SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Gradient Sparsification for Efficient Wireless Federated Learning with Differential Privacy

Kang Wei[1], Jun Li[1*], Chuan Ma[2,7], Ming Ding[4], Feng Shu[3,1], Haitao Zhao[5], Wen Chen[6] & Hongbo Zhu[5]

[1] *Nanjing University of Science and Technology, Nanjing* 210096*, China;*
[2] *Zhejiang Lab, Hangzhou* 311121*, China;*
[3] *Hainan University, Haikou* 570228*, China;*
[4] *CSIRO, Sydney* 2015*, Australia;*
[5] *Nanjing University of Posts and Telecommunications, Nanjing* 210003*, China;*
[6] *Shanghai Jiao Tong University, Shanghai* 200240*, China*
[7] *Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing* 211189*, China;*

Appendix A Adjusting the Clipping Value

Considering the sparsification process, we have $\boldsymbol{g}_i^t(\mathcal{D}_{i,m}) = \boldsymbol{g}_i^t(\mathcal{D}_{i,m}) \odot \boldsymbol{m}_i^t$, where $\mathcal{D}_{i,m}$ is the m-th sample of the i-th client, $\odot$ represents the element-wise product process, $\boldsymbol{m}_i^t$ is a binary mask vector and its element $\forall m_{i,k}^t \in \{0,1\}$, $k \in \{1,\dots,K\}$ and $K = |\boldsymbol{m}_i^t|$. Because probabilities of $m_{i,k}^t = 0$ and $m_{i,k}^t = 1$ are equal to s_i^t and $(1-s_i^t)$, respectively, we have

$$\begin{aligned}(\mathbb{E}[\|\boldsymbol{g}_i^t(\mathcal{D}_{i,m}) \odot \boldsymbol{m}_i^t\|])^2 &\leqslant \mathbb{E}[\|\boldsymbol{g}_i^t(\mathcal{D}_{i,m}) \odot \boldsymbol{m}_i^t\|^2] = \mathbb{E}\left[\sum_{k=1}^{K}(\boldsymbol{g}_{i,k}^t(\mathcal{D}_{i,m})m_{i,k}^t)^2\right] = \sum_{k'=0}^{K}\frac{k'\binom{K}{k'}}{K}(s_i^t)^{k'}(1-s_i^t)^{K-k'}\sum_{k=1}^{K}(g_{i,k}^t(\mathcal{D}_{i,m}))^2 \\ &= \sum_{k'=0}^{K}\binom{K-1}{k'-1}(s_i^t)^{k'}(1-s_i^t)^{K-k'}\sum_{k=1}^{K}(g_{i,k}^t(\mathcal{D}_{i,m}))^2 = s_i^t\|\boldsymbol{g}_i^t(\mathcal{D}_{i,m})\|^2.\end{aligned} \tag{A1}$$

Therefore, we can obtain $\mathbb{E}\left[\left\|\boldsymbol{g}_i^t(\mathcal{D}_{i,m}) \odot \boldsymbol{m}_i^t\right\|\right] \leqslant \sqrt{s_i^t}\left\|\boldsymbol{g}_i^t(\mathcal{D}_{i,m})\right\|$. This completes the proof. □

Appendix B Convergence Analysis

First, we define $\mathbb{1}_i^t \triangleq \sum_{j=1}^{N} a_{i,j}^t$ to denote whether the i-th client has been allocated to an available channel. We can note that

$$\boldsymbol{w}^{t+1} - \boldsymbol{w}^t = \sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \boldsymbol{w}_i^{t,\tau} - \boldsymbol{w}_i^{t,0} = -\eta \sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}(\nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\ell} \odot \boldsymbol{m}_i^t). \tag{B1}$$

Using the L-Lipschitz smoothness, we can obtain

$$\begin{aligned}&\mathbb{E}[F(\boldsymbol{w}^{t+1}) - F(\boldsymbol{w}^t)] \leqslant \mathbb{E}[\langle \nabla F(\boldsymbol{w}^t), \boldsymbol{w}^{t+1} - \boldsymbol{w}^t\rangle] + \frac{L}{2}\mathbb{E}[\|\boldsymbol{w}^{t+1} - \boldsymbol{w}^t\|^2] \\ &= \underbrace{-\mathbb{E}\left[\left\langle \nabla F(\boldsymbol{w}^t), \eta \sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}(\nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\ell} \odot \boldsymbol{m}_i^t)\right\rangle\right]}_{E_1} + \frac{\eta^2 L}{2} \underbrace{\mathbb{E}\left[\left\|\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}(\nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\ell} \odot \boldsymbol{m}_i^t)\right\|^2\right]}_{E_2}.\end{aligned} \tag{B2}$$

Then, we can rewrite E_1 as

$$E_1 = -\eta \sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\langle \nabla F(\boldsymbol{w}^t), \mathbb{E}[\nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t]\rangle - \eta \sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\langle \nabla F(\boldsymbol{w}^t), \mathbb{E}[\boldsymbol{n}_i^{t,\ell} \odot \boldsymbol{m}_i^t]\rangle. \tag{B3}$$

Because $\mathbb{E}[\boldsymbol{n}_i^{t,\ell}] = 0$ and $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \frac{1}{2}(\|\boldsymbol{y}\|^2 + \|\boldsymbol{x}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2)$, we have

$$E_1 = -\frac{\eta}{2}\sum_{\ell=0}^{\tau-1}\|\nabla F(\boldsymbol{w}^t)\|^2 - \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t\|^2] + \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F(\boldsymbol{w}^t) - \nabla F_i(\boldsymbol{w}_i^{t,\ell}) \odot \boldsymbol{m}_i^t\|^2]. \tag{B4}$$

* Corresponding author (email: jun.li@njust.edu.cn)

Further, we have

$$E_1 = -\frac{\eta}{2}\sum_{\ell=0}^{\tau-1}\|\nabla F(\boldsymbol{w}^t)\|^2 - \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\odot \boldsymbol{m}_i^t\|^2]$$
$$+\frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F(\boldsymbol{w}^t) - \nabla F_i(\boldsymbol{w}^t) + \nabla F_i(\boldsymbol{w}^t) - \nabla F_i(\boldsymbol{w}_i^{t,\ell}) + \nabla F_i(\boldsymbol{w}_i^{t,\ell}) - \nabla F_i(\boldsymbol{w}_i^{t,\ell})\odot \boldsymbol{m}_i^t\|^2]. \tag{B5}$$

Due to Jensen's inequality and (A1), we obtain

$$E_1 \leqslant -\frac{\eta}{2}\sum_{\ell=0}^{\tau-1}\|\nabla F(\boldsymbol{w}^t)\|^2 - \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2] + \frac{3\eta\tau}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \varepsilon_i$$
$$+\frac{3\eta L^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\underbrace{\mathbb{E}[\|\boldsymbol{w}^t - \boldsymbol{w}_i^{t,\ell}\|^2]}_{E_{11}} + \frac{3\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t (1-s_i^t)\sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2]. \tag{B6}$$

Then, we can bound E_{11} as

$$E_{11} = \mathbb{E}[\|\boldsymbol{w}^t - (\boldsymbol{w}_i^{t,\ell-1} - \eta\nabla F_i(\boldsymbol{w}_i^{t,\ell-1})\odot \boldsymbol{m}_i^t - \eta \boldsymbol{n}_i^{t,\ell-1}\odot \boldsymbol{m}_i^t)\|^2] = \eta^2\mathbb{E}[\|\sum_{\kappa=0}^{\ell-1}(\nabla F_i(\boldsymbol{w}_i^{t,\kappa})\odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\kappa}\odot \boldsymbol{m}_i^t)\|^2]$$
$$\leqslant \eta^2\ell\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\kappa})\odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\kappa}\odot \boldsymbol{m}_i^t\|^2] = \eta^2\ell s_i^t\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\kappa})\|^2] + \eta^2\ell s_i^t\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\boldsymbol{n}_i^{t,\kappa}\|^2]. \tag{B7}$$

Hence, we can obtain

$$E_1 \leqslant -\frac{\eta}{2}\sum_{\ell=0}^{\tau-1}\|\nabla F(\boldsymbol{w}^t)\|^2 + \frac{3\eta\tau}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \varepsilon_i + \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t (3-4s_i^t)\sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2]$$
$$+\frac{3\eta^3L^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\ell\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\kappa})\|^2] + \frac{3\eta^3L^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\ell\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\boldsymbol{n}_i^{t,\kappa}\|^2] \tag{B8}$$

and

$$E_2 \leqslant \tau\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\odot \boldsymbol{m}_i^t + \boldsymbol{n}_i^{t,\ell}\odot \boldsymbol{m}_i^t\|^2] = \tau\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2] + \tau\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\boldsymbol{n}_i^{t,\ell}\|^2]. \tag{B9}$$

Combining E_1 and E_2, we can obtain

$$\mathbb{E}[F(\boldsymbol{w}^{t+1}) - F(\boldsymbol{w}^t)] \leqslant -\frac{\eta}{2}\sum_{\ell=0}^{\tau-1}\|\nabla F(\boldsymbol{w}^t)\|^2 + \frac{\eta}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t (3-4s_i^t)\sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2]$$
$$+\frac{3\eta^3L^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\ell\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\kappa})\|^2] + \frac{3\eta^3L^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\ell\sum_{\kappa=0}^{\ell-1}\mathbb{E}[\|\boldsymbol{n}_i^{t,\kappa}\|^2] \tag{B10}$$
$$+\frac{\eta^2L\tau}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\nabla F_i(\boldsymbol{w}_i^{t,\ell})\|^2] + \frac{\eta^2L\tau}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t \sum_{\ell=0}^{\tau-1}\mathbb{E}[\|\boldsymbol{n}_i^{t,\ell}\|^2] + \frac{3\eta\tau}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \varepsilon_i.$$

To ensure the training performance, we will select a proper DP noise variance to have $\mathbb{E}[\|\boldsymbol{n}_i^{t,\ell}\|^2] = \mathbb{E}[\|\boldsymbol{n}_i^{t,\kappa}\|^2] \leqslant \Theta$. Due to the bounded gradient, by setting $\eta L\tau < 1$ and $\eta^3L^2 \ll 1$, we obtain

$$\mathbb{E}[F(\boldsymbol{w}^{t+1}) - F(\boldsymbol{w}^t)] \leqslant -\frac{\eta\tau}{2}\|\nabla F(\boldsymbol{w}^t)\|^2 + \frac{3\eta\tau G^2}{2}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t (1-s_i^t) + \frac{\eta^2L\tau^2\Theta(1+3\eta L\tau)}{2} + \frac{3\eta\tau}{2}\mathbb{E}\Big[\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t \varepsilon_i\Big]. \tag{B11}$$

Rearranging and summing t from 0 to $T-1$, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla F(\boldsymbol{w}^t)\|^2] \leqslant \frac{2(F(\boldsymbol{w}^0) - F(\boldsymbol{w}^T))}{\eta\tau T} + 3\varepsilon + \frac{3G^2}{T}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t (1-s_i^t) + \eta\tau^2\Theta(1+3\eta L\tau). \tag{B12}$$

This completes the proof. □

Appendix C Solution of the Optimal Gradient Sparsification Rate

To obtain the optimal gradient sparsification rate, we first derive the relation between s_i^t and V^t. Hence, we first consider the condition $Q^{t,\mathrm{de}} > 0$ and have

$$V^t = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}}(Q_i^{t,\mathrm{fa}} - \lambda p_i^t s_i^t)a_{i,j}^t + Q^{t,\mathrm{de}}\left(d^t - d^{\mathrm{Avg}}\right) - \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}}\beta_i = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}}(Q_i^{t,\mathrm{fa}} - \lambda p_i^t s_i^t)a_{i,j}^t$$
$$+ Q^{t,\mathrm{de}}\max_{i\in\mathcal{N}}\left\{\sum_{j\in\mathcal{N}} a_{i,j}^t\left(\frac{Zp_i^t s_i^t}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)} + d_i^{t,\mathrm{do}} + \frac{\tau|\mathcal{D}_i|\Phi_i}{f_i^t}\right)\right\} - Q^{t,\mathrm{de}}d^{\mathrm{Avg}} - \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{N}}\beta_i. \tag{C1}$$

Due to the maximizing process, this problem can be divided into N subproblems based on the client with the maximum delay. First, let us discuss the condition that the delay of the client owning the j-th channel is the maximum one among all clients. We assume the j-th channel is allocated to the i-th client and its delay is the maximum one. Thus, we can obtain

$$d_{i,j}^t = \frac{Zp_i^t s_i^t}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)} + d_i^{t,\mathrm{do}} + \frac{\tau|\mathcal{D}_i|\Phi_i}{f_i^t} \geqslant \max_{i'\in\mathcal{U}/i}\left\{\sum_{j'\in\mathcal{N}} a_{i',j'}^t\left(\frac{Zp_{i'}^t s_{i'}^t}{B\log_2\left(1+\frac{P_{i'}^t h_{i',j'}^t}{\sigma^2}\right)} + d_{i'}^{t,\mathrm{do}} + \frac{\tau|\mathcal{D}_{i'}|\Phi_{i'}}{f_{i'}^t}\right)\right\}. \tag{C2}$$

From the above inequation, we have

$$s_i^t \geqslant s_i^{t,\min} \triangleq B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)\frac{\max\limits_{i'\in\mathcal{U}/i}\left\{\sum_{j'\in\mathcal{N}} a_{i',j'}^t\left(\frac{Zp_{i'}^t s_{i'}^{\mathrm{th}}}{B\log_2\left(1+\frac{P_{i'}^t h_{i',j'}^t}{\sigma^2}\right)} + d_{i'}^{t,\mathrm{do}} + \frac{\tau|\mathcal{D}_{i'}|\Phi_{i'}}{f_{i'}^t}\right)\right\} - d_i^{t,\mathrm{do}} - \frac{\tau|\mathcal{D}_i|\Phi_i}{f_i^t}}{Zp_i^t}. \tag{C3}$$

If $s_i^{t,\min} > 1$, there is no solution to this subproblem. Otherwise, we can derive the first derivative of V^t with respect to s_i^t of the j-th subproblem as follows:

$$\frac{\partial V^t}{\partial s_i^t} = -\lambda + \frac{Zp_i^t Q^{t,\mathrm{de}}}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)}. \tag{C4}$$

If $\frac{Zp_i^t Q^{t,\mathrm{de}}}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)} \leqslant \lambda$, it can be found that as the value of s_i^t increases, the objective V^t decreases. Hence, we have $s_i^{t,*} = 1$. For other fast clients, i.e., $i' \in \mathcal{U}/i$, we have

$$s_{i'}^t \leqslant s_{i'}^{\max} \triangleq \frac{Zp_{i'}^t\left(\frac{Zp_i^t s_i^t}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)} + d_i^{t,\mathrm{do}} + \frac{\tau|\mathcal{D}_i|\Phi_i}{f_i^t} - d_{i'}^{t,\mathrm{do}} - \frac{\tau|\mathcal{D}_{i'}|\Phi_i}{f_{i'}^t}\right)}{\sum_{j'=1}^N a_{i',j'}^t B\log_2\left(1+\frac{P_{i'}^t h_{i',j'}^t}{\sigma^2}\right)}. \tag{C5}$$

We can also derive the first derivative of V^t with respect to $s_{i'}^t$ as $\frac{\partial V^t}{\partial s_{i'}^t} = -p_{i'}^t\lambda$. Therefore, we have $s_{i'}^{t,*} = \min\{s_{i'}^{\max}, 1\}$.

If $\frac{Zp_i^t Q^{t,\mathrm{de}}}{B\log_2\left(1+\frac{P_i^t h_{i,j}^t}{\sigma^2}\right)} > \lambda$, we can find that as the value of s_i^t increases, the objective V^t increases. Therefore, the system need to select a small gradient sparsification rate s_i^t in $[s^{\mathrm{th}}, 1]$ for the i-th client. However, for the i'-th client, i.e., $i' \in \mathcal{U}/i$, we want to select a large gradient sparsification rate in $[s^{\mathrm{th}}, \min\{1, s_{i'}^{\max}\}]$ because the first derivative of V^t with respect to $s_{i'}^t$ is negative. We can decrease the s_i^t from 1 and then $s_{i'}^{\max}$ may be selected. Therefore, the first derivative of V^t with respect to s_i^t should be modified because $s_{i'}^{\max}$ is related to s_i^t. When the first derivative of V^t with respect to s_i^t become negative, let us stop decreasing the value of s_i^t. We can note that this way can obtain the optimal $s_i^{t,*}$ and $s_{i'}^{t,*} = \min\{s_{i'}^{\max}, 1\}$.

Other subproblems can be addressed using the same method. Overall, after addressing all N subproblems, the optimal solution can be obtained as the final output. This completes the proof. □

Appendix D Feasibility Analysis

We first introduce the Lyapunov function $\Gamma(\boldsymbol{Q}^t) = \frac{1}{2}(Q^{t,\mathrm{de}})^2 + \frac{1}{2}\sum_{i\in\mathcal{U}}(Q_i^{t,\mathrm{fa}})^2$, in which the drift from one communication round can be given as

$$\begin{aligned}\Gamma\left(\boldsymbol{Q}^{t+1}\right) - \Gamma\left(\boldsymbol{Q}^t\right) &= \frac{1}{2}\left(\max\left\{Q^{t,\mathrm{de}} + d^{\mathrm{Avg}} - d^t, 0\right\}\right)^2 - \frac{1}{2}\left(Q^{t,\mathrm{de}}\right)^2 + \frac{1}{2}\sum_{i\in\mathcal{U}}\left(\max\left\{Q_i^{t-1,\mathrm{fa}} + \mathbb{1}_i^t - \beta_i, 0\right\}\right)^2 - \frac{1}{2}\sum_{i\in\mathcal{U}}(Q_i^{t,\mathrm{fa}})^2 \\ &\leqslant \frac{1}{2}\left(d^{\mathrm{Avg}} - d^t\right)^2 + Q^{t,\mathrm{de}}\left(d^{\mathrm{Avg}} - d^t\right) + \frac{1}{2}\sum_{i\in\mathcal{U}}\left(\mathbb{1}_i^t - \beta_i\right)^2 + \sum_{i\in\mathcal{U}} Q_i^{t,\mathrm{fa}}\left(\mathbb{1}_i^t - \beta_i\right).\end{aligned} \tag{D1}$$

Because

$$V^t(\boldsymbol{P}^t, \boldsymbol{s}^t, \boldsymbol{a}^t) = -\lambda\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t + \sum_{i\in\mathcal{U}} Q_i^{t,\mathrm{fa}}(\mathbb{1}_i^t - \beta_i) + Q^{t,\mathrm{de}}\left(d^t - d^{\mathrm{Avg}}\right), \tag{D2}$$

we have

$$V^t(\boldsymbol{P}^t, \boldsymbol{s}^t, \boldsymbol{a}^t) \leqslant -\lambda\sum_{i\in\mathcal{U}} p_i^t \mathbb{1}_i^t s_i^t + \frac{1}{2}\left(d^{\mathrm{Avg}} - d^t\right)^2 + Q^{t,\mathrm{de}}\left(d^{\mathrm{Avg}} - d^t\right) + \frac{1}{2}\sum_{i\in\mathcal{U}}\left(\mathbb{1}_i^t - \beta_i\right)^2 + \sum_{i\in\mathcal{U}} Q_i^{t,\mathrm{fa}}\left(\mathbb{1}_i^t - \beta_i\right). \tag{D3}$$

Due to $s_i^t, p_i^t \leqslant 1$, $\boldsymbol{a}_{i,j}^t \in \{0,1\}$, $\beta_i \leqslant \frac{N}{U}$, $d^t = \max_{i\in\mathcal{U}} \mathbb{1}_i^t d_{i,j}^t$ and $d_{i,j}^t = d_i^{t,\mathrm{do}} + d_i^{t,\mathrm{lo}} + d_{i,j}^{t,\mathrm{up}}$, $\forall i \in \mathcal{U}, j \in \mathcal{N}$, we have

$$\mathbb{E}\left[V^t(\boldsymbol{P}^t, \boldsymbol{s}^t, \boldsymbol{a}^t)|\boldsymbol{Q}^t\right] \leqslant C_1 + Q^{t,\mathrm{de}}\mathbb{E}\left[d^{\mathrm{Avg}} - d^t|\boldsymbol{Q}^t\right] + \sum_{i\in\mathcal{U}} Q_i^{t,\mathrm{fa}}\mathbb{E}\left[\mathbb{1}_i^t - \beta_i|\boldsymbol{Q}^t\right]. \tag{D4}$$

Based on Theorem 4.5 in [1] and Lemma 1 in [2], existing $\zeta^{\text{opt}} > 0$, we can obtain the following inequality:

$$\mathbb{E}\left[V^t(\boldsymbol{P}^t, \boldsymbol{s}^t, \boldsymbol{a}^t)\right] \leqslant C_1 + \zeta^{\text{opt}}. \tag{D5}$$

By summing this equation over $t = 0, 1, \ldots, T$, we obtain

$$\lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[V^t(\boldsymbol{P}^t, \boldsymbol{s}^t, \boldsymbol{a}^t)\right] \leqslant C_1 + \zeta^{\text{opt}} < \infty. \tag{D6}$$

This completes the proof. □

References

1 Neely M. J. Stochastic network optimization with application to communication and queueing systems. Synthesis Lectures on Communication Networks, 2010

2 Deng X., Li J., Shi L., et al. Wireless powered mobile edge computing: Dynamic resource allocation and throughput maximization. IEEE Trans. Mob. Comput., 2022, 21(6): 2271-2288