assignment protocol for wireless FL systems. In particular, we propose a multi-agent multi-armed bandit (MAMAB)

based policy jointly minimizing the training time and the learning performance for wireless FL systems. Our main contributions are summarized as follows.

- We investigate the problem of delay minimization for FL over wireless communication networks considering potentially different privacy protections and data imbalance. We solve this problem by a MAMAB framework with constraints on the overall training performance and each client's DP requirement.
- We transform the long-term constraint of the MAMAB problem on the training performance and DP requirements into a virtual queue based on the Lyapunov technique, and then use the upper confidence bound (UCB) method to estimate rewards in MAMAB. Based on this estimate, we can schedule clients at each communication round by solving a max-min bipartite matching problem with two efficient solutions. The first solution can achieve the optimal matching result but a high complexity and the second one can achieve a lower complexity with little performance loss.
- We perform an analysis of the feasibility of the proposed MAMAB based FL framework. We show that the optimality gap of the proposed agent-based collaborative MAMAB framework is given by $O(V^2 N \log T)$, where V^2 is attributed to the impact of the client participation ratio constraint, and N and $\log T$ are, respectively, attributed to the cost of communication and computation dynamics in the learning process. In particular, N and T are denoted as the number of available channels and the number of communication rounds, respectively.
- Extensive experimental results are provided to demonstrate the effectiveness of our proposed algorithm in terms of feasibility. Moreover, we show that our proposed algorithms can fully exploit the interplay between communication and computation to outperform the baselines.

The rest of this paper is organized as follows. In Section II, we give a system moder for wireless FL with imbalanced resources among clients. Then, we formulate the joint channel assignment problem as an optimization problem whose goal is to minimize the time delay and propose the MAMAB based algorithm in Section IV. In Section V, we propose solutions to the max-min weighted bipartite matching problem. We perform an analysis of the feasibility of the proposed algorithm in VI. Experimental results are described in Section VII. Finally, conclusions are drawn in Section VIII.

II. System Model

Fig. 1 shows a dynamic FL system consisting of a base station (BS) with a central server, U clients and N available channels, to train a global model by fully utilizing all clients' data during T communication rounds. In each communication round, the selected clients can communicate with the BS through assigned channels, which are subject to interference caused by wireless equipment in other service areas. We can note that each chosen client i consists of a local dataset $\mathcal{D}_i$, $\forall i \in \mathcal{U}$, $\mathcal{U} = \{1, 2, \ldots, U\}$. These datasets are independent from client to client. The training process of such an
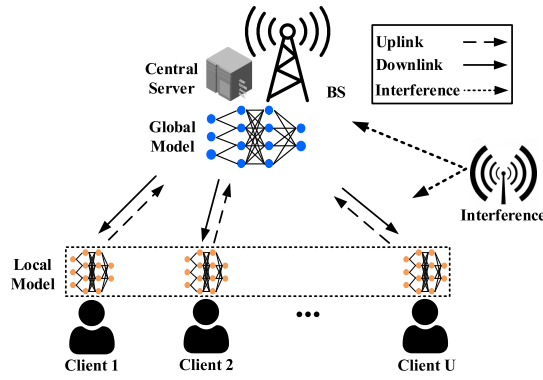
Fig. 1. The framework of wireless FL consists of multiple clients and a BS with multiple channels, where it is non-trivial to characterize the available computing capability for each client and the dynamic wireless channel gains caused by interference.

FL system can be summarized as follows: 1) The central server broadcasts the global model and training information (e.g., channel assignment) to all clients. 2) Each client updates their respective models with the global model, and tests the performance of the updated models. 3) Each client performs the local training procedure based on its local dataset, and sends the trained local model parameters to the central server. 4) The server performs aggregation over the uploaded models from clients, and begins the next communication round until the terminal condition is reached. The central server aggregates the models received from the clients, which can be given as

$$\boldsymbol{w}(t) = \sum_{i \in \mathcal{S}(t)} p_i \boldsymbol{w}_i(t), \qquad (1)$$

where $\mathcal{S}(t)$ is the set of clients that successfully upload their local models, $p_i = |\mathcal{D}_i| / \sum_{i \in \mathcal{S}(t)} |\mathcal{D}_i|$, $\boldsymbol{w}(t)$ is the global model at the t-th communication round, and $\boldsymbol{w}_i(t)$ is the uploaded model of the i-th client at the t-th communication round. The goal of FL is to learn a model over data that resides at the U associated clients. Formally, this FL task can be expressed as

$$\boldsymbol{w}^\star = \arg\min_{\boldsymbol{w}} F(\boldsymbol{w}), \qquad (2)$$

where $F(\boldsymbol{w}) = \sum_{i \in \mathcal{U}} p_i F_i(\boldsymbol{w})$ and $F_i(\cdot)$ is the local objective function of the i-th client. From this FL training procedure, we observe that all clients with the same data structure collaboratively learn a ML model with the help of a central server. After a sufficient number of local training and update exchanges between the server and its associated clients, the solution to the optimization problem is able to converge to the global optimal learning model.

A. Transmission Model

Due to the core network connection, we assume that the overall training delay does not include the transmission time between the BS and the central server, i.e., the delay in backhauls is ignored. We can note that the interference caused by the wireless equipments that are located in other service areas in the uplink or downlink channel is governed by the applied channel and the location of the receiver. Thus, in this FL system, the uplink data rate of the i-th client transmitting its local model to the BS via the j-channel at the t-th communication round can be expressed as

$$R_{i,j}^{\mathrm{U}}(t) = B^{\mathrm{U}} \log_2 \left\{ 1 + \frac{P^{\mathrm{U}} h_{i,j}(t)}{I_{i,j}^{\mathrm{U}}(t) + \sigma^2} \right\}, \qquad (3)$$

where P^{U} is the transmit power for clients, $h_{i,j}(t)$ is the average channel gain from the i-th client to the BS via the j-th channel, $I_{i,j}^{\mathrm{U}}(t)$ is the co-channel interference caused by the wireless equipments that are located in other service areas, σ^2 is the noise power spectral density, and B^{U} is the bandwidth of the uplink channel.

Similarly, for the downlink, when the central server transmits the global model parameters to the i-th client, the data rate of the BS via the j-th channel at the t-th communication round is given by

$$R_{i,j}^{\mathrm{D}}(t) = B^{\mathrm{D}} \log_2 \left\{ 1 + \frac{P^{\mathrm{D}} h_{i,j}(t)}{I_{i,j}^{\mathrm{D}}(t) + \sigma^2} \right\}, \qquad (4)$$

where P^{D} is the transmit power for the BS, B^{D} is the bandwidth of the downlink channel and $I_{i,j}^{\mathrm{D}}(t)$ is the interference caused by other wireless equipments not participating in the FL training process. Given the uplink data rate in (3) and the downlink data rate in (4), we can derive the uplink and downlink transmission delays, respectively. Since the number of elements in the local FL model $\boldsymbol{w}_i(t)$ is equal to that of the global FL model $\boldsymbol{w}(t)$. The transmission delay between client i and the j-th channel over both uplink and downlink at the t-th communication round can be, respectively, computed as

$$d_{i,j}^{\mathrm{U}}(t) = \frac{Z(\boldsymbol{w}_i(t))}{R_{i,j}^{\mathrm{U}}(t)}, \ d_{i,j}^{\mathrm{D}}(t) = \frac{Z(\boldsymbol{w}(t))}{R_{i,j}^{\mathrm{D}}(t)}, \qquad (5)$$

where function $Z(\boldsymbol{w})$ is the data size of $\boldsymbol{w}$ which is defined as the number of bits that the client or the BS require to transmit vector $\boldsymbol{w}$ over wireless links. To this end, the scheduling policy plays a crucial role of assigning the resource-limited radio available channels to the appropriate clients.

B. Computation Model

In the computation model, we assume each client is equipped with an CPU for the training task. We measure the computation capacity of each client by its CPU frequency, denoted by $f_i(t)$ (in CPU cycle/s), which is changed at the t-th communication round. Moreover, let Φ_i denote the number of CPU cycles performing the forward-backward propagation algorithm with one data for the i-th client. Due to the fact that CPU operates in the serial mode, the local gradient calculation latency is given by $d_i^{\mathrm{L}}(t) = \tau |\mathcal{D}_i| \Phi_i / f_i(t)$, where τ is the local training iterations. We can note that there exists a tradeoff between the learning performance and local training delay.

C. Differential Privacy

The DP mechanism with parameters ϵ and δ provides a strong criterion for the privacy preservation of distributed data processing systems. Here, $\epsilon > 0$ is the distinguishable bound on all outputs on neighboring datasets x, x' in a database $\mathcal{X}$, and δ represents the probability of the event that the ratio of the

probabilities for two adjacent datasets x, x' cannot be bounded by e^ϵ after adding a privacy-preserving mechanism. With an arbitrarily given δ, a larger ϵ gives a clearer distinguishability of neighboring datasets and thus a higher risk of privacy violation. Now, we will formally define DP as follows.

Definition 1 ((ϵ, δ)-DP [30]): A randomized mechanism $\mathcal{M}$ satisfies (ϵ, δ)-DP: $\mathcal{X} \to \mathcal{R}$ with domain $\mathcal{X}$ and range $\mathcal{R}$, if for all measurable sets $\mathcal{Y} \subseteq \mathcal{R}$ and any two adjacent datasets, i.e., $\forall x, x' \in \mathcal{X}$, we have $\Pr[\mathcal{M}(x) \in \mathcal{Y}] \leq e^\epsilon \Pr[\mathcal{M}(x') \in \mathcal{Y}] + \delta$.

We can notice that central DP (CDP) and local DP (LDP) are both discussed in FL [31] and they are applied for different threat models. CDP in FL is utilized for the condition that there may exist an adversary, who can access to the global model, and infer that whether a client exists in the whole training process. In our model, we consider the curious-but-honest server and intend to preserve the privacy of each user's data x_i in the local setting, which belongs to LDP. Thus, we define the neighborhood dataset x'_i as adding or removing one record in the dataset x_i, $\forall x_i, x'_i$ in a database $\mathcal{X}_i$. The formal definition of (ϵ_i, δ_i)-LDP for the i-th user is as follows.

Definition 2 ((ϵ_i, δ_i)-LDP [32]): For the i-th client, a randomized mechanism $\mathcal{M}_i$ satisfies (ϵ_i, δ_i)-LDP: $\mathcal{X}_i \to \mathcal{R}_i$ with domain $\mathcal{X}_i$ and range $\mathcal{R}_i$, if for all measurable sets $\mathcal{Y}_i \subseteq \mathcal{R}_i$ and any two adjacent datasets $\forall x_i, x'_i \in \mathcal{X}_i$, we have $\Pr[\mathcal{M}_i(x_i) \in \mathcal{Y}_i] \leq e^{\epsilon_i} \Pr[\mathcal{M}_i(x'_i) \in \mathcal{Y}_i] + \delta_i$.

The perturbation mechanism $\mathcal{M}_i$ is applied to each user's dataset independently. Gaussian mechanism has been widely used in the privacy preserving stochastic gradient descent (SGD) algorithms [4] to protect users' privacy during training. Therefore, we adopt the Gaussian mechanism in this paper, which can satisfy (ϵ_i, δ_i)-LDP for the i-th client when we properly select the value of the standard deviation (STD) σ_i. Based on [27], each client can achieve the (ϵ_i, δ_i)-LDP requirement with a proper σ_i, where (ϵ_i, δ_i) is the LDP parameter for the i-th client and $\Delta\ell_i$ is the sensitivity of local training process.

Due to the varying channel, the unknown interference and the stochastic scheduling scheme, the exposure time for each client cannot be obtained in advance. Therefore, we only consider the LDP for each client in each communication round. Nevertheless, we can obtain the composition of leakage $\bar{\epsilon}_i$ based on the local privacy leakage according to [33] in each communication round as follows:

$$\bar{\epsilon}_i = \sqrt{\frac{E_i \ln(\frac{1}{\delta_i})}{\ln(\frac{2}{\delta_i})}} \epsilon_i, \qquad (6)$$

where E_i is the times of the model uploading for the i-th client and can be obtained at each communication round.

D. Problem Formulation

According to the aforementioned analysis, the time cost of the i-th client at the t-th communication round depends on three main components: broadcasting time, local training time, and uploading time, denoted by $d_{i,j}^D(t)$, $d_i^L(t)$, and $d_{i,j}^U(t)$, respectively. Since the central server has abundant computational resources compared to the clients, the latency

incurred by global model aggregation is negligible. Thus, the total time consumed by the i-th client at the t-th round is given as $d_{i,j}(t) = d_{i,j}^D(t) + d_{i,j}^U(t) + d_i^L(t)$ with the j-th channel which has been selected. We can observe that our model is consistent with the asynchronous FL described in [34]. This asynchronous setting is to guarantee the predefined system delay requirement. The time that the clients and the BS with N available channels require to jointly complete an update of their respective local and global FL models at t-th communication round is given by $d(t) = \max_{i \in \mathcal{U}} \{\min\{d_{i,j}(t), d_{\max}\}\}$. Here, we define $d_{\max}$ as the maximal interval of each communication round, which is used to avoid an endless waiting time caused by possible stragglers. Thus, it is crucial to control the largest delay among all clients. Having defined the system model, the next step is to design the dynamic channel assigning mechanisms in this work to minimize the time delay while competing the FL training. This optimization problem is formulated as follows:

P1: $$\min_{\boldsymbol{a}(1)...\boldsymbol{a}(T)} \sum_{t \in \mathcal{T}} d(t),$$

$$\text{s.t. } \mathbf{C1:} \; a_{i,j}(t) \in \{0, 1\}, \quad \forall i \in \mathcal{U}, j \in \mathcal{N},$$
$$\mathbf{C2:} \; \sum_{j \in \mathcal{N}} a_{i,j}(t) \leq 1, \quad \forall i \in \mathcal{U},$$
$$\mathbf{C3:} \; \sum_{i \in \mathcal{U}} a_{i,j}(t) = 1, \quad \forall j \in \mathcal{N},$$
$$\mathbf{C4:} \; \lim_{T \to \infty} \sup \frac{1}{T} \sum_{t \in \mathcal{T}} \mathbb{1}_i(t) \geq \beta_i, \quad \forall i \in \mathcal{U},$$

where β_i is the participation ratio for the i-th client determined by its DP requirement (ϵ_i, δ_i) and local training model $\boldsymbol{w}_i(t)$, which will be discussed detailly in the following section, $\boldsymbol{a}(t)$ is a selection matrix with size $U \times N$ at t-th communication round,[1] in which $a_{i,j}(t) = 1$ means the i-th client is assigned to the j-th channel and $\mathbb{1}_i(t)$ denotes an indicator function, which means whether the local training model of the i-th client has successfully received by the server at the t-th communication round, that is, if it is true, $\mathbb{1}_i(t) = 1$, otherwise $\mathbb{1}_i(t) = 0$. The domains of the variables in $\boldsymbol{a}(t)$ are defined by **C1**. Constraints **C2** and **C3** are maximum matching constraints, which control collisions among different clients [27]. The constraint **C4** is adopted to guarantee the participation ratio for various clients. From **P1**, we can find that the time used for the update of the local and global FL models is determined by the allocation matrix $\boldsymbol{a}(t)$ from clients to available channels. We consider a practical condition that the central server is unaware of wireless channel state information and statistical characteristics. In consequence, the time duration of local training and the time delay induced by model transmission are unavailable. Hence, it is challenging to solve problem **P1**.

III. LEARNING PERFORMANCE BOUND WITH DP AND DATA IMBALANCE

In this section, we will derive a learning convergence bound to determine the client's participation ratio. Investigating

---

[1]Here we assume that $U \geq N$. If $U < N$, we can use a transposed matrix, i.e., $\boldsymbol{a}(t)^\top$.

the convergence performance of FL training relies on three aspects: the number of rounds taken to reach a target accuracy, the number of participant clients for each training round and the data distribution of participant clients. The rationale behind FL is that when local epochs $\tau = 1$, i.e., when we perform global aggregation after every local update step, the distributed gradient descent is equivalent to the centralized gradient descent. Consequently, we evaluate the FL training performance by formulating the divergence between the i-th uploaded model $\widetilde{\boldsymbol{w}}_i(t)$ and the centralized training model $\widehat{\boldsymbol{w}}(t)$ at the t-th communication round, referred to as distributed model divergence. To this end, we formally bound the distributed model divergence as the following theorem.

Theorem 1: We assume that $F_i(\cdot)$ is λ_m-*Lipschitz smooth for m-class samples for all clients, where $m \in \{1, \ldots, M\}$. The divergence between $\widetilde{\boldsymbol{w}}_i(t)$ and $\widehat{\boldsymbol{w}}(t)$ at the t-th communication round can be written as*

$$
\begin{aligned}
&\mathbb{E}\{\|\widetilde{\boldsymbol{w}}_i(t) - \widehat{\boldsymbol{w}}(t)\|\} \\
&\leq \Theta_i \triangleq \sum_{j=0}^{\tau-1} (1 + \eta\lambda_{\max})^j \\
&\quad \times \left(\eta C \sum_{m=1}^{M} \|p_{i,m} - \overline{p}_m\| + \frac{4\eta C q_i \sqrt{2\tau \ln(1/\delta_i)}}{\sqrt{\pi} b_i \epsilon_i} \right), \quad (7)
\end{aligned}
$$

where η is the learning rate for SGD, b_i is the sampling size for each local epoch, $q_i = b_i/|\mathcal{D}_i|$ is the sampling rate, C is an upper bound on the norm of the gradient $\|\nabla F(\cdot)\|$, $\lambda_{\max} = \max_{m \in \{1, \ldots, M\}} \lambda_m$, $\overline{p}_m$ is the ratio of m-class samples for all clients and $p_{i,m}$ is the ratio of m-class samples in the i-th client, respectively.

Proof: Please see Appendix A. ∎

From **Theorem 1**, we can note that the divergence between $\widetilde{\boldsymbol{w}}_i(t)$ and $\widehat{\boldsymbol{w}}(t)$ is governed by two terms. The first term indicates that if $p_{i,m}$ is similar to $\overline{p}_m$, for all m and i, we will have a small divergence. An client's data is more useful for the learning if its distribution better represents the overall data distribution. Besides, we can find that the divergence increases with the value of τ. The second term shows the effect of DP requirement, in which a larger ϵ_i will lead to a larger divergence.

Then, we take the user-specific measurements out of (7), and then the participation ratio [22]–[24] for the i-th client can be determined by

$$
\beta_i = \min \left\{ N \cdot \frac{1/\Theta_i}{\sum_{u \in \mathcal{U}} 1/\Theta_u}, 1 \right\}. \quad (8)
$$

Combining with (7), we can note that if ϵ_i is larger, then β_i will be larger due to less noise on the i-th local model. This participation ratio β_i derived by the convergence bound of the i-th client is introduced to show the number of communication rounds in which the i-th client participates. Based on this participation ratio β_i for each client, we can optimize the communication resources while taking learning performance into account by utilizing the participation ratio as the constraint. In other word, we apply this bound as the constraint in optimizing the accumulated delay to achieve a balance between the training performance and delay. Superior

to the fairness guarantee [22]–[24], we can note that this constraint can not only avoid the occurrence of abandoning the slow but important clients due to the pursuit of low delay, but also involve the important clients in more rounds by setting a large participation ratio.

IV. Collaborative MAMAB Based FL

Our goal is to design allocation strategies online to minimize the accumulated transmission delay over a time horizon T when wireless channel conditions and the computation capability (i.e., available CPU frequencies in II-B) are unknown. In our framework, we adopt multi-armed bandit (MAB) to optimize the client scheduling by learning statistical property of the interference and the computation capability from the unknown environment, thereby minimizing the overall delay. Since there are multiple clients conornting different unknown transmission environments, we adopt the MAMAB framework to deal with this situation. Thus, in this section, we will model this FL training optimization problem as a sequential multi-agent decision making problem and reformulate this problem as a MAMAB problem.

A. Multi-Agent Multi-Armed Bandit Reformulation

At the t-th communication round, the chosen j-th channel for the i-th client is referred to as a super arm. Hence, the i-th client, $i \in \mathcal{U}$, is assigned to the j-th channel, $j \in \mathcal{N}$, which specifies the selected channel and observes its reward $r_{i,j}(t)$. The reward can be defined as $r_{i,j}(t) \triangleq \max\{1 - d_{i,j}(t)/d_{\max}, 0\}$. The reward $r_{i,j}(t)$ transforms the time delay $d_{i,j}(t)$ into a selection reward and if $d_{i,j}(t)$ is larger than $d_{\max}$, the system is set to receive a zero-reward. Clients who have timed will be dropped and will not participant in the aggregation in each communication round. The server will start the next communication round and transmit the global model to all clients after completing the aggregation process. The general reward $r(t)$ related with all users at the t-th communication round can be given as

$$
\hat{r}(t) = \min_{i \in \mathcal{U}} \sum_{j \in \mathcal{N}} r_{i,j}(t) a_{i,j}(t). \quad (9)
$$

Moreover, selecting clients is not only to find clients with good channels to quickly complete the communication round but also to identify clients with valuable data contributions to the learning itself. Thus, by considering the participation ratio β_i, the objective function can be reformulated as

P2: $\displaystyle \max_{\boldsymbol{a}(1)\ldots\boldsymbol{a}(T)} \sum_{t \in \mathcal{T}} \hat{r}(t)$, s.t. **C1**, **C2**, **C3**, **C4**.

We can note that optimization problem **P2** is a MAMAB problem, where each client acts as an agent and each channel acts as an arm. The server determines the channel assigning strategy based on its cumulative knowledge to maximize the accumulated reward with unknown distributions of the reward. We can solve the optimization problem **P2** by estimating the reward and then designing the assigning strategy at each communication round.

B. Virtual Queue and One-Slot Optimization Problem

In this subsection, we aim to solve **P2**. It can be noted that **P2** is a stochastic optimization problem under an unknown environment with a long-term constraint **C4**. According to [35], the long-term constraint **C4** in **P2** can be transformed into part of the objective function in **P3**. This transformation is sufficient and necessary. First, we can leverage the Lyapunov technique and transform constraint **C4** into queue stability constraints [36]. In detail, we introduce virtual queues Q_i with the following update equation: $Q_i(t) = [Q_i(t-1) + \beta_i - \mathbb{1}_i(t-1)]^+$, where $[x]^+ \triangleq \max\{x, 0\}$. We can note $\mathbb{1}_i(t-1)$ increases when the i-th client's local model is received by the server timely. Then, we define

$$\widetilde{r}_i(t) \triangleq V \sum_{j \in \mathcal{N}} r_{i,j}(t) a_{i,j}(t) + Q_i(t), \qquad (10)$$

where $V \geq 0$ strikes a balance between learning performance and delay experience. As can be seen in the above queue-length evolution, the value of the virtual queue to the i-th client increases by β_i in each round as β_i is the minimum selection fraction, and it decreases by one if the i-th client is uploading its model successfully.

Under the framework of Lyapunov optimization, we further resort to the drift-plus-penalty algorithm and solve the following optimization problem to determine channel allocation as

$$\textbf{P3:} \quad \max_{\boldsymbol{a}(1)\dots\boldsymbol{a}(T)} \sum_{t=1}^{T} \min_{i \in \mathcal{U}} \widetilde{r}_i(t), \text{ s.t. } \textbf{C1}, \textbf{C2}, \textbf{C3}.$$

Note that **P3** can be divided into T independent sub-optimization problems, for each communication round, the client selection strategy is optimized with the estimated reward based on the historical reward observations. The objective of **P3** is to maximize the accumulated reward over a time horizon based on its cumulative knowledge. Therefore, there exists a tradeoff between exploration (i.e., assigning all available channels a sufficient number of times to estimate the reward more accurately) and exploitation (i.e., assigning available channels to maximize the estimated reward).

C. Solution for Agent-Based Collaborative MAMAB

In order to solve **P3**, we now consider this MAMAB problem with i.i.d. rewards wherein multiple clients select available channels at the same time. The server and clients have no information about expectations or distribution of rewards from various available channels. Therefore, it is observed that the total reward function can be decomposed into a cumulative sum of minimum local agent-dependent reward functions for overall communication rounds. Instead of utilizing the average reward straightforwardly, we define the estimated reward by adding a perturbed term to the average reward [37], which can achieve a good tradeoff between exploration and exploitation. For the i-th client and the j-th channel, we can obtain the estimated reward at the end of the t-th communication

round as

$$e_{i,j}(t) = Q_i(t) + V\overline{r}_{i,j}(t)$$
$$+ V\sqrt{\frac{(U+2)\log(\sum_{u=0}^{t-1}\sum_{j\in\mathcal{N}} a_{i,j}(u))}{\sum_{u=0}^{t-1} a_{i,j}(u)}}, \quad (11)$$

where $\overline{r}_{i,j}(t)$ is the sample mean of rewards from the j-th channel for the i-th client at the t-th communication round and given by

$$\overline{r}_{i,j}(t) = \frac{\sum_{u=0}^{t-1} a_{i,j}(u) r_{i,j}(u)}{\sum_{u=0}^{t-1} a_{i,j}(u)}. \qquad (12)$$

Note that this perturbed term corresponds to the UCB in MAMAB, which is utilized in the combinatorial MAMAB problem [37].

With the estimated reward, we can optimize the channel assigning strategy by maximizing the total estimated reward at each time slot t. In this case, we summarize the detailed steps in **Algorithm 1**. Different from the conventional FL framework, **Algorithm 1** is designed that central server will transmit the matching result $\boldsymbol{a}(t)$ to all clients at the beginning of each communication round along with the global model based on the estimated reward $e_{i,j}(t)$, $i \in \mathcal{U}$, $j \in \mathcal{N}$. In addition, each client can upload an extra bit of information to the server to indicate the received time of the global model along with uploading local model, and we neglect the extra time consuming here. Meanwhile, the server is able to update the estimated reward $e_{i,j}(t)$. We can observe that T_0 is utilized to achieve a tradeoff between the exploitation of learned knowledge and the exploration of more potential actions.

Compared with the single-agent MAB problem, the main difficulty of the MAMAB problem is that there exists collisions among different agents. At this point, with the estimated reward $e_{i,j}(t)$ for all clients, we can note that the key of this algorithm is how to complete the max-min weighted bipartite matching at each communication round (obtain $\boldsymbol{a}(t)$). In the following section, we will propose solutions for this problem.

V. SOLUTIONS FOR MAX-MIN WEIGHTED MATCHING

At each communication round, with the estimated reward $e_{i,j}(t)$, we can complete clients scheduling by solving a bipartite matching for all clients and available channels to obtain a required delay. However, different from the conventional matching problem for a maximum cumulative reward, the target matching problem aims to maximize the minimum $R_i(t)$, $i \in \mathcal{U}$. Formally, a mathematical model for this problem transformed from **P3** can be expressed as

$$\textbf{P4:} \quad \max_{\boldsymbol{a}(t)} \min_{i \in \mathcal{U}} \sum_{j \in \mathcal{N}} a_{i,j}(t) e_{i,j}(t), \quad \text{s.t. } \textbf{C1}, \textbf{C2}, \textbf{C3}.$$

The objective function in **P4** maximizes the value of the minimum estimated reward of all clients, i.e., $\sum_{j \in \mathcal{N}} a_{i,j}(t) e_{i,j}(t)$, over all possible matchings. We can note that, in this problem, there is a complete weighted bipartite graph $\mathcal{G} = (\mathcal{U}, \mathcal{N}, \mathcal{E})$, where $\mathcal{U}$ and $\mathcal{N}$ are the sets of clients and channels, respectively, and $\mathcal{E}$ is the set of edges which value is corresponding to $e_{i,j}(t)$.

Algorithm 1 Collaborative MAMAB Based FL

Require: The number of participant clients U, the number of available channels N, and the exploitation and exploration parameter T_0

Ensure: The global model parameter $\boldsymbol{w}$

1: Initialize: $\boldsymbol{w}(0)$ and $t = 0$
2: **while** $t < T$ **do**
3:    The server broadcasts global model $\boldsymbol{w}(t)$ to all clients along with the channel assigning result $\boldsymbol{a}(t)$
4:    **for** all i in $\mathcal{U}$ **do**
5:       Local update for τ iterations and obtain: $\boldsymbol{w}_i(t+1)$
6:    **end for**
7:    All chosen clients send their local models to the server;
8:    The server receives all punctual models with $d_{\max}$;
9:    Perform the aggregation by (1);
10:    Updates the estimated reward $e_{i,j}(t+1)$ for all clients;
11:    Generate a random number κ;
12:    **if** $\kappa \geq 1 - e^{-\frac{t}{T_0}}$ **then**
13:       Generate a random bipartite matching result $\boldsymbol{a}(t)$;
14:    **else**
15:       Obtain the bipartite matching result $\boldsymbol{a}(t)$ by the solutions in the following section;
16:    **end if**
17:    Set $t = t+1$;
18: **end while**
19: **return** $\boldsymbol{w}(t)$

A. Modified Hungarian Algorithm

With a standard Hungarian algorithm [38], we can find the perfect matching with maximum cumulative rewards. Therefore, for the max-min weighted matching, we can prune the minimum edge of graph $\mathcal{G}$, and then try to search a perfect matching with the updated $\mathcal{G}'$. If we obtain the perfect matching successfully, we will prune the minimum edge and search a perfect matching consecutively. Otherwise, we output the perfect matching as the final result. In this way, we can complete the max-min weighted matching and term it as optimal matching (OM).

B. Greedy Matching With a Better Alternative

In this subsection, we introduce a modified greedy algorithm, termed greedy matching with a better alternative (GMBA) algorithm, which can achieve a verified low-complexity. **Algorithm 2** characterizes the procedure of the GMBA algorithm. At the t-th communication round, the server can possess the previous matching result $\boldsymbol{a}(t-1)$ and the estimated reward $e_{i,j}(t)$, $\forall i \in \mathcal{U}$. At the beginning, the server initializes a client set $\mathcal{A} = \{1, \ldots, N\}$, which includes all clients and an all zero matrix $\hat{\boldsymbol{a}}(t)$. The server randomly selects an greedy order from the set of all orders, i.e., $\boldsymbol{o} \in \mathcal{O}$ and assigns available channels to all clients. Here, we define $\boldsymbol{o}$ as a sequence of clients $(\boldsymbol{o}_1, \ldots, \boldsymbol{o}_U)$ such that $\boldsymbol{o}_i \in \mathcal{U}$ and $\boldsymbol{o}_i \neq \boldsymbol{o}_j$ for any $i \neq j$, and $\mathcal{O}$ denotes the set of all orders. Then, via the estimated reward, the i'-th client in $\boldsymbol{o}$ is assigned to an optimized channel from all available channels as: $\hat{a}_{i',j'}(t) = 1$ and $j' = \arg\max_{j \in \mathcal{A}} e_{i',j}(t)$.

Algorithm 2 Greedy Matching With a Better Alternative

Require: The previous matching result $\boldsymbol{a}(t-1)$, the estimated reward $e_{i,j}(t)$, $\forall i \in \mathcal{U}$, $\forall j \in \mathcal{N}$;

Ensure: The matching matrix $\boldsymbol{a}(t)$

1: Initialize: $\mathcal{A} = \{1, \ldots, N\}$, all zero matrix $\hat{\boldsymbol{a}}(t)$;
2: Select $\boldsymbol{o} \in \mathcal{O}$ uniformly at random;
3: **for** $i = 1$ to U **do**
4:    Find the i-th value in the order $\boldsymbol{o}$: $i' \leftarrow \boldsymbol{o}_i$;
5:    Select the optimized channel for the i-th value in the order $\boldsymbol{o}$: $j' = \arg\max_{j \in \mathcal{A}} e_{i',j}(t)$ and $\hat{a}_{i',j'}(t) = 1$;
6:    Update $\mathcal{A}$ by removing j';
7: **end for**
8: Update $\boldsymbol{a}(t) = \arg\max_{\boldsymbol{a} \in \{\boldsymbol{a}(t-1), \hat{\boldsymbol{a}}(t)\}} \min_{i \in \mathcal{U}} \sum_{j \in \mathcal{N}} a_{i,j} e_{i,j}(t)$;
9: **return** $\boldsymbol{a}(t)$

After assigning, the selected available channels j' is removed from the set of available channels $\mathcal{A}$. Until all clients finish the channel selection. Finally, we update $\boldsymbol{a}(t)$ as the better one in $\{\boldsymbol{a}(t-1), \hat{\boldsymbol{a}}(t)\}$. We can note that with an increasing communication round t, the estimated reward $e_{i,j}(t)$ will be more accurate and the performance of the GMBA algorithm will improve. In the following theorem, we will show the superiority of the proposed GMBA algorithm.

Theorem 2: The set of all greedy matching results $\mathcal{R}$ includes at least an optimal matching, i.e., $\boldsymbol{a}^\star \in \mathcal{R}$.

Proof: First, we assume that no client selects its best channel in the optimal matching $\boldsymbol{a}^\star$. From [39], we can find an equal matching with the same result with $\boldsymbol{a}^\star$, which is in conflict with the optimal claim. We can conclude that there is at least one client who selects its best channel (i.e., the channel with the highest estimated reward) in the optimal matching $\boldsymbol{a}^\star$ or its equal matching. Besides, the matching results for the set of clients that selecting their best available channels is not important under the greedy algorithm, because each client will choose a different channel. Hence, we can first process the set of clients that selects their best available channels, and then the remaining clients can reselect their best available channels within the remaining available channels. Repeating the procedure, we can obtain an order $\boldsymbol{o}$ that yields $\boldsymbol{a}^\star$ through the greedy algorithm. ∎

Via this theorem, we can note that if $e_{i,j}(t+1)$ is unchanged, finding the optimal matching $\boldsymbol{a}^\star$ by the greedy algorithm requires searching overall $U!$ permutations for U clients. However, we cannot obtain the mean reward directly, and thus $e_{i,j}(t+1)$ varies at different communication rounds. We can note that with a comparable large t, the performance of the proposed GMBA algorithm will be close to the optimal value, where more results can be found in the experimental results.

C. Complexity Analysis

At the beginning of each communication round, the server will calculate estimated rewards for all clients and available channels, which has a $O(UN)$ time complexity. The computational complexity of the Hungarian algorithm and the greedy matching algorithm with a given order $\boldsymbol{o}$ can be given by

$O(UN)$ and $O(UN)$, respectively. Then, for the Hungarian algorithm, we need at most $(N-1)U$ pruning operations. Therefore, under T communication rounds, the overall computational complexity of the OM and the GMBA algorithm can be expressed as $O\left(T(N-1)U^3N^2\right)$ and $O\left(TU^2N^2\right)$, respectively. From expressions, we can notice that the relationship between computational complexity and the key variables, i.e., N, U and T, is not exponential. Furthermore, the client scheduling is conducted in the server side and can be decomposed into multithreaded parallel computing tasks. Therefore, our proposed client scheduling schemes will be able to work well when N, U and T go large.

VI. PERFORMANCE EVALUATION

A. Feasibility and Regret Bound for MAMAB Algorithm

In this subsection, we first show **Algorithm 1** can satisfy the selection utility constraint for any minimum selection fraction vector and the constraints are satisfied as long as the virtual queue system defined is mean rate stable. Then, we show that the proposed algorithm also achieves the rate-optimality, i,e,. the logarithmic growth of the expected total regret with respect to time $\log T$. We state the first result as follows.

Theorem 3: The proposed **Algorithm 1** *is feasibility-optimal. Specifically, for any minimum selection fraction, the virtual queue system defined is strongly stable.*

Proof: See Appendix B. ∎

Theorem 3 implies that the constraints may be unsatisfied even after a sufficiently long time. This theorem states that our proposed algorithm can satisfy the participation ratio constraint as long as the requirement is feasible.

Then, we aim to bound the expected regret of the proposed algorithm. Let $\mu_{i,j}$ denote the average reward of the i-th client with the j-th channel. Let $\boldsymbol{a}^\star$ and $\mu^\star$ denote an optimal bipartite matching result (i.e., $\boldsymbol{a}^\star = \arg\max_{\boldsymbol{a}} \min_{i\in\mathcal{U}} \sum_{j\in\mathcal{N}} a_{i,j}\mu_{i,j}$) and the expected reward corresponding to $\boldsymbol{a}^\star$ (i.e., $\mu^\star = \max_{\boldsymbol{a}} \min_{i\in\mathcal{U}} \sum_{j\in\mathcal{N}} a_{i,j}\mu_{i,j}$), respectively. Then, we define

$$\Delta_{\min} \triangleq \mu^\star - \max_{\boldsymbol{a}, \boldsymbol{a}\neq\boldsymbol{a}^\star}\left\{\min_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}} a_{i,j}\mu_{i,j}\right\} \quad (13)$$

and

$$\Delta_{\max} \triangleq \mu^\star - \min_{\boldsymbol{a}}\left\{\min_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}} a_{i,j}\mu_{i,j}\right\}. \quad (14)$$

Under the assumption $\Delta_{\min} > 0$, we have the following expected regret shown in **Theorem 4**.

Theorem 4: Let ε be the precision of the bipartite matching algorithm and choose ε such that $\varepsilon < \Delta_{\min}-Q_{\max}$. Then, the expected regret of **Algorithm 1** *is given by*

$$R_{\text{reg}} \leq \Delta_{\max}\left(\frac{4V^2N(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2} + (2U+1)N\right), \quad (15)$$

where $Q_{\max}$ is the maximum value of $Q_i(t)$, $i\in\mathcal{U}$.

Proof: See Appendix C. ∎

It is obvious that the upper bound for the proposed algorithm is quite appealing as it separately captures the impact of the utility selection constraint and the impact of the uncertainty in the mean rewards for any finite number of communication rounds T. Specifically, when V is small, the proposed algorithm gives a higher priority to meeting the client participating ratio requirement by favoring a match with a larger virtual queue length, even if this match has a small estimated reward. Similarly, a larger V leads to a smaller regret, but it will take a longer time to converge which satisfies the participation ratio constraint. The part of $O(N \log T)$ in this regret corresponds to the notion of regret in typical MAMAB problems and is attributed to the cost that needs to be paid in the learning/exploration process.

B. Convergence Analysis for the Proposed MAMAB Based FL

We note that the wireless channel will influence the transmitting time for each client, and will lead to different participating clients of aggregation in each communication round. However, it is not easy to obtain the set of participating clients of aggregation in each communication round due to the varying channel, the unknown interference and the stochastic scheduling scheme. Thus, we define the set of participating clients of the t-th aggregation as $\mathcal{B}(t)$ and derive a convergence bound based on this predefined set.

We first mention the customary assumptions required for both convex and non-convex settings.

Assumption 1: We assume the following for all i:

2) *For any i, $F_i(\cdot)$ is λ_m-smooth for m-class samples for all clients, where $m \in \{1,\dots,M\}$;*

3) *For the learning rate η and $\lambda_{\max} = \max_{m\in\{1,\dots,M\}} \lambda_m$, $\eta\lambda_{\max} < 1$.*

Based on **Assumption 1**, we further analyze the convergence performance of the proposed client scheduling scheme.

Theorem 5: If we assume that the loss function $F_i(\boldsymbol{w})$ of the i-th client is convex, the convergence bound of **Algorithm 1** *is given by*

$$\mathbb{E}\left\{F(\boldsymbol{w}^{\mathcal{B}}(T))\right\} - F(\boldsymbol{w}^\star) \leq \frac{1}{T\tau\eta\varphi - \frac{T\rho\Xi(\tau)+\max_{t\in\mathcal{T}}\widehat{\Xi}(t)}{\varepsilon_0^2}}, \quad (16)$$

where $\rho \triangleq \omega(1-\frac{\lambda_{\max}}{2})$, $\omega \triangleq \min \frac{1}{\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}^\star\|}$, $\varepsilon_0 \triangleq \min_{t\in\mathcal{T}} \frac{1+\sqrt{1+4\eta\rho T^2\tau(\rho\Xi(\tau)+|\mathcal{B}(t)|)}}{2\eta\rho T\tau}$, (17) and (18), as shown at the bottom of the next page.

$D_{\min} = \min_{i\in\mathcal{U}} |\mathcal{D}_i|$ *and $\mathbb{P}_{\max}$ is the maximum participant probability of aggregation for all clients.*

Proof: See Appendix D. ∎

From **Theorem 5**, we can note that the proposed client scheduling scheme is converged when there exists $|\mathcal{B}(t)| > 0$. If all clients have a good enough channel and can upload their local models to the server successfully, the system can achieve a satisfied learning performance. If $|\mathcal{B}(t)| = 0$, it means that the t-th communication round has failed and is missing in the whole process. Moreover, the convergence bound is also governed by the maximum participant probability of aggregation $\mathbb{P}_{\max}$, which is determined by the computation capacity and channel quality for all clients. Hence, at each

communication round, if all clients can upload their training models successfully, the FL system will obtain a satisfied performance.

Moreover, we also provide a convergence bound for the case of non-convex loss functions in the following theorem.

Theorem 6: If we assume that the loss function $F_i(\boldsymbol{w})$ of the i-th client is non-convex, the convergence bound of **Algorithm 1** is given by

$$
\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{\tau-1}\nabla F\left(\widehat{\boldsymbol{w}}^j(t-1)\right)
$$
$$
\leq \frac{F\left(\boldsymbol{w}(0)\right)-F\left(\boldsymbol{w}^{\star}\right)}{T\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)}
$$
$$
+\frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)|\mathcal{D}|^2|\mathcal{B}|^2 D_{\min}^2}
$$
$$
+\frac{\lambda_{\max}\Xi(\tau)}{\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)}. \tag{19}
$$

Proof: See Appendix E.   ∎

The bound in **Theorem 6** implies that our proposed algorithm can achieve an overall convergence rate of $O\left(\frac{1}{T}\right)$ for non-convex losses. Similarly, when all clients can upload their training models successfully and the local epoch $\tau = 1$, i.e., equal to centralized learning, the FL system will obtain the best performance.

VII. EXPERIMENTAL RESULTS

A. Experimental Settings

We examine the results of the proposed algorithm, specifically the performance of **Algorithm 1**, on the following two neural networks and datasets: multi-layer perceptron (MLP) with FahionMNIST and convolutional neural network (CNN) with CIFAR-10.

- **MLP with FahionMNIST.** MLP is conducted on the FahionMNIST dataset [40]. MLP is a simple feed-forward deep neural network with ReLU units and soft-max of 10 classes (corresponding to the 10 categories) with cross-entropy loss. FahionMNIST is a dataset of fashion products consisting of $60,000$ training examples and $10,000$ testing examples formatted as 28×28 size gray scale images;

- **CNN with CIFAR-**10**.** The CNN model consists of three 3×3 convolution layers (the first with 64 filters, the second with 128 filters, the third with 256 filters, each followed with 2×2 max pooling and ReLu activation), two fully connected layers (the first with 128 units, the second with 256 units, each followed with ReLu activation), and a final softmax output layer. The CIFAR-10 dataset [41] consists of $60,000$ color images in 10 object classes such as deer, airplane, and dog with $6,000$ images included per class. The complete dataset is pre-divided into $50,000$ training images and $10,000$ test images. For CIFAR-10, we also use softmax of 10 classes with cross-entropy loss.

To evaluate the performance, we compare the proposed algorithm, i.e., **MAMAB-OM** (modified Hungarian algorithm) and **MAMAB-GMBA** (**Algorithm 2**), with the following baselines:

- **Random Scheduling [17]:** In each communication round, the BS uniformly select U associated clients at random for parameter update, each selected client is assigned a dedicated subchannel to transmit the trained parameter.

- **Round Robin [17]:** The BS arranges all the clients into $\lceil\frac{U}{N}\rceil$ groups and consecutively assigns each group to access the radio channels and update their parameters per communication round.

- **Single-UCB [23]:** The BS selects a subset consisting of N clients from U associated clients with the maximum total rewards via a single UCB policy.

Besides, simulations are performed in a square area of $2,000\times2,000\ m^2$. Both available channels and clients are uniformly distributed in this plane. Unless otherwise stated, the system parameters are set as follows. We set the number of clients to 10, the number of available channels to 4, the degree of non-IID to 0.8 and the number of local iterations to 5, respectively. For the wireless channel, we set the bandwidth B^{U} and B^{D} to 15 KHz, the transmission power of available channels to 23 dBm, Gaussian white noise power to -107 dBm, and path loss exponent model to PL, where $PL[dB] = 128.1+37.6\log(k)$ with k representing the distance in km. The uplink and downlink interferences are generated by the Gaussian distribution with different variances. The computing capability $f_i(t)$ of each client is uniformly distributed in $[10*i+10, 100*i+30]$ KHz, and the maximal

$$
\widehat{\Xi}(t)=\begin{cases}\dfrac{\lambda_{\max}\mathbb{P}_{max}(1-\mathbb{P}_{max})\sum\limits_{i\in\mathcal{U}}\sum\limits_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2+\Theta_j^2\right)}{2|\mathcal{D}|^2|\mathcal{B}(t)|^2 D_{min}^2}+\lambda_{max}\Xi(\tau), & |\mathcal{B}(t)|>0,\\[4mm]\eta\tau C+\Xi(\tau), & |\mathcal{B}(t)|=0,\end{cases} \tag{17}
$$

$$
\Xi(\tau)=\eta C\left\|\sum_{i\in\mathcal{U}}(p_i-p_i(t))\right\|+\left(\frac{4\eta C\sqrt{2\tau}}{\sqrt{\pi}}\sqrt{\sum_{i\in\mathcal{U}}\frac{p_i^2 q_i^2\ln(1/\delta_i)}{b_i^2\epsilon_i^2}}\right.
$$
$$
\left.+\eta C\sum_{i\in\mathcal{U}}p_i(t)\sum_{m\in\mathcal{M}}\|p_{i,m}-q_m\|\right)\sum_{j=1}^{\tau-1}\left(1+\eta\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\right)^j, \tag{18}
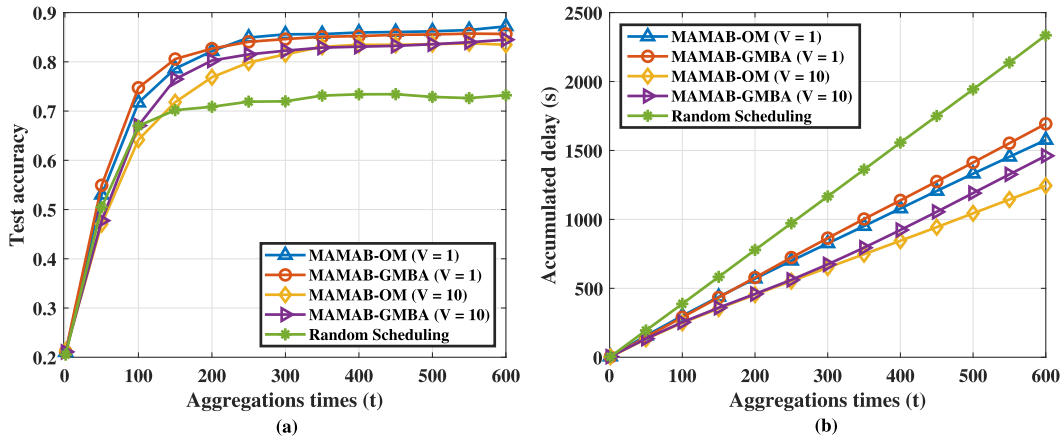$$

Fig. 2. Instantaneous results of the proposed MAMAB algorithm, i.e., MAMAB-OM and MAMAB-Greedy, along with increasing communication rounds on MLP with FashionMNIST. (a) and (b) show the test accuracy and accumulated delay, respectively.

interval $d_{max} = 2$ seconds for the CIFAR-10 and $d_{max} = 5$ seconds for the FahionMNIST, respectively. In addition, the required DP parameters for all clients are set to $\epsilon_i = 25$ and $\delta_i = 0.001, \forall i \in \mathcal{U}$, respectively. In **Algorithm 1**, we adopt various V, i.e., $V = 1$, $V = 10$ and $V = 100$, and $T_0 = 100$ to balance the learning performance and training delay.

B. Evaluation of the Proposed Algorithm

Fig. 2 illustrates the test accuracy of our proposed MAMAB algorithms with different values of V, and random scheduling on MLP with FahionMNIST. As seen from Fig. 2(a), the proposed MAMAB algorithm using different values of V achieve a better accuracy than the random one. We can also observe that the performance gap decreases as the number of communication rounds increases. The reason is that the test accuracy is close to the limitation with a sufficiently large number of communication rounds. Moreover, the value of V is a key factor to balance the trade-off between the latency and the client selection rate. As shown in Fig. 2(b), our proposed MAMAB algorithms with larger values of V bring out lower latencies. The intuition is that a larger V can lead to a higher consideration for the FL training delay but a smaller consideration for the learning performance. From Figs. 2(a) and 2(b), we can also notice that our proposed MAMAB algorithms with $V = 1$ show the better test accuracy than the one with $V = 10$, but lead to higher latencies. The reason is the trade-off between the latency and training accuracy, that is: guaranteeing the chosen rate for each client can achieve a satisfied accuracy but with larger latency for a long term, while reducing the latency per round but degrading the training accuracy.

In addition, we also evaluate this MLP based FL system on FahionMNIST by setting a small privacy level, i.e, $\epsilon_i = 0.8, \forall i \in \mathcal{U}$ and a small data sampling size, i.e., $b_i = 10, \forall i \in \mathcal{U}$. Similar to Fig. 2, our proposed MAMAB algorithms using different values of V achieve a better accuracy than the random one in Fig. 3(a). Meanwhile, our proposed MAMAB algorithms with larger values of V incur lower latencies as shown in Fig. 3(b). Besides, we can also calculate the composition of

leakage and obtain the maximum privacy leakage among all clients, i.e., $\overline{\epsilon} = \max_{i \in \mathcal{U}} \overline{\epsilon}_i = 12.05$ using (6). Due to the high privacy level, i.e., a small ϵ, we can observe that the training performance in Fig. 3(a) is worse than that in Fig. 2(a).

Fig. 4 shows aggregation times achieved at a variety of maximum tolerable delay of our proposed MAMAB algorithms on FashionMNIST and CIFAR-10, respectively. We see that our proposed MAMAB-OM algorithm can achieve more aggregation times than the MAMAB-GMBA one. Furthermore, with a larger V, our proposed algorithms can complete more training epochs with a fixed maximum tolerable delay. The reason is that a larger V can lead to a higher consideration for the FL training delay.

Then the client selection rates for each client with our proposed MAMAB-OM algorithm under various values of V are shown in Fig. 5. In Figs. 5(a) and 5(b), we consider two different DP schemes: 1) *uniform*, where all clients adopt the same DP parameter, i.e, $\epsilon_i = 25$ and $\delta = 0.001$, 2) *non-uniform*, where the DP parameter for the ith client is set to $\epsilon_i = 5*(\lfloor \frac{i-1}{2} \rfloor + 3)$ and $\delta = 0.001$. From Fig. 5(b), we can note that when i is small, the privacy level is high and then leads to a small selection requirement. Furthermore, we can observe that the value of V has a significant effect on the selection rate for each client. This result reveals the tradeoff between the guarantee for the selection rate and the training latency.

C. Comparison of Different Scheduling Policies

To show the effectiveness of **Algorithm 1**, we compare our proposed MAMAB-OM and MAMAB-GMBA algorithms with a set of baseline policies, i.e., Random Scheduling, Round Robin, Single-UCB-0.01 and Single-UCB-0.1.

Fig. 6 shows the changes of test accuracy along with the communication rounds t using our proposed MAMAB algorithms and baseline policies on FashionMNIST and CIFAR-10, respectively. We can find that the proposed MAMAB algorithms achieves a better performance than baseline policies. The reason/intuition is that the proposed MAMAB algorithms keep learning the statistical information of each client in
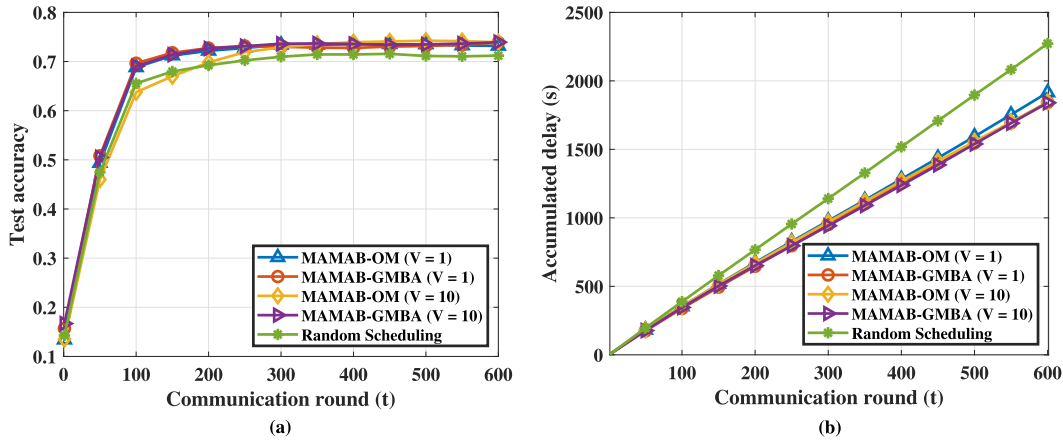
Fig. 3. Instantaneous results of the proposed MAMAB algorithm with $\epsilon_i = 0.8$, i.e., MAMAB-OM and MAMAB-Greedy, along with increasing communication rounds on MLP on FashionMNIST. (a) and (b) show the test accuracy and accumulated delay, respectively.
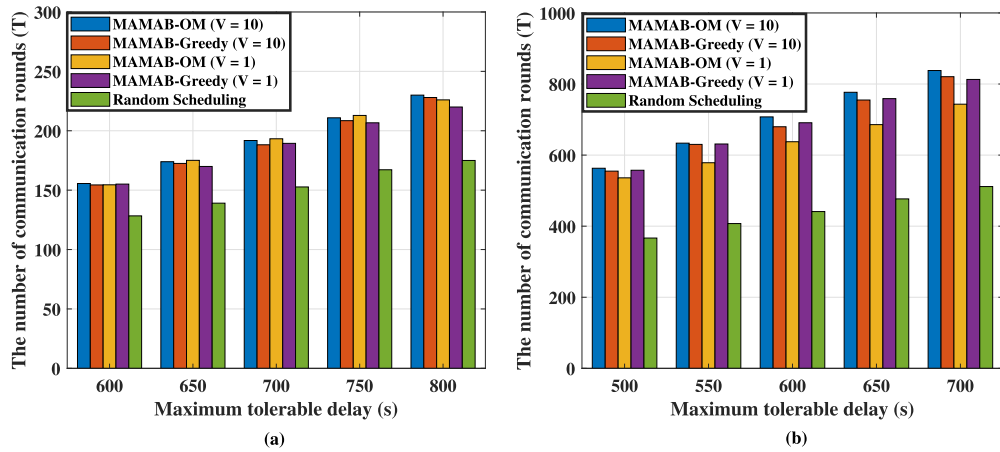

Fig. 4. Achievable number of communication rounds of the proposed MAMAB algorithm, i.e., MAMAB-OM and MAMAB-GMBA, along with maximum tolerable delay on different datasets. (a) and (b) show the MLP on FashionMNIST and CNN on CIFAR-10, respectively.
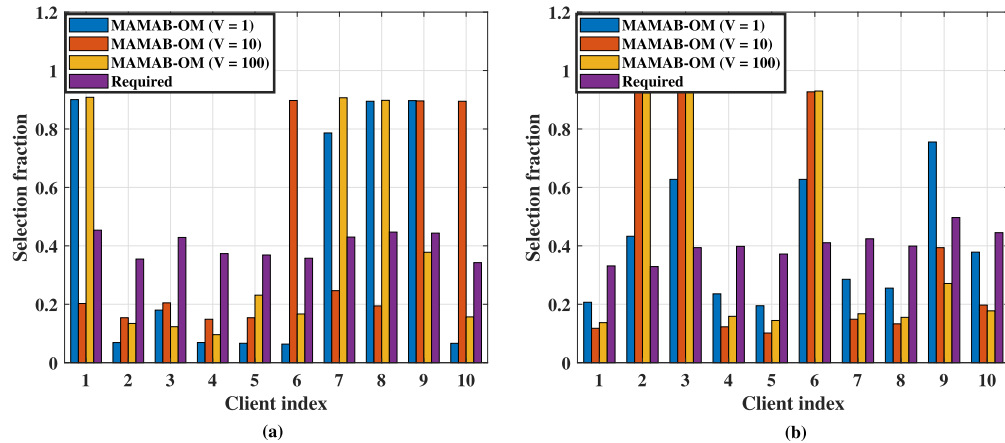

Fig. 5. The selection fraction of the proposed MAMAB-OM algorithm with various clients on different datasets. (a) and (b) show the MLP on FashionMNIST and CNN on CIFAR-10, respectively.

the execution process and leverages a tradeoff between the exploitation of learned knowledge and the exploration of more potential actions. Although Single-UCB-0.01 and Single-UCB-0.1 also have the ability to exploit the statistical information of channel state and computation capacity for each client, but they cannot overcome the difference of interference for different clients.

Fig. 7 illustrates the changes of cumulative delay along with the communication rounds t using our proposed MAMAB algorithms and baseline policies on FashionMNIST and CIFAR-10, respectively. As seen from Fig. 7, the proposed MAMAB algorithms achieve lower cumulative delay than baseline policies and the superiority is more and more obvious with increasing number of communication rounds.
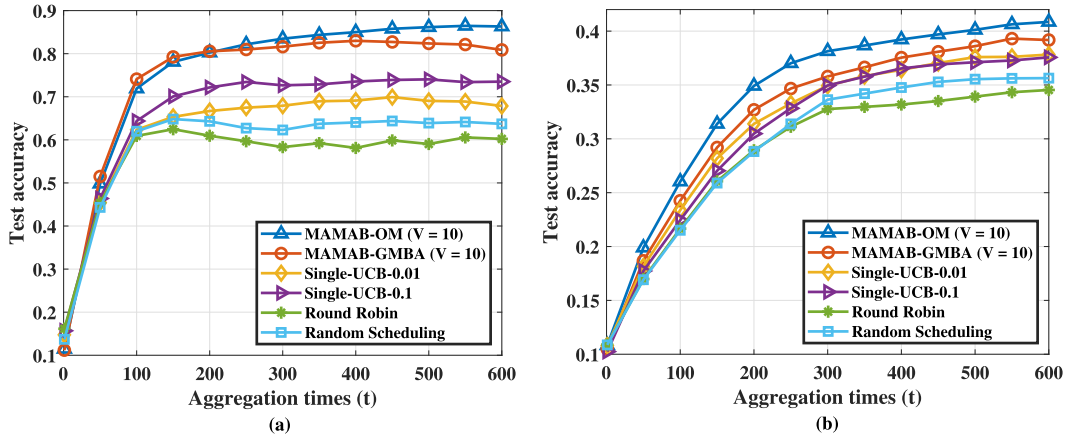
Fig. 6. The test accuracy of the proposed MAMAB algorithm, i.e., MAMAB-OM and MAMAB-GMBA, along with increasing communication rounds. (a) and (b) show the MLP on FashionMNIST and CNN on CIFAR-10, respectively.
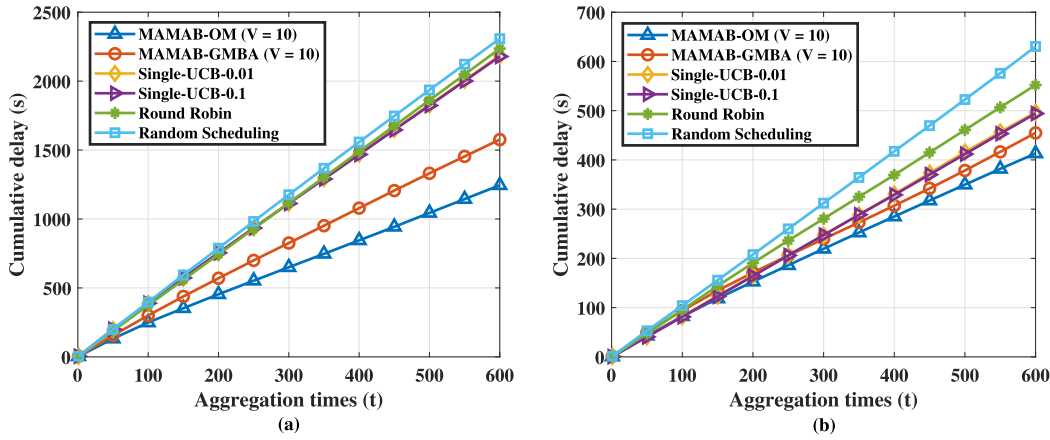

Fig. 7. The test accuracy of the proposed MAMAB algorithm, i.e., MAMAB-OM and MAMAB-GMBA and baselines, along with increasing communication rounds. (a) and (b) show the MLP on FashionMNIST and CNN on CIFAR-10, respectively.

The advantage of our proposed MAMAB algorithms is twofold: Firstly, our proposed MAMAB algorithms schedules the clients with better channel conditions and computation capabilities with an appropriate matching according to the estimated reward, and thus can reduce the per round latency compared to baseline policies. Secondly, our proposed MAMAB algorithms achieve a better trade-off between the latency per round and the training performance.

VIII. CONCLUSION

In this paper, we have developed a realistic implementation of FL over a wireless network with imbalanced resources and DP requirements among clients, i.e., stochastic training frequencies, DP guarantee, uncertain interference of wireless channels, non-IID data and various data sizes for clients. Then, we have formulated the joint client selection and channel assignment problem within a MAMAB framework to minimize the time delay while taking FL convergence into account. We have developed a convergence bound for each client to derive a client participation ratio to constrain the above MAMAB problem. In order to address this constraint, we have introduced a virtual queue inspired by Lyapunov

optimization. We then have adopted the UCB method to estimate the model upload time for each client and channel to convert this MAMAB to a max-min weighted bipartite graph matching problem, and then proposed the modified Hungarian and GMBA algorithms, to schedule clients at each communication round. An upper bound on the expected regret of the proposed MAMAB based FL has been derived and shows a linear growth over the logarithm of communication rounds, justifying its theoretical feasibility. Extensive experimental results have been provided to validate the effectiveness of our proposed algorithms, and outperforms state-of-the-art baseline scheduling policies. We have observed that if the maximum waiting time is large enough to ensure all the clients' models participate in the aggregation of every communication round, then we can obtain the optimal training performance.

APPENDIX A
PROOF OF THEOREM 1

In this proof, we first denote the i-local model between the $(t-1)$-th and the t-th communication round by $\boldsymbol{w}_i^\tau(t-1)$. For the decentralized training, we define $\widehat{\boldsymbol{w}}^\tau(t-1) = \sum_{i=1}^U p_i \widehat{\boldsymbol{w}}_i^\tau(t-1)$. Based on the update rule of

gradient decent, we have

$$\widetilde{\boldsymbol{w}}(t) = \sum_{i \in \mathcal{U}} p_i(t) \sum_{m \in \mathcal{M}} p_{i,m}(\boldsymbol{w}_i^{\tau-1}(t-1)$$
$$-\eta \nabla F_{i,m}(\boldsymbol{w}_i^{\tau-1}(t-1))) + \sum_{i \in \mathcal{U}} p_i(t) \boldsymbol{n}_i, \quad (20)$$

and

$$\widehat{\boldsymbol{w}}(t) = \sum_{i \in \mathcal{U}} p_i \sum_{m \in \mathcal{M}} p_{i,m}(\widehat{\boldsymbol{w}}_i^{\tau-1}(t-1)$$
$$-\eta \nabla F_{i,m}(\widehat{\boldsymbol{w}}_i^{\tau-1}(t-1)))$$
$$= \widehat{\boldsymbol{w}}^{\tau-1}(t-1) - \eta \nabla F_{i,m}(\widehat{\boldsymbol{w}}_i^{\tau-1}(t-1)), \quad (21)$$

where $p_{i,m}$ is the ratio of m-class samples in the i-th client. Then, we need to bound $\|\widetilde{\boldsymbol{w}}_i^{\tau}(t-1) - \widehat{\boldsymbol{w}}^{\tau}(t-1)\|$ by

$$\|\widetilde{\boldsymbol{w}}_i^{\tau}(t-1) - \widehat{\boldsymbol{w}}^{\tau}(t-1)\| \leq \left\| \boldsymbol{w}_i^{\tau-1}(t-1) - \widehat{\boldsymbol{w}}^{\tau-1}(t-1) \right\|$$
$$+\eta \left\| \sum_{m \in \mathcal{M}} p_{i,m} \nabla F_{i,m}(\boldsymbol{w}_i^{\tau-1}(t-1)) - p_{i,m} \nabla F_m(\widehat{\boldsymbol{w}}^{\tau-1}(t-1)) \right\|$$
$$+\eta \left\| \sum_{m=1}^{M} (p_{i,m} - \overline{p}_m) \nabla F_m(\boldsymbol{w}_i^{\tau-1}(t-1)) \right\| + \|\boldsymbol{n}_i\|. \quad (22)$$

Considering uniformly Lipschitz continuous, we can obtain the following inequation:

$$\|\widetilde{\boldsymbol{w}}_i^{\tau}(t-1) - \widehat{\boldsymbol{w}}^{\tau}(t-1)\|$$
$$\leq \left(1 + \eta \sum_{m=1}^{M} p_{i,m} \lambda_m \right) \left\| \boldsymbol{w}_i^{\tau-1}(t-1) - \widehat{\boldsymbol{w}}^{\tau-1}(t-1) \right\|$$
$$+\eta \sum_{m=1}^{M} \|p_{i,m} - \overline{p}_m\| \left\| \nabla F_m(\widehat{\boldsymbol{w}}^{\tau-1}(t-1)) \right\| + \|\boldsymbol{n}_i\|, \quad (23)$$

where $\overline{p}_m$ is the ratio of m-class samples for all clients. Applying (23) recursively, we have

$$\|\widetilde{\boldsymbol{w}}_i^{\tau}(t-1) - \widehat{\boldsymbol{w}}^{\tau}(t-1)\|$$
$$\leq \left(1 + \eta \sum_{m=1}^{M} p_{i,m} \lambda_m \right)^{\tau} \|\boldsymbol{w}(t-1) - \widehat{\boldsymbol{w}}(t-1)\|$$
$$+\eta C \sum_{m=1}^{M} \|p_{i,m} - \overline{p}_m\| \sum_{j=0}^{\tau-1} \left(1 + \eta \sum_{m=1}^{M} p_{i,m} \lambda_m \right)^{j} + \|\boldsymbol{n}_i\|, \quad (24)$$

where C is an clipping upper bound on any gradient vector $\nabla F(\boldsymbol{w})$. We can remark that gradient clipping is a popular ingredient of SGD and ML. We can note that $\boldsymbol{w}(t-1) = \widehat{\boldsymbol{w}}(t-1)$ at the beginning of each communication round. Thus, we have

$$\mathbb{E}\{\|\widetilde{\boldsymbol{w}}_i(t) - \widehat{\boldsymbol{w}}(t)\|\}$$
$$\leq \eta C \sum_{m=1}^{M} \|p_{i,m} - \overline{p}_m\| \sum_{j=0}^{\tau-1} (1 + \eta \lambda_{\max})^{j}$$
$$+\frac{4C \eta q_i \sqrt{2\tau \ln(1/\delta_i)}}{\sqrt{\pi} b_i \epsilon_i} \sum_{j=0}^{\tau-1} (1 + \eta \lambda_{\max})^{j}, \quad (25)$$

where η is the learning rate for stochastic gradient descent (SGD), b_i is the batch size, $q_i = b_i/|\mathcal{D}_i|$ is the sampling rate,

C is the upper bound on the norm of the gradient $\|\nabla F(\cdot)\|$, $\lambda_{max} = \max_{m \in \{1, \dots, M\}} \lambda_m$, $\overline{p}_m$ is the ratio of m-class samples for all clients and $p_{i,m}$ is the ratio of m-class samples in the i-th client. This completes the proof. $\square$

APPENDIX B
PROOF OF THEOREM 3

We first introduce the Lyapunov function $L(\boldsymbol{Q}(t)) = \frac{1}{2} \sum_{i \in \boldsymbol{S}} Q_i(t)^2$, of which the drift from one slot is given as

$$L(Q_i(t+1)) - L(Q_i(t))$$
$$= \frac{1}{2} \sum_{i \in \mathcal{S}(t)} (\max\{Q_i(t) + \beta_i - \mathbb{1}_i(t), 0\})^2 - \frac{1}{2} \sum_{i \in \mathcal{S}(t)} Q_i(t)^2$$
$$\leq \frac{1}{2} \sum_{i \in \mathcal{S}(t)} (\beta_i - \mathbb{1}_i(t))^2 - \sum_{i \in \mathcal{S}(t)} Q_i(t)(\beta_i - \mathbb{1}_i(t)). \quad (26)$$

Because both β_i and $\mathbb{1}_i(t)$ are within $[0, 1]$, we have

$$L(Q_i(t+1)) - L(Q_i(t)) \leq \frac{U}{2} + \sum_{i \in \mathcal{S}(t)} Q_i(t)(\beta_i - \mathbb{1}_i(t)). \quad (27)$$

Further, we define the conditional Lyapunov drift for the t-th communication round as

$$\mathbb{E}\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t))|\boldsymbol{Q}(t)\}$$
$$\leq \frac{U}{2} + \sum_{i \in \mathcal{S}(t)} Q_i(t)\beta_i - \mathbb{E}\left\{ \sum_{i \in \mathcal{S}(t)} Q_i(t)\mathbb{1}_i(t)|\boldsymbol{Q}(t) \right\}$$
$$= \frac{U}{2} + \sum_{i \in \mathcal{S}(t)} Q_i(t)\beta_i - \mathbb{E}\left\{ \sum_{i \in \mathcal{S}(t)} R_i(t)|\boldsymbol{Q}(t) \right\}$$
$$+V\mathbb{E}\left\{ \sum_{i \in \mathcal{S}(t)} \mathbb{1}_i(t) \sum_{j \in \mathcal{N}} r_{i,j}(t)\boldsymbol{a}_{i,j}(t)|\boldsymbol{Q}(t) \right\}. \quad (28)$$

Due to $r_{i,j}(t) \leq 1$, $\boldsymbol{a}_{i,j}(t) \in \{0, 1\}$ and $\mathbb{1}_i(t) \in [0, 1]$, $\forall i \in \mathcal{S}(t), j \in \mathcal{N}$, we have

$$\mathbb{E}\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t))|\boldsymbol{Q}(t)\}$$
$$\leq \frac{U}{2} + UV + \sum_{i \in \mathcal{S}(t)} Q_i(t)\beta_i - \mathbb{E}\left\{ \sum_{i \in \mathcal{S}(t)} \widetilde{r}_i(t)|\boldsymbol{Q}(t) \right\}. \quad (29)$$

Based on Theorem 4.5 in [35] and Lemma 1 in [42], for any $\zeta > 0$, we can make the following inequalities that hold:

$$\mathbb{E}\left\{ \sum_{i \in \mathcal{S}(t)} \widetilde{r}_i(t)|\boldsymbol{Q}(t) \right\} \geq \sum_{i \in \mathcal{S}(t)} Q_i(t)(\beta_i + \epsilon). \quad (30)$$

Therefore, $\mathbb{E}\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t))|\boldsymbol{Q}(t)\} \leq \frac{U}{2} + UV - \zeta \sum_{i \in \mathcal{S}(t)} Q_i(t)$. Finally, invoking Theorem 4.5 in [35] (Lyapunov Drift Theorem) and with the condition that $\zeta > 0$, we can conclude that all the virtual queues defined is not only mean rate stable, but also strongly stable, i.e.,

$$\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left\{ \min_{i \in \mathcal{U}} \widetilde{r}_i(t) \right\} \leq \frac{U(1+2V)}{2\zeta} < \infty. \quad (31)$$

This completes the proof. $\square$

APPENDIX C
PROOF OF THEOREM 4

Under the MAMAB, the time-average regret is defined as

$$R_{\text{reg}} = T\mu^\star - \mathbb{E}\left\{\frac{1}{T}\sum_{t=0}^{T-1}\min_{i\in\mathcal{U}}\widetilde{r}_i(t)\right\}. \tag{32}$$

We first upper bound the number of sub-optimal plays. We define $x_{i,j}(t)$ as follows: $x_{i,j}(t) = \sum_{u=0}^{t-1}a_{i,j}(u)$. Let $\boldsymbol{x}(t)$ denote the total number of suboptimal plays, where $\boldsymbol{x}(t) = \sum_{i\in\mathcal{U}}\sum_{j=1}^{N}x_{i,j}(t)$. Let $\hat{\nu}_{i,j}^t$ be the time at which the i-th client makes the t-th transition to the j-th channel from another channel and $\tilde{\nu}_{i,j}^t$ be the time at which the agent makes the t-th transition from the j-th channel to another channel. We assume $\tilde{\nu}_{i,j}^t = \min\{\tilde{\nu}_{i,j}^t, T\}$. Therefore, in order to bound $\boldsymbol{x}(T)$, we first obtain a bound on $x_{i,j}(T)$ as

$$x_{i,j}(T) \le 1 + \sum_{t=0}^{T-1}\boldsymbol{I}\left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}e_{i,j}(t)a_{i,j}^\star(t)\right\}\right.$$
$$\le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}e_{i,j}(t)a_{i,j}(t)\right\}\right\}$$
$$\le l + \sum_{t=0}^{T-1}\boldsymbol{I}\left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}e_{i,j}(t)a_{i,j}^\star(t)\right\}\right.$$
$$\le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}e_{i,j}(t)a_{i,j}(t)\right\}, x_{i,j}(t)\ge l\right\}, \tag{33}$$

where $\boldsymbol{I}\{\cdot\}$ is the event in $\{\cdot\}$. According to (11), we can obtain

$$e_{i,j}(t) = V\overline{r}_{i,j}(t) + c_{i,j}(t), \tag{34}$$

where

$$c_{i,j}(t) = Q_i(t) + V\sqrt{\frac{(U+2)\log(\sum_{u=0}^{t-1}\sum_{j\in\mathcal{N}}a_{i,j}(u))}{\sum_{u=0}^{t-1}a_{i,j}(u)}}. \tag{35}$$

Substituting (34) into (33), we can obtain

$$x_{i,j}(T) \le l + \sum_{t=0}^{T-1}\boldsymbol{I}\left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right.$$
$$\le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)(\overline{r}_{i,j}(t)\right.$$
$$\left.+c_{i,j}(t))\right\}, x_{i,j}(t-1)\ge l\right\}. \tag{36}$$

Because $1 \le \sum_{u=0}^{t-1}a_{i,j}(u) < t$, and then we have

$$x_{i,j}(T) \le l + \sum_{t=0}^{T-1}\boldsymbol{I}\left\{\min_{c_{i,j}(t)}\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right.$$
$$\le \varepsilon + \max_{c_{i,j}(t)}\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)(\overline{r}_{i,j}(t)\right.$$
$$\left.+c_{i,j}(t))\right\}, x_{i,j}(t-1)\ge l\right\}. \tag{37}$$

We define

$$s_i \triangleq \sum_{u=0}^{t-1}\sum_{j\in\mathcal{N}}a_{i,j}^\star(u), \quad s_i' \triangleq \sum_{u=0}^{t-1}\sum_{j\in\mathcal{N}}a_{i,j}(u). \tag{38}$$

Due to $1 \le s_i, s_i' \le t$, we can obtain

$$x_{i,j}(T) \le l + \sum_{t=1}^{\infty}\sum_{s_1=1}^{t}\cdots\sum_{s_U=1}^{t}\sum_{s_1'=l}^{t}\cdots\sum_{s_U'=l}^{t}$$
$$\times\boldsymbol{I}\left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right.$$
$$\left.\le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right\}. \tag{39}$$

Now, it is easy to observe that the event

$$\boldsymbol{I}\left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right.$$
$$\left.\le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)(\overline{r}_{i,j}(t)+c_{i,j}(t))\right\}\right\} \tag{40}$$

implies at least one of the following events:

$$A_i : \left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)\overline{r}_{i,j}(t) \le \sum_{j\in\mathcal{N}}a_{i,j}^\star(t)(\mu_{i,j}-c_{i,j}(t))\right\},$$
$$B_i : \left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)\overline{r}_{i,j}(t) \ge \sum_{j\in\mathcal{N}}a_{i,j}(t)(\mu_{i,j}+c_{i,j}(t))\right\},$$
$$C : \left\{\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)\mu_{i,j}\right\} \le \varepsilon + \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)\mu_{i,j}\right\}\right.$$
$$\left.+2\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)c_{i,j}(t)\right\}\right\}. \tag{41}$$

Invoking the Chernoff-Hoeffding inequality [43], we can obtain

$$\mathbb{P}(A_i) \le e^{-2c_{i,j}^2(t)} \le t^{-2V^2(U+2)},$$
$$\mathbb{P}(B_i) \le t^{-2V^2(U+2)}, \quad i\in\mathcal{U}. \tag{42}$$

Then, if $l \ge \left\lceil\frac{4V^2(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2}\right\rceil$, we can obtain

$$\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)\mu_{i,j}\right\}-\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)\mu_{i,j}\right\}$$
$$-2\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)c_{i,j}\right\}-\varepsilon$$
$$\ge \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)\mu_{i,j}\right\}-\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)\mu_{i,j}\right\}$$
$$-2V\sqrt{\frac{(U+2)\log(t)}{l}}-Q_{\max}-\varepsilon$$
$$\ge \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}^\star(t)\mu_{i,j}\right\}-\min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}a_{i,j}(t)\mu_{i,j}\right\}$$
$$-(\Delta_{\min}-Q_{\max}-\varepsilon)-\varepsilon \ge 0. \tag{43}$$

Furthermore, we can obtain

$$\mathbb{E}[x_{i,j}(T)] \le \left\lceil\frac{4V^2(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2}\right\rceil$$
$$+2U\sum_{t=1}^{\infty}\sum_{s_1=1}^{t}\cdots\sum_{s_U=1}^{t}\sum_{s_1'=l}^{t}\cdots\sum_{s_U'=l}^{t}t^{-2V^2(U+2)}$$

$$\leq \frac{4V^2(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2}+2U+1 \tag{44}$$

and

$$\mathbb{E}[\boldsymbol{x}(T)]$$
$$= \min_{i\in\mathcal{U}}\left\{\sum_{j\in\mathcal{N}}x_{i,j}(T)\right\}\frac{4V^2N(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2}+(2U+1)N. \tag{45}$$

Now, we can bound the regret as

$$R_{\text{reg}} \leq \Delta_{\max}\left(\frac{4V^2N(U+2)\log T}{(\Delta_{\min}-Q_{\max}-\varepsilon)^2}+(2U+1)N\right). \tag{46}$$

This completes the proof. $\qquad\square$

APPENDIX D
PROOF OF THEOREM 5

We define a subset $\mathcal{B}$ with a size $|\mathcal{B}|$, which is determined by user selection policy, available computation resources and channel states, and $\mathcal{B}$ is changed during the whole training process. Due to $\lambda_{\max}$-Lipschitz smoothness, we have $\mathbb{E}\left\{F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\boldsymbol{w}(t))\right\} \leq \frac{L}{2}\mathbb{E}\left\{\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}(t)\|^2\right\}$. Then, we can obtain

$$\mathbb{E}\left\{\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}(t)\|^2\right\}$$
$$= \mathbb{E}\left\{\left\|\frac{\sum_{i\in\mathcal{B}}|\mathcal{D}_i|(\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t))}{\sum_{i\in\mathcal{B}}|\mathcal{D}_i|}\right\|^2\right\}$$
$$\leq \frac{\mathbb{E}\left\{\left\|\sum_{i\in\mathcal{U}}\mathbb{1}_i(t)|\mathcal{D}_i|(\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t))\right\|^2\right\}}{|\mathcal{B}|^2D_{\min}^2}, \tag{47}$$

where $D_{\min} = \min\{|\mathcal{D}_1|\ldots|\mathcal{D}_i|\ldots|\mathcal{D}_U|\}$. Moreover,

$$\mathbb{E}\left\{\left\|\sum_{i\in\mathcal{U}}\mathbb{1}_i(t)|\mathcal{D}_i|(\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t))\right\|^2\right\}$$
$$\leq \mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}|\mathcal{D}_i|^2\|\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t)\|^2. \tag{48}$$

Furthermore, we bound the term $\sum_{i\in\mathcal{U}}|\mathcal{D}_i|^2\|\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t)\|^2$ as follows:

$$\sum_{i\in\mathcal{U}}|\mathcal{D}_i|^2\|\widetilde{\boldsymbol{w}}_i(t)-\boldsymbol{w}(t)\|^2$$
$$\leq \sum_{i\in\mathcal{U}}|\mathcal{D}_i|^2\left\|\frac{\sum_{j\in\mathcal{U}}|\mathcal{D}_j|(\widetilde{\boldsymbol{w}}_i(t)-\widetilde{\boldsymbol{w}}_j(t))}{\sum_{j\in\mathcal{U}}|\mathcal{D}_j|}\right\|^2$$
$$\leq \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\frac{|\mathcal{D}_i|^2|\mathcal{D}_j|^2}{|\mathcal{D}|^2}\|\widetilde{\boldsymbol{w}}_i(t)-\widetilde{\boldsymbol{w}}_j(t)\|^2$$
$$+\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\frac{|\mathcal{D}_i|^2|\mathcal{D}_j|^2}{|\mathcal{D}|^2}\left(\|\widetilde{\boldsymbol{w}}_i(t)-\widehat{\boldsymbol{w}}(t)\|^2+\|\widetilde{\boldsymbol{w}}_j(t)-\widehat{\boldsymbol{w}}(t)\|^2\right)$$
$$\leq \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}\frac{|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Xi_i^2(\tau)+\Xi_j^2(\tau)\right)}{|\mathcal{D}|^2}, \tag{49}$$

where

$$\|\widetilde{\boldsymbol{w}}_i(t)-\widehat{\boldsymbol{w}}(t)\|$$
$$\leq \Theta_i(\tau) \triangleq \frac{4\eta Cq_i\sqrt{2\tau\ln(1/\delta_i)}}{\sqrt{\pi}b_i\epsilon_i}$$

$$+\eta C\sum_{m\in\mathcal{M}}\|p_{i,m}-q_m\|\sum_{j=0}^{\tau-1}\left(1+\eta\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\right)^j \tag{50}$$

derived by (24). Overall, we have

$$\mathbb{E}\left\{\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}(t)\|^2\right\}$$
$$\leq \frac{\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{|\mathcal{D}|^2|\mathcal{B}|^2D_{\min}^2} \tag{51}$$

and

$$\mathbb{E}\left\{F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\boldsymbol{w}(t))\right\}$$
$$\leq \frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{2|\mathcal{D}|^2|\mathcal{B}|^2D_{\min}^2}. \tag{52}$$

Then, we want to bound $\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|$ as

$$\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|$$
$$\leq \left\|\sum_{i\in\mathcal{U}}p_i(t)\boldsymbol{w}_i^{\tau-1}(t-1)-\widehat{\boldsymbol{w}}^{\tau-1}(t-1)\right\|$$
$$+\eta\left\|\sum_{i\in\mathcal{U}}p_i(t)\sum_{m\in\mathcal{M}}p_{i,m}\nabla F_{i,m}(\boldsymbol{w}_i^{\tau-1}(t-1))\right.$$
$$\left.-\sum_{i\in\mathcal{U}}p_i\sum_{m\in\mathcal{M}}p_{i,m}\nabla F_{i,m}(\widehat{\boldsymbol{w}}^{\tau-1}(t-1))\right\|. \tag{53}$$

To establish the Lipschitz continuous, it suffices to have $\|\nabla F_m(\boldsymbol{w})-\nabla F_m(\boldsymbol{w}')\| \leq \lambda_m\|\boldsymbol{w}-\boldsymbol{w}'\|$, and then we have

$$\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|$$
$$\leq \left\|\sum_{i\in\mathcal{U}}p_i(t)\boldsymbol{w}_i^{\tau-1}(t-1)-\widehat{\boldsymbol{w}}^{\tau-1}(t-1)\right\|$$
$$+\eta\sum_{i\in\mathcal{U}}p_i(t)\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\left\|\boldsymbol{w}_i^{\tau-1}(t-1)-\widehat{\boldsymbol{w}}^{\tau-1}(t-1)\right\|$$
$$+\eta\left\|\sum_{i\in\mathcal{U}}(p_i-p_i(t))\nabla F_i(\widehat{\boldsymbol{w}}^{\tau-1}(t-1))\right\| \tag{54}$$

Using triangle inequality, we have

$$\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|$$
$$\leq \sum_{i\in\mathcal{U}}p_i(t)\left\|\boldsymbol{w}_i^{\tau-1}(t-1)-\widehat{\boldsymbol{w}}^{\tau-1}(t-1)\right\|$$
$$+\eta\sum_{i\in\mathcal{U}}p_i(t)\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\left\|\boldsymbol{w}_i^{\tau-1}(t-1)-\widehat{\boldsymbol{w}}^{\tau-1}(t-1)\right\|$$
$$+\eta\left\|\sum_{i\in\mathcal{U}}(p_i-p_i(t))\right\|\left\|\nabla F_i(\widehat{\boldsymbol{w}}^{\tau-1}(t-1))\right\|. \tag{55}$$

Substituting (24) into (55), we can obtain

$$\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|$$
$$\leq \|\boldsymbol{w}(t-1)-\widehat{\boldsymbol{w}}(t-1)\|\sum_{i\in\mathcal{U}}p_i(t)\left(1+\eta\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\right)^{\tau-1}$$
$$+\eta C\left\|\sum_{i\in\mathcal{U}}(p_i-p_i(t))\right\|\eta C\sum_{i\in\mathcal{U}}p_i(t)\left(1+\eta\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\right)$$

$$\sum_{m\in\mathcal{M}}\|p_{i,m}-q_m\|\sum_{j=0}^{\tau-2}\left(1+\eta\sum_{m\in\mathcal{M}}p_{i,m}\lambda_m\right)^j$$

$$+\frac{4\eta C\sqrt{2\tau}}{\sqrt{\pi}}\sqrt{\sum_{i\in\mathcal{U}}\frac{p_i^2 q_i^2\ln(1/\delta_i)}{b_i^2\epsilon_i^2}}\triangleq\Xi(\tau). \quad (56)$$

We assume that $\boldsymbol{w}(t-1)=\widehat{\boldsymbol{w}}(t-1)$ at the beginning of each communication round. We have

$$\mathbb{E}\{F(\boldsymbol{w}(t))-F(\widehat{\boldsymbol{w}}(t))\}\le\lambda_{\max}\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|\le\lambda_{\max}\Xi(\tau). \quad (57)$$

Because $\mathbb{E}\{F(\boldsymbol{w}(t))-F(\widehat{\boldsymbol{w}}(t))\}\le\lambda_{\max}\Xi(\tau)$, we have $\mathbb{E}\{F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\widehat{\boldsymbol{w}}(t))\}\le\widehat{\Xi}(\tau)$, where

$$\widehat{\Xi}(\tau)$$
$$=\frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{2|\mathcal{D}|^2|\mathcal{B}|^2 D_{\min}^2}$$
$$+\lambda_{\max}\Xi(\tau). \quad (58)$$

According to [18], [44], we can obtain

$$\mathbb{E}\{F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\boldsymbol{w}^\star)\}\le\frac{1}{T\left(\eta\varphi-\frac{\rho\Xi(\tau)+\max_{t\in\mathcal{T}}\widehat{\Xi}_t(\tau)}{\tau\varepsilon_0^2}\right)}, \quad (59)$$

where $\rho\triangleq\omega\left(1-\frac{\lambda_{\max}}{2}\right)$, $\omega\triangleq\min_{t\in\mathcal{T}}\frac{1}{\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}^\star\|}$ and $\varepsilon_0\triangleq\min_{t\in\mathcal{T}}\frac{1+\sqrt{1+4\eta\rho T^2\tau(\rho\Xi(\tau)+|\mathcal{B}(t)|)}}{2\eta\rho T\tau}$. This completes the proof. □

APPENDIX E
PROOF OF THEOREM 6

First, using the $\lambda_{\max}$-Lipschitz smoothness, we can obtain the following inequality:

$$F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\widehat{\boldsymbol{w}}(t))$$
$$\le\lambda_{\max}\left(\|\boldsymbol{w}^{\mathcal{B}}(t)-\boldsymbol{w}(t)\|^2+\|\boldsymbol{w}(t)-\widehat{\boldsymbol{w}}(t)\|^2\right). \quad (60)$$

Substituting (51) and (55) into (60), we can obtain

$$F(\boldsymbol{w}^{\mathcal{B}}(t))-F(\widehat{\boldsymbol{w}}(t))$$
$$\le\frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{|\mathcal{D}|^2|\mathcal{B}|^2 D_{\min}^2}$$
$$+\lambda_{\max}\Xi(\tau). \quad (61)$$

Due to the $\lambda_{\max}$-Lipschitz smoothness, we can obtain

$$F(\widehat{\boldsymbol{w}}(t))-F(\widehat{\boldsymbol{w}}(t-1))\le-\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)\sum_{j=0}^{\tau-1}\nabla F\left(\widehat{\boldsymbol{w}}^j(t-1)\right). \quad (62)$$

We can note that $\boldsymbol{w}^{\mathcal{B}}(t-1)=\widehat{\boldsymbol{w}}(t-1)$ at the beginning of each communication round. Thus, we have

$$F\left(\boldsymbol{w}^{\mathcal{B}}(t)\right)-F\left(\boldsymbol{w}^{\mathcal{B}}(t-1)\right)$$
$$\le F\left(\boldsymbol{w}^{\mathcal{B}}(t)\right)-F\left(\widehat{\boldsymbol{w}}(t)\right)+F\left(\widehat{\boldsymbol{w}}(t)\right)-F\left(\widehat{\boldsymbol{w}}(t-1)\right)$$
$$\le\frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{|\mathcal{D}|^2|\mathcal{B}|^2 D_{\min}^2}$$

$$+\lambda_{\max}\Xi(\tau)-\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)\sum_{j=0}^{\tau-1}\nabla F\left(\widehat{\boldsymbol{w}}^j(t-1)\right). \quad (63)$$

Now summing above equation over $t=1,\dots,T$ and rearranging the terms yield that

$$\frac{1}{T}\sum_{t=1}^T\sum_{j=0}^{\tau-1}\nabla F\left(\widehat{\boldsymbol{w}}^j(t-1)\right)\le\frac{F(\boldsymbol{w}(0))-F(\boldsymbol{w}^\star)}{T\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)}$$
$$+\frac{\lambda_{\max}\mathbb{P}_{\max}(1-\mathbb{P}_{\max})\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}}|\mathcal{D}_i|^2|\mathcal{D}_j|^2\left(\Theta_i^2(\tau)+\Theta_j^2(\tau)\right)}{\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)|\mathcal{D}|^2|\mathcal{B}|^2 D_{\min}^2}$$
$$+\frac{\lambda_{\max}\Xi(\tau)}{\eta\left(1-\frac{\eta\lambda_{\max}}{2}\right)}. \quad (64)$$

This completes the proof. □

REFERENCES

[1] J. Li, S. Chu, F. Shu, J. Wu, and D. N. K. Jayakody, "Contract-based small-cell caching for data disseminations in ultra-dense cellular networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1042–1053, May 2019.

[2] S. Shaham, M. Ding, B. Liu, S. Dang, Z. Lin, and J. Li, "Privacy preservation in location-based services: A novel metric and attack model," *IEEE Trans. Mobile Comput.*, vol. 20, no. 10, pp. 3006–3019, Oct. 2021.

[3] S. Shaham, M. Ding, B. Liu, S. Dang, Z. Lin, and J. Li, "Privacy preserving location data publishing: A machine learning approach," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 9, pp. 3270–3283, Sep. 2021.

[4] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proc. ACM Conf. Comput. Commun. Secur. (CCS)*, Vienna, Austria, Oct. 2016, pp. 308–318.

[5] D. C. Nguyen *et al.*, "Enabling AI in future wireless networks: A data life cycle perspective," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 553–595, 1st Quart., 2021.

[6] Z. Liu, L. Zhang, W. Ni, and I. B. Collings, "Uncoordinated pseudonym changes for privacy preserving in distributed networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 6, pp. 1465–1477, Jun. 2020.

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

[8] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 72–80, Feb. 2020.

[9] L. U. Khan *et al.*, "Federated learning for edge networks: Resource optimization and incentive mechanism," *IEEE Commun. Mag.*, vol. 58, no. 10, pp. 88–93, Oct. 2020.

[10] C. Ma *et al.*, "On safeguarding privacy and security in the framework of federated learning," *IEEE Netw.*, vol. 34, no. 4, pp. 242–248, Jul./Aug. 2020.

[11] X. Deng, J. Li, L. Shi, Z. Wei, X. Zhou, and J. Yuan, "Wireless powered mobile edge computing: Dynamic resource allocation and throughput maximization," *IEEE Trans. Mobile Comput.*, early access, Oct. 28, 2020, doi: 10.1109/TMC.2020.3034479.

[12] K. Guo, R. Gao, W. Xia, and T. Q. S. Quek, "Online learning based computation offloading in MEC systems with communication and computation dynamics," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1147–1162, Feb. 2021.

[13] X. Xu, M. Tao, and C. Shen, "Collaborative multi-agent multi-armed bandit learning for small-cell caching," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2570–2585, Apr. 2020.

[14] G. Zhu, Y. Du, D. Gunduz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2120–2135, Mar. 2021.

[15] Z. Wang, T. Alpcan, J. S. Evans, and S. Dey, "Truthful mechanism design for wireless powered network with channel gain reporting," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7966–7979, Nov. 2019.

[16] Z. Xiong, Y. Zhang, D. Niyato, R. Deng, P. Wang, and L. Wang, "Deep reinforcement learning for mobile 5G and beyond: Fundamentals, applications, and challenges," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 44–52, Jun. 2019.

[17] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317–333, Jan. 2020.

[18] S. Wang *et al.*, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 1205–1221, Jun. 2019.

[19] 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 Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[20] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, "Communication-efficient federated learning," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 17, Apr. 2021, Art. no. e2024789118.

[21] M. Chen, H. V. Poor, W. Saad, and S. Cui, "Convergence time optimization for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2457–2471, Apr. 2021.

[22] X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu, and E. Dutkiewicz, "Optimal online data partitioning for geo-distributed machine learning in edge of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2393–2406, Oct. 2019.

[23] W. Xia, T. Q. S. Quek, K. Guo, W. Wen, H. H. Yang, and H. Zhu, "Multi-armed bandit-based client scheduling for federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7108–7123, Nov. 2020.

[24] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya, "An efficiency-boosting client selection scheme for federated learning with fairness guarantee," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1552–1564, Jul. 2021.

[25] A. Triastcyn and B. Faltings, "Federated learning with Bayesian differential privacy," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 2587–2596.

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

[27] K. Wei *et al.*, "User-level privacy-preserving federated learning: Analysis and performance optimization," *IEEE Trans. Mobile Comput.*, early access, Feb. 4, 2021, doi: 10.1109/TMC.2021.3056991.

[28] M. Seif, R. Tandon, and M. Li, "Wireless federated learning with local differential privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 2604–2609.

[29] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 170–185, Jan. 2021.

[30] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[31] M. Naseri, J. Hayes, and E. De Cristofaro, "Local and central differential privacy for robustness and privacy in federated learning," 2020, *arXiv:2009.03561*.

[32] S. Wang *et al.*, "Local differential private data aggregation for discrete distribution estimation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 9, pp. 2046–2059, Sep. 2019.

[33] Z. Huang, R. Hu, Y. Guo, E. Chan-Tin, and Y. Gong, "DP-ADMM: ADMM-based distributed learning with differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1002–1012, 2019.

[34] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6G communications: Challenges, methods, and future directions," *China Commun.*, vol. 17, no. 9, pp. 105–118, Sep. 2020.

[35] M. J. Neely, *Stochastic Network Optimization With Application to Communication and Queueing Systems* (Synthesis Lectures on Communication Networks). Morgan & Claypool, 2010.

[36] S. Kang and C. Joo, "Low-complexity learning for dynamic spectrum access in multi-user multi-channel networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Apr. 2018, pp. 1367–1375.

[37] D. Kalathil, N. Nayyar, and R. Jain, "Decentralized learning for multiplayer multiarmed bandits," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2331–2345, Apr. 2014.

[38] R. Jonker and T. Volgenant, "Improving the Hungarian assignment algorithm," *Oper. Res. Lett.*, vol. 5, no. 4, pp. 171–175, Oct. 1986.

[39] J. A. Bondy and U. S. R. Murty, *Graph Theory With Applications*. New York, NY, USA: Macmillan, 1976.

[40] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.

[41] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," M.S. thesis, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2009. [Online]. Available: http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

[42] F. Li, J. Liu, and B. Ji, "Combinatorial sleeping bandits with fairness constraints," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1799–1813, Jul. 2020.

[43] D. Pollard, *Convergence of Stochastic Processes* (Springer Series in Statistics). New York, NY, USA: Springer, 1984.

[44] W. Shi, S. Zhou, Z. Niu, M. Jiang, and L. Geng, "Joint device scheduling and resource allocation for latency constrained wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 453–467, Jan. 2021.

Kang Wei (Graduate Student Member, IEEE) received the B.S. degree in information engineering from Xidian University, Xian, China, in 2014, and the M.Sc. degree from the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include data privacy and security, differential privacy, machine learning, information theory, and channel coding theory in NAND flash memory.

Jun Li (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. From January 2009 to June 2009, he worked with the Department of Research and Innovation, Alcatel Lucent Shanghai Bell, as a Research Scientist. From June 2009 to April 2012, he was a Post-Doctoral Fellow at the School of Electrical Engineering and Telecommunications, University of New South Wales, Australia. From April 2012 to June 2015, he was a Research Fellow at the School of Electrical Engineering, The University of Sydney, Australia. Since June 2015, he has been a Professor at the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. He was a Visiting Professor at Princeton University from 2018 to 2019. His research interests include network information theory, game theory, distributed intelligence, multiple agent reinforcement learning, and their applications in ultra-dense wireless networks, mobile edge computing, network privacy and security, and the Industrial Internet of Things. He has coauthored more than 200 papers in IEEE journals and conferences, and holds one U.S. patent and more than ten Chinese patents in these areas. He was serving as a TPC member for several flagship IEEE conferences. He received the Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2018 and the Best Paper Award from IEEE International Conference on 5G for Future Wireless Networks in 2017. He was serving as an Editor for IEEE COMMUNICATIONS LETTERS.

Chuan Ma (Member, IEEE) received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2013, and the Ph.D. degree in telecommunication from The University of Sydney, in 2018. He is currently working as a Lecturer at the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. He has published more than ten transactions and conference papers, including the Best Paper in WCNC 2018. His research interests include stochastic geometry, device-to-device communication, wireless caching networks, and machine learning. He is currently working on the big data analysis and privacy preservation in distributed learning.

Ming Ding (Senior Member, IEEE) received the B.S. and M.S. degrees (Hons.) in electronics engineering and the Doctor of Philosophy (Ph.D.) degree in signal and information processing from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2004, 2007, and 2011, respectively. From April 2007 to September 2014, he worked at the Sharp Laboratories of China, Shanghai, China, as a Researcher/a Senior Researcher/a Principal Researcher. Currently, he is a Senior Research Scientist at the Data61, CSIRO, Sydney, NSW, Australia. His research interests include information technology, data privacy and security, machine learning, and AI. He has authored over 140 papers in IEEE journals and conferences, all in recognized venues, and around 20 3GPP standardization contributions, as well as a Springer book *Multi-point Cooperative Communication Systems: Theory and Applications*. Also, he holds 21 U.S. patents and co-invented another more than 100 patents on 4G/5G technologies in CN, JP, KR, and EU. Currently, he is an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. Besides, he has served as a guest editor/the co-chair/the co-tutor/a TPC member for multiple IEEE top-tier journals/conferences and received several awards for his research work and professional services.

Cailian Chen (Member, IEEE) received the B.Eng. and M.Eng. degrees in automatic control from Yanshan University, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Hong Kong, in 2006.

She has been with the Department of Automation, Shanghai Jiao Tong University, since 2008. She is now a Distinguished Professor. Her research interests include industrial wireless networks and computational intelligence and Internet of Vehicles. She has authored three research monographs and over 100 refereed international journal articles. She is the inventor of more than 30 patents.

Dr. Chen received the prestigious IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2008 and five conference best paper awards. She won the Second Prize of the National Natural Science Award from the State Council of China in 2018, the First Prize of Natural Science Award from the Ministry of Education of China in 2006 and 2016, respectively, and the First Prize of Technological Invention of Shanghai Municipal, China, in 2017. She was honored "National Outstanding Young Researcher" by NSF of China in 2020 and "Changjiang Young Scholar" in 2015. She has been actively involved in various professional services. She also served as the TPC Chair for ISAS'19, the Symposium TPC Co-Chair for IEEE GLOBECOM 2016, the Track Co-Chair for VTC2016-Fall and VTC2020-Fall, the Workshop Co-Chair for WiOpt'18, and a Guest Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. She serves as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *IET Cyber-Physical Systems: Theory and Applications*, and *Peer-to-Peer Networking and Applications* (Springer).

Shi Jin (Senior Member, IEEE) received the B.S. degree in communications engineering from the Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently a Faculty with the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He and his coauthors have been awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and the 2010 Young Author Best Paper Award by the IEEE Signal Processing Society. He serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, and *IET Communications*.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer at JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an Assistant Professor at Boise State University, Idaho. Currently, he is a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as with the Computer Science Department, University of Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid.

Dr. Han received the NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the Best Paper Award for the *EURASIP Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications systems (best paper award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS) in 2016, and several best paper awards in IEEE conferences. He was an IEEE Communications Society Distinguished Lecturer (from 2015 to 2018), an AAAS fellow (since 2019), and an ACM Distinguished Member (since 2019). Since 2017, he has been 1% highly cited researcher according to Web of Science. He is also the Winner of 2021 IEEE Kiyo Tomiyasu Award for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."
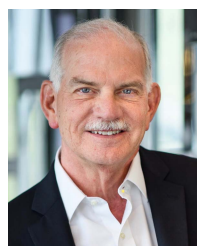
H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University in 1977.

From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor. From 2006 to 2016, he also served as the Dean of Princeton's School of Engineering and Applied Science. His research interests are in the areas of information theory, machine learning and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book *Machine Learning and Wireless Communications* (Cambridge University Press).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and a foreign member of the Chinese Academy of Sciences and the Royal Society, and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.