Client Selection and Bandwidth Allocation for Federated Learning: An Online Optimization Perspective

Yun Ji\textsuperscript{1}, Zhourbin Kou\textsuperscript{1}, Xiaoxiong Zhong\textsuperscript{1,2}, Hangfan Li\textsuperscript{1}, Fan Yang\textsuperscript{1}, and Sheng Zhang\textsuperscript{1,*}
\textsuperscript{1}Graduate School in Shenzhen, Tsinghua University, Shenzhen, 518055, China
\textsuperscript{2}Peng Cheng Laboratory, Shenzhen 518000, P.R. China
*Corresponding author: Sheng Zhang, email: zhangsh@sz.tsinghua.edu.cn

Abstract—Federated learning (FL) can train a global model from clients’ local data set, which can make full use of the computing resources of clients and performs more extensive and efficient machine learning on clients with protecting user information requirements. Many existing works have focused on optimizing FL accuracy within the resource constrained in each individual round, however there are few works comprehensively consider the optimization for latency, accuracy and energy consumption over all rounds in wireless federated learning. Inspired by this, in this paper, we investigate FL in wireless networks where client selection and bandwidth allocation are two crucial factors which significantly affect the latency, accuracy and energy consumption of clients. We formulate the optimization problem as a mixed-integer problem, which is to minimize the cost of time and accuracy within the long-term energy constrained over all rounds. To address this optimization problem, we propose a per-round energy drift plus cost (PEDPC) algorithm from an online perspective, and the performance of the PEDPC algorithm is verified in simulation results in terms of latency, accuracy and energy consumption in IID and NON-IID data distributions.

Index Terms—Federated learning, client selection, bandwidth allocation, wireless networks.

I. INTRODUCTION

Federated learning (FL), which trains machine learning (ML) models on clients, has been regarded as one of the most promising distributed ML techniques for its high secrecy [1].

However, to deploy FL on wireless networks, there are many problems to be addressed: 1) The computing and communication capabilities of different clients are not identical, hence some stragglers will tremendously prolong the latency of each learning round [2]. 2) The total bandwidth of all clients are finite, hence some clients with bad communication condition will lose their data packages or become stragglers. 3) For each client, the sum energy of all learning rounds is limited, hence to solve the FL optimization problem should take all the rounds into consideration, such that the problem is a long-term problem.

To address the above problems, it is essential to design an jointly optimal clients selection and bandwidth allocation method. However most existing works mainly focus on improving FL accuracy in individual round while putting less emphasis on comprehensively considering the latency, energy consumption and accuracy under all rounds [3]–[5]. Particularly, early effort on long-term FL optimization is presented in [6]. The authors maximize the linear function which is empirically proportional to the accuracy within the energy constraint, but the latency of FL isn’t considered in their paper.

In this paper, we consider the FL wireless networks subject to long-term energy constraint. To minimize the cost of time and accuracy, we jointly optimize the client selection and bandwidth allocation processes. The main contributions are summarized as follows:

- We propose a per-round energy drift plus cost (PEDPC) algorithm based on Lyapunov theory to transform the long-term energy constraint problem to a series of online problems. The PEDPC algorithm can achieve an $O(1/V, \sqrt{V})$ cost-energy tradeoff where $V$ is an algorithm parameter.
- To solve the mixed-integer problem introduced by the online procedures, we propose an increasing time-maximum client selection (ITMCS) algorithm.
- We consider the high heterogeneity of FL in the experiments, and the performance of our algorithms is verified by extensive simulations.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a FL system as shown in Fig. 1, which consists of a cloud server and $K$ clients indexed by the set $\mathcal{K} = \{1, \ldots, K\}$. In each FL round, the cloud server firstly selects a set of clients and distributes the global model to them; then the clients train the local models on their own datasets and upload the models to the cloud server; finally the cloud server aggregates the uploaded models and updates the global model. Each client $k \in \mathcal{K}$ has a local dataset $\mathcal{D}_k$ with size of $D_k \in \mathbb{R}_+$. We assume that there are total $R$ rounds in the FL process. In the $r$-th round, we adopt $x_k(r)$ to denote whether the client $k$ is selected (if the client $k$ is selected then $x_k(r) = 1$, otherwise $x_k(r) = 0$), where $r \leq R$. We define $\mathbf{x}(r) \triangleq [x_1(r), \ldots, x_K(r)]^T$ denoting the overall selection strategy. Then we assume that the total bandwidth allocated to all clients is $B$ MHz, and the proportion allocated to the client $k$ is $b_k(r)$, where $\sum_{k=1}^K b_k(r) = 1$. We also define $\mathbf{b}(r) \triangleq [b_1(r), \ldots, b_K(r)]^T$ denoting the overall bandwidth allocation strategy. In practical wireless networks, the bandwidth allocated to each client generally cannot be 0 Hz.
due to the finite resource block size [6], so we let $b_k(r) \geq b_{\text{min}}$ for $r = 1, \ldots, R$.

### A. Energy Consumption Model

In each FL round, the local training process leads to the consumption of computational energy. We denote the CPU frequency of client $k$ as $f_k$, and the number of CPU cycles for training one bit data as $c_k$. We assume that client $k$ conducts totally $U_k(r)$ training iterations in the $r$-th round. Then the computational energy consumption can be expressed as

$$E_k^{\text{cmp}} = U_k(r)\delta_k c_k D_k f_k^2$$

where $\delta_k$ is the effective capacitance coefficient of computing chipset for client $k$ [3]. After training, the selected clients upload their model parameters to the server. The transmission rate is

$$R_k(r) = b_k(r) B \log \left(1 + \frac{p_k(r) h_k^2(r)}{N_0}\right)$$

where $p_k(r)$ is the transmission power of client $k$ and $N_0$ is the power of the additive white Gaussian noise. $h_k(r)$ denotes the channel gain between client $k$ and the server. We let $S_k$ denote the data size of model parameters, then the communication latency and the communication energy can be calculated by $T_k^{\text{com}}(r) = S_k/R_k(r)$ and $E_k^{\text{com}}(r) = p_k(r)S_k/R_k(r)$, respectively. Therefore, the total energy consumption is given by

$$E_k(r) = E_k^{\text{cmp}}(r) + E_k^{\text{com}}(r)$$

### B. Latency Model

We let $T_k^{\text{cmp}}(r)$ denote the time consumed by client $k$ on training local model, which can be expressed as

$$T_k^{\text{cmp}}(r) = U_k(r)\frac{c_k D_k}{f_k}$$

Then the total latency of client $k$ in the $r$-th round is given by

$$T_k(r) = T_k^{\text{com}}(r) + T_k^{\text{cmp}}(r)$$

For each round, the total latency depends on the maximum latency of all selected clients, such that

$$T_0(x(r), b(r)) = \max_{k=1, \ldots, K} \{x_k(r)T_k(r)\}$$

where $T_0(x(r), b(r))$ denotes the latency of the $r$-th round.

### C. Accuracy and Cost Model

According to [4], FL accuracy depends on the data size of selected clients, such that

$$\Phi(x(r)) = \sum_{k=1}^{K} \log [1 + \mu D_k x_k(r)]$$

where $\Phi(x(r))$ denotes the FL accuracy of the $r$-th round, and $\mu$ is the system global parameter. In this work, we denote cost function as follows:

$$y_0((x(r), b(r))) = T_0(x(r), b(r)) - \Phi(x(r))$$

### D. Problem Formulation

A joint problem of client selection and bandwidth allocation is considered in this subsection, which minimizes the rounds average cost function with energy constraint of each client. The problem is formulated as follows:

$$\text{P1} : \min_{x(0), b(0), \ldots, x(R-1), b(R-1)} \frac{1}{R} \sum_{r=0}^{R-1} y_0(x(r), b(r))$$

s.t. $\sum_{r=0}^{R-1} x_k(r)E_k(r) \leq H_k, \forall k$ \hspace{1cm} (10)

$b_k(r) \geq b_{\text{min}}, \forall k, \forall r$ \hspace{1cm} (11)

$\sum_{k=1}^{K} b_k(r) = 1, \forall r$ \hspace{1cm} (12)

$x_k(r) \in \{1, 0\}, \forall k, \forall r$ \hspace{1cm} (13)

Constraint (10) guarantees that the energy consumption of each client is limited by $H_k$. Constraint (11) ensures that the bandwidth of each client is lower bounded by $b_{\text{min}}$. Constraint (12) indicates that the total bandwidth equals to $B$. Constraint (13) indicates the selection decisions.

To solve P1, firstly we divide $R$ rounds into $F$ frames, and each frame includes $L = R/F$ rounds, then we transform P1 to a series of P2 for $f = 1, 2, \ldots, F$.

$$\text{P2} : \min_{x(r), b(r)} c_f \triangleq \frac{1}{L} \sum_{r=0}^{L-1} y_0(x(r), b(r))$$

s.t. $\sum_{r=0}^{L-1} x_k(r)E_k(r) \leq H_k/F, \forall k$ \hspace{1cm} (14)

Constraints (11), (12), (13)

We denote $c_f^*$ as the optimal value of P2 by L-round lookahead algorithm which solve P2 by foreseeing full offline knowledge over the frame [7].

### III. ONLINE OPTIMIZATION ALGORITHMS

In this section, we propose the PEDPC algorithm to turn P2 to a series of online problems and develop the Iterative Algorithm to solve these problems. Specifically, we propose the ITMCS algorithm to select clients, and adopt the Barrier Method algorithm to allocate bandwidth to the selected clients in each round.
Assume that \( y \) only requires the current energy deficit and channel state to make decisions without foreseeing any information. In Algorithm \( D \) we denote the energy drift function as \( \mathbf{z} = [z_1(r), ..., z_k(r)]^T \), where \( z_k(r) \) is the queue backlog of client \( k \) in the \( r \)-th round, and represents the energy deficit of client \( k \) over \( r \) rounds. \( z_k(r) \) is updated as follows:

\[
  z_k(r + 1) = \max[z_k(r) + x_k(r)E_k(r) - H_k/R, 0]
\]

(16) represents the queue congestion in the network. Besides, we denote the energy drift function as \( Y(Z(r)) = Z_k(r)(r + 1) - Y(Z(r)) \), which is upper-bounded by

\[
  Y(Z(r + 1)) - Y(Z(r)) \\
  \leq \frac{1}{2} \sum_{k=1}^{K} \left[ x_k(r)E_k(r) - \frac{H_k}{R} + z_k(r) \right] \left[ x_k(r)E_k(r) - \frac{H_k}{R} \right] \\
  \leq D + \sum_{k=1}^{K} z_k(r) \left[ x_k(r)E_k(r) - \frac{H_k}{R} \right]
\]

(18)

where \( D = \frac{1}{2} \sum_{k=1}^{K} \max[(y_k^{\min})^2, (y_k^{\max})^2] \).

We now present PEDPC in Algorithm 1. In each round, PEDPC only requires the current energy deficit and channel state to make decisions without foreseeing any information. In each round, we aim to solve the problem:

\[
P_3 : \min_{\mathbf{x}(r), \mathbf{b}(r)} \sum_{k=1}^{K} z_k(r) \left[ x_k(r)E_k(r) - \frac{H_k}{R} \right] \text{ subject to Constraints (11), (12), (13)}
\]

In \( P_3 \), by considering the right-hand side of the inequality (18) of the inequality, the system takes into account the energy deficit of the clients during the current round optimization. If client \( k \) is often selected, \( z_k \) will be large which incurs that reducing the energy consumption is more significant. Thus the energy deficit queue can guide the system towards meeting the energy constraint without foreseeing any information. \( V \) is a hyper-parameter to achieve the tradeoff between minimizing the cost function and energy consumption over \( R \) rounds, and more details will be discussed in simulation results. As a consequence, PEDPC decouples the energy constraint (15) and turns \( P_2 \) to a series of per-round problems.

Algorithm 2 Iterative Algorithm

Input:
A feasible solution \( x^0(r), b^0(r) \) of \( P_3, i = 0 \) and \( I > 0 \);

1: \textbf{repeat}
2: With given \( b^0(r) \), figure out the optimal \( x^{i+1}(r) \) of \( P_4 \) through Algorithm 3;
3: With given \( x^{i+1}(r) \), figure out the optimal \( b^{i+1}(r) \) of \( P_5 \) through the Barrier method algorithm;
4: set \( i = i + 1; \)
5: until \( i \geq I \)

B. Iterative Algorithm

Notice that \( P_3 \) is a mixed-integer problem and there is no polynomial-time algorithm. So we develop the Iterative Algorithm in Algorithm 2 with low complexity to solve \( P_3 \).

Iterative Algorithm contains two steps at each iteration: 1) we fix \( b(r) \) and optimize \( x(r) \) in \( P_3 \) which is a 0-1 integer problem given in \( P_4 \); 2) we optimize \( b(r) \) in \( P_3 \) with updated \( x(r) \) by step 1), which is a convex problem given in \( P_5 \). The Iterative Algorithm guarantees that the objective value of \( P_3 \) is non-increasing in each step and always converges to a local optimal solution since it’s lower bounded by zero when we choose no client in the \( r \)-th round.

Fixing \( b(r) \) and plugging (6), (7) and (8) into \( P_3 \), we have

\[
P_4 : \min_{\mathbf{x}(r)} \sum_{k=1}^{K} (Z_k(r)x_k(r) - V \log (1 + v_k x_k(r))) \quad \text{s.t. Constraints (13)}
\]

where \( Z_k(r) = Z_k(r) \left( \frac{E_{cmp}(r)}{b_k(r) + S_k(r)} \right) \).

We propose ITMCS in Algorithm 3 to efficiently solve \( P_4 \).

In ITMCS, we let \( q_k(r) = Z_k(r)x_k(r) - V \log (1 + v_k x_k(r)) \), and denote a set \( S_0 \) collecting all clients which satisfy \( q_k(r) < 0 \). We sort \( S_0 \) in time latency ascending order and add clients into the selection set \( S \) sequentially until we loop the entire \( S_0 \). Finally, the implemented optimal selection is \( S^* = \arg \min_{S \subseteq S_0} \sum_{k \in S} q_k \). We assume there are \( m \) clients being selected, i.e., \( |S^*| = m \).

With fixed \( x(r) \), replacing \( y_0(x(r), b(r)) \) by \( \ln \left( \sum_{k \in S^*} \exp(T_k(r)) \right) - \Phi(x(r)) \) and plugging (5), (7) into (19) yields:

\[
P_5 : \min_{\mathbf{b}(r)} \sum_{k \in S^*} \exp(T_k(r) + \frac{S_k(r) - G_k(r)}{b_k(r)}) \quad \text{s.t. Constraints (11), (12)}
\]
where \( S_k'(r) = S_k/G_k(r), G_k'(r) = p_k(r)Z_k(r)S_k/G_k(r) \). Clearly, the objective function (21) is convex according to the vector composition theorem [8]. Hence the problem P5 is convex. We adopt Barrier Method [8, Chapter 11] to solve it.

C. Algorithm Performance and Time Complexity Analysis

In this section, we prove the performance guarantee and analyze the time complexity of PEDPC.

According to (16), we have

\[
Z_k(R) - Z_k(0) = \sum_{r=0}^{R-1} Z_k(r + 1) - Z_k(r)
\]

\[
= \sum_{r=0}^{R-1} \max \left[ x_k(r)E_k(r) - \frac{H_k}{R}, -Z_k(r) \right]
\]

\[
\ge \sum_{r=0}^{R-1} \left( x_k(r)E_k(r) - \frac{H_k}{R} \right)
\]

\[
= \sum_{r=0}^{R-1} x_k(r)E_k(r) - H_k
\]

(22)

**Theorem 1:** The PEDPC returns a sub-optimal solution to P1 which is upper bounded by \( \frac{1}{\gamma} \sum_{r=0}^{F-1} c_f^* + 2DL + \frac{Y(Z(0))}{V R} \) and the total energy consumption of client \( k \) is upper bounded by \( H_k + \sqrt{2DRL + 2VL} \sum_{f=0}^{F-1} (c_f^* - y_0^{\min}) \).

**Proof.** According to [7, Sec 4.9.2], we have

\[
Y(Z(FL)) - Y(Z(0)) + \frac{FL-1}{2} \sum_{r=0}^{FL-1} y_0(x(r), b(r))
\]

\[
\le DL^2 F + VL \sum_{f=0}^{F-1} c_f^*(23)
\]

Dividing by \( VLF \), using the fact that \( Y(Z(FL)) \ge 0 \), \( R = FL \), and rearranging terms yields:

\[
\frac{1}{R} \sum_{r=0}^{FL-1} y_0(x(r), b(r)) \le \frac{1}{R} \sum_{f=0}^{F-1} c_f^* + \frac{DL}{V} + \frac{Y(Z(0))}{V R}
\]

(24)

When \( Z(0) = 0 \) the solution of P1 is within \( O(1/V) \) of the average \( c_f^* \) values where \( f = 1, 2, ..., F \). Then by rearranging terms of (23), we have

\[
Y(Z(FL)) - Y(Z(0)) \le DL^2 F + V \sum_{f=0}^{F-1} c_f^*(25)
\]

According to \( y_0(x(r), b(r)) \ge y_0^{\min} \), we have

\[
Y(Z(FL)) \le DL^2 F + VL \sum_{f=0}^{F-1} (c_f^* - y_0^{\min}) + Y(Z(0))(26)
\]

**Algorithm 3 ITMCS**

**Input:**

1. Set \( S_0 = \emptyset, S = S_0, S = \{S_0\} \);
2. Calculate \( q_k = Z_k'(r) - V \log (1 + v_k), \forall k \);
3. Find \( k \) to satisfy \( q_k < 0, \forall k \) and Update \( S_0 = S_0 \cup \{k\} \);
4. Rank the clients in \( S_0 \) according to \( T_k(r) \). Hence we have \( T_k(r) \le T_k(r) \); \( \le T_k[r_0](r) \);
5. for \( i \in S_0 \) do
6. Update \( S = S \cup \{k\} \), where \( T_k(r) \le T_k(r), \forall k \in S_0 \);
7. Update \( S = S \cup \{S\} \);
8. Calculate \( W(S) = VT_k(r) + \sum_{k \in S} q_k \);
9. Set \( S = \emptyset \);
10. end for
11. Find \( S^* = \arg \min_{S \subseteq S} W(S) \);
12. Return \( x^* \), where \( x_k^* = 1 \{k \in S^*\} \), \( \forall k \);

Due to \( Y(Z(FL)) = \frac{1}{2} \sum_{k=1}^{K} Z_k(FL)^2 \) we have

\[
Z_k(FL)^2 \le 2Y(Z(FL)) \le 2DL^2 F + 2VL \sum_{f=0}^{F-1} (c_f^* - y_0^{\min}) + 2Y(Z(0))(27)
\]

Dividing \( Z_k(FL) \) by \( FL \), we have

\[
\frac{Z_k(FL)}{FL} \le \sqrt{2D + \frac{2V \sum_{f=0}^{F-1} (c_f^* - y_0^{\min})}{F^2 L} + \frac{2Y(Z(0))}{F^2 L}}(28)
\]

According to (22), we have

\[
Z_k(FL) \ge \sum_{r=0}^{R-1} x_k(r)E_k(r) - H_k + Z_k(0)(29)
\]

Plugging (28) into (29) and assuming \( Y(Z(0)) = 0 \), we have

\[
\sum_{r=0}^{R-1} x_k(r)E_k(r) \le H_k + \sqrt{2DRL + 2VL} \sum_{f=0}^{F-1} (c_f^* - y_0^{\min})(30)
\]

Thus the energy consumption of each client is upper bounded by \( H_k + \sqrt{2DRL + 2VL} \sum_{f=0}^{F-1} (c_f^* - y_0^{\min}) \).

**Theorem 1** implies that PEDPC can achieve \( O(1/V, \sqrt{V}) \) cost-energy tradeoff in our system model.

According to [8], the time complexity of the Barrier Method is \( O \left( \log \frac{2m/(\epsilon_{2 \alpha}(0))}{\log \mu} \right) \). Obviously the complexity of Algorithm 2 is \( O(m) \), hence the time complexity of PEDPC is \( O \left( RLm \frac{\log \mu}{\log \mu^{1/2}} \right) \). We can find that the complexity grows linearly with the product of the number of clients and FL iterative rounds.
TABLE I
SIMULATION PARAMETERS

| Parameters                  | Value                  |
|-----------------------------|------------------------|
| Number of clients, $K$      | 100                    |
| Number of global rounds, $R$| 300                    |
| Number of local training iterative, $U_k(r)$ | 5                      |
| CPU cycles for training 1 bit data, $c_k$ | $1 \sim 10$ cycles/bit |
| CPU frequency, $f_k$        | $0.01 \sim 1$ GHz      |
| Effective switched capacitance, $\delta_k$ | $10^{-28}$          |
| System bandwidth, $B$       | 10 MHz                 |
| Square of channel gain, $h_k^2(r)$ | $10^{-9} \sim 10^{-11}$ |
| Transmission power, $p_k(r)$| $10 \sim 20$ dBm       |
| Noise power, $N_0$          | $10^{-13}$ W           |
| System parameter, $\mu$    | $1.7 \times 10^{-8}$   |
| Energy budget, $H_k$        | 1.5 J                  |

IV. SIMULATION RESULTS

A. Experiment Settings

We consider the hand-written digit classification task on MNIST dataset [9]. We apply a multi-layer perceptron which has two hidden layers with 10 hidden nodes each (data size of model parameters $S_B = 0.24$ Mbits). Besides, we consider two data distribution cases: 1) IID case, where the data of training set is shuffled and uniformly distributed over all clients. 2) NON-IID case, where the data size of each client is randomly distributed in $[1.2, 2.4, 3.6, 4.8, 6]$ Mbits. In addition, we sort the training data by labels and divide them into 300 groups, then each client chooses $1 \sim 5$ groups according to its own data size and at most contains five kinds of digit labels. FedAvg [1] is used as the learning algorithm in FL. The parameters setting is shown in TABLE I.

Furthermore, to verify the effectiveness of our proposed algorithm we consider the following four benchmarks:

- **Select all**: All clients are selected in each round, and the bandwidth is allocated uniformly.
- **Randomly**: For a given probability $Pr$, randomly select $Pr \times K$ clients in each round, and the bandwidth is allocated evenly for the selected clients.
- **Greedyly**: Maximize the number of chosen clients in each round within limited energy. Specifically, select clients by solving:

$$\max_{x(r), b(r)} \sum_{k=1}^{K} x_k(r) \quad (31)$$

$$\text{s.t. } x_k(r)E_k(r) \leq E_{max}, \forall k, \forall r \quad (32)$$

Constraints (11), (12), (13)

- **FedCS** [10]: Maximize the number of chosen clients in each round within limited energy.

B. Impact of $V$

In PEDPC, $V$ is used to control the key optimizing factors of $\mathbf{P3}$. In this section, we simulate the impact of $V$ on the average number of clients, latency and energy overflow in FL, where the energy overflow equals the energy consumption minus the energy maximum constraint.

C. Performance Comparison

To verify the performance of PEDPC, we experiment under IID and NON-IID cases.

Fig. 3 shows the energy overflow, latency and accuracy of PEDPC and benchmarks in IID case. For comparison purpose, we set the average client number of above algorithms to be 40 other than Select all. Compared with Greedily, PEDPC saves about $5.5 \sim 6.5$ times energy and latency, meanwhile achieves a higher accuracy. For selecting more clients, Greedily assigns each client the minimum bandwidth meeting the energy constraint which leads to an huge latency. Compared with Randomly, PEDPC saves about $5.8 \sim 8.0$ times energy and time latency while achieving a higher accuracy. This is because Randomly takes no account of the energy and time consumption in client selection and bandwidth allocation. Even though FedCS saves three times latency of our proposed algorithm, it uses $26.6$ times energy of ours. This is because FedCS takes no account of energy consumption. The Select all algorithm picks up all clients, and its energy consumption and latency are much larger than ours.

Fig. 4 shows the energy overflow, latency and accuracy of PEDPC and benchmarks in NON-IID case. For comparison purpose, we set the average number of clients 90 other than Select all. Because of the similar number of selected clients, the accuracy of all algorithms does not differ by more than 3 percents. Compared with Greedily and Randomly, the time latency of PEDPC achieves $33\%$ of above two algorithms under a similar energy consumption. This is because the bandwidth allocation of these two algorithms without regarding to the latency. As the number of clients increasing, the advantage of FedCS in saving time gradually diminishes, and the energy consumption is $4.5$ times of ours. Select all takes much more energy and time latency than ours.

According to the results from Fig. 3 and Fig. 4, PEDPC could achieve a good balance between energy consumption, latency and accuracy under both IID and NON-IID situations.

V. CONCLUSION

In this paper, we propose a joint optimization algorithm PEDPC with considering client selection and bandwidth allocation for wireless federated leaning. In PEDPC, we exploit the Lyapunov-based energy deficit queue to solve the optimization problem with ITMC algorithm from an online optimization perspective. Then we prove that PEDPC can achieve an $O(1/V, \sqrt{V})$ cost-energy tradeoff. The extensive simulations demonstrate that the proposed PEDPC algorithm can achieve a better performance than the other algorithms, i.e., Select all, Randomly, Greedily, and FedCS in terms of energy consumption, time latency and FL accuracy.
Fig. 2. The impact of V on the average number of clients, time latency and energy overflow

(a) Average number of clients
(b) Time latency
(c) Energy Overflow

Fig. 3. The energy overflow, time latency and test accuracy of PEDPC and benchmarks in IID case

(a) Energy overflow
(b) Time Latency
(c) Test accuracy

Fig. 4. The energy overflow, time latency and test accuracy of PEDPC and benchmarks in NON-IID case

(a) Energy overflow
(b) Time Latency
(c) Test accuracy

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. of the 20th International Conference on Artificial Intelligence and Statistics(AISTATS 2017). PMLR, Apr. 2017, pp. 1273–1282.

[2] 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, 2020.

[3] J. Zheng, K. Li, E. Tovar, and M. Guizani, “Federated learning for energy-balanced client selection in mobile edge computing,” in 2021 International Wireless Communications and Mobile Computing(IWCMC 2021). IEEE, Jun. 2021, pp. 1942–1947.

[4] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, “Energy efficient federated learning over wireless communication networks,” IEEE Trans. Wireless Commun., vol. 20, no. 3, pp. 1935–1949, Mar. 2021.

[5] C. T. Dinh, N. H. Tran, M. N. Nguyen, C. S. Hong, W. Bao, A. Y. Zomaya, and V. Gramoli, “Federated learning over wireless networks:

Convergence analysis and resource allocation,” IEEE/ACM Trans. Netw., vol. 29, no. 1, pp. 398–409, Feb. 2021.

[6] J. Xu and H. Wang, “Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective,” IEEE Trans. Wireless Commun., vol. 20, no. 2, pp. 1188–1200, Feb. 2021.

[7] M. Neely, “Stochastic network optimization with application to communication and queueing systems,” Synth. Lectures Commun. Netw., vol. 3, no. 1, pp. 1–211, Jan. 2010.

[8] S. Boyd, S. P. Boyd, and L. Vandenberghe, Convex optimization. Cambridge university press, 2004.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE Proc. IRE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[10] T. Nishio and R. Yonetani, “Client selection for federated learning with heterogeneous resources in mobile edge,” in Proc. IEEE International Conference on Communication(ICC 2019). IEEE, May. 2019, pp. 1–7.