

Supplementary Materials for Delay-efficient Multi-hop Federated Learning with Opportunistic In-network Aggregation

Jiaming Du, Haichuan Ding, Xianhao Chen, Jinpeng Song, Chengwen Xing, and
Kaibin Huang, *Fellow, IEEE*

APPENDIX A

THE PROOF OF LEMMA 1

To start with, we first linearize the product of two binary variables, and we have Lemma 3.

Lemma 3 *θ_1 and θ_2 are two different binary variables. If $\theta_1\theta_2 = \sigma$, where σ is a continuous variable, we obtain the equivalent expression*

$$\begin{aligned}\sigma &\geq \theta_1 + \theta_2 - 1, \quad \sigma \geq 0, \\ \sigma &\leq \theta_1, \quad \sigma \leq \theta_2.\end{aligned}$$

Jiaming Du, Haichuan Ding, and Jinpeng Song are with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: dujiaming.ee@gmail.com; hcding@bit.edu.cn; jinpeng.song@outlook.com).

Chengwen Xing is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: xingchengwen@gmail.com).

Xianhao Chen and Kaibin Huang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: xchen@eee.hku.hk; huangkb@eee.hku.hk).

Then, we can rewrite (21) as follow:

$$\begin{aligned}
& \sum_{j \in \mathcal{A}} \theta_j(l) \left(\sum_{l' \in \mathcal{L}} \theta_j(l') - 1 \right) \frac{c_j D_0}{u_j} \\
&= \sum_{j \in \mathcal{A}} \left[\left(\theta_j(l) \sum_{l' \in \mathcal{L}} \theta_j(l') \right) \frac{c_j D_0}{u_j} \right] - \sum_{j \in \mathcal{A}} \theta_j(l) \frac{c_j D_0}{u_j}, \\
&= \sum_{j \in \mathcal{A}} \left[\left(\theta_j(l) \left(\theta_j(l) + \sum_{l' \in \mathcal{L} \setminus \{l\}} \theta_j(l') \right) \right) \frac{c_j D_0}{u_j} \right] - \sum_{j \in \mathcal{A}} \theta_j(l) \frac{c_j D_0}{u_j}, \\
&\stackrel{(b)}{=} \sum_{j \in \mathcal{A}} \left[\left(\theta_j(l) + \sum_{l' \in \mathcal{L} \setminus \{l\}} \theta_j(l) \theta_j(l') \right) \frac{c_j D_0}{u_j} \right] - \sum_{j \in \mathcal{A}} \theta_j(l) \frac{c_j D_0}{u_j}, \\
&= \sum_{j \in \mathcal{A}} \sum_{l' \in \mathcal{L} \setminus \{l\}} \theta_j(l) \theta_j(l') \frac{c_j D_0}{u_j}, \tag{51}
\end{aligned}$$

where (b) is obtained based on the observation that $\theta^2 = \theta$ when θ is a binary variable. Note that $\sum_{l' \in \mathcal{L} \setminus \{l\}} \theta_j(l) \theta_j(l')$ is the sum of $\theta_j(l) \theta_j(l')$, and each $\theta_j(l')$ is different from $\theta_j(l)$. Therefore, we can use Lemma 3 to obtain the $\theta_j(l) \theta_j(l') = \sigma_j(l')$ with constraints (22) and (23). Substituting these results into (51), Lemma 1 can be obtained.

APPENDIX B

THE PROOF OF LEMMA 2

It can be observed that $\phi \geq \theta T$ is equivalent to

$$\phi \geq \begin{cases} T & \text{if } \theta = 1, \\ 0 & \text{if } \theta = 0. \end{cases} \tag{52}$$

On the other hand, according to constraint (24), we have $\phi \geq 0$ and $\phi \geq -\Phi + T$ when $\theta = 0$, and $\phi \geq -\Phi$ and $\phi \geq T$ when $\theta = 1$. As Φ is a very large positive constant and T is a finite continuous variable, it is easy to confirm that $0 \geq -\Phi + T$ and $T \geq -\Phi$. Therefore, we have $\phi \geq 0$ when $\theta = 0$, and $\phi \geq T$ when $\theta = 1$, which is equivalent to (52), i.e., $\phi \geq \theta T$. The proof is completed.

APPENDIX C

THE LINEARIZATION OF (28) AND (29)

(28) and (29) can be rewritten in the logarithmic form

$$\log_2(t_1(l)) + \log_2(T^{\max}) \geq 2 \log_2(T^{T_1}(l)), \quad (53)$$

$$\log_2(t_2(l)) + \log_2(T^{\max}) \geq 2 \log_2\left(\sum_{j \in \mathcal{A}} \phi_j(l)\right), \quad (54)$$

where the $\log_2(t)$ can be replaced by the N_y -piecewise function $\mathcal{Z}_1(t)$ and $\mathcal{Z}_2(t)$, which is similar to $\mathcal{G}(\cdot)$ except the value range, satisfying the following constraints:

$$\mathcal{Z}_1(t) \leq Y_i^-(t), \quad \forall Y_i^- \in \mathcal{Y}, \quad (55)$$

$$\mathcal{Z}_2(t) \geq \min \{Y_i^+(t)\}, \quad \forall Y_i^+ \in \mathcal{Y}, \quad (56)$$

$$Y_i^-(t) = c_i^- t + d_i^-, \quad t \in \{t^-, t^+\}, \quad (57)$$

$$Y_i^+(t) = c_i^+ t + d_i^+, \quad t \in \{t^-, t^+\}, \quad (58)$$

where $\mathcal{Z}_1(t)$ is the lower bound of the approximation for $\log_2(t)$, and $\mathcal{Z}_2(t)$ is the upper bound of the approximation for $\log_2(t)$. Specifically, $c_i^- = \frac{N_y}{t^+ - t^-} \log_2 \left[\frac{it^+ + (N_y - i)t^-}{(i-1)t^+ + (N_y - i + 1)t^-} \right]$, $d_i^- = \log_2 \left[\frac{it^+ + (N_y - i)t^-}{N_y} \right] - \frac{it^+ + (N_y - i)t^-}{t^+ - t^-} \log_2 \left[\frac{it^+ + (N_y - i)t^-}{(i-1)t^+ + (N_y - i + 1)t^-} \right]$, $c_i^+ = \frac{2N_y}{[(2i-1)t^+ + (2N_y - 2i + 1)t^-] \ln 2}$, $d_i^+ = \log_2 \left[\frac{(2i-1)t^+ + (2N_y - 2i + 1)t^-}{2N_y} \right] - \frac{1}{\ln 2}$. For the minimum value t^- in (57) and (58), it is infeasible for $\log_2(t)$ to set $t^- = 0$. Thus, we set $t^- = \tilde{t}$, where \tilde{t} is a small constant close to zero. For the maximum value t^+ , t^+ is obtained by $t^+ = 2T^{\text{Direct}}$. T^{Direct} is the total delay obtained by solving the problem that only considers the no-aggregation case with relaxed constraints, which is detailed in Section V-D. Notice that the maximum delay experienced in the no-aggregation case should be higher than T^{\max} . Given the relaxed constraints, T^{Direct} is the lower bound of the delay experienced in the no-aggregation case. Thus, we set $t^+ = 2T^{\text{Direct}}$ to ensure that the value of T^{\max} is covered by the interval $\{t^-, t^+\}$. As (56) is not convex, we rewrite $\mathcal{Z}_2(t) \geq \min \{Y_i^+(t)\}$ as

$$\mathcal{Z}_2(t) \geq \sum_{i=1}^{N_y} \zeta_i \{Y_i^+(t)\}, \quad (59)$$

and we have

$$\sum_{i=1}^{N_y} \zeta_i = 1, \quad \zeta_i \in \{0, 1\}, \quad (60)$$

$$\zeta_i t_i^- \leq t \leq \zeta_i t_i^+ + (1 - \zeta_i) \tilde{T}, \quad (61)$$

where t_i^- and t_i^+ are the known piecewise points of $\mathcal{Z}_2(t)$, i.e., $t_i^- = t^- + (t^+ - t^-)(i - 1)/N_y$, $t_i^+ = t^- + i(t^+ - t^-)/N_y$, $i = 1, 2, \dots, N_y$, and \tilde{T} is a large constant. However, the constraints (59) has the product of variables ζ_i 's and $Y_i^+(t)$'s, which is still nonlinear. Thus, we reformulate them by using Lemma 2, and we obtain

$$\mathcal{Z}_2(t) \geq \sum_{i=1}^{N_y} \kappa_i^1(l), \quad (62)$$

where $\kappa_i(l)$'s satisfy the following constraints

$$\kappa_i^1(l) \geq -\zeta_i^1(l)\tilde{T}, \quad (63)$$

$$\kappa_i^1(l) \geq -(1 - \zeta_i^1(l))\tilde{T} + \{Y_i^+(t)\}. \quad (64)$$

As $\mathcal{Z}_1(t)$ and $\mathcal{Z}_2(t)$ are the lower bound and the upper bound of the approximation for $\log_2(t)$ respectively, we have $\mathcal{Z}_1(t_1(l)) + \mathcal{Z}_1(T_{max}) \leq \log_2(t_1(l)) + \log_2(T_{max})$ and $2\log_2(T^{T1}(l)) \leq 2\mathcal{Z}_2(T^{T1}(l))$. With the properly designed piecewise functions, the gap between the values of $\mathcal{Z}_1(t)$, $\mathcal{Z}_2(t)$ and the values of $\log_2(t)$ is small [1]. Therefore, by using the piecewise functions with constraints (55)-(64), (53) and (54) can be reformulated as

$$\mathcal{Z}_1(t_1(l)) + \mathcal{Z}_1(T_{max}) \geq 2\mathcal{Z}_2(T^{T1}(l)), \quad (65)$$

$$\mathcal{Z}_1(t_2(l)) + \mathcal{Z}_1(T_{max}) \geq 2\mathcal{Z}_2\left(\sum_{j \in \mathcal{A}} \phi_j(l)\right). \quad (66)$$

The proof is completed.

REFERENCES

- [1] X. Li, Y. Sun, Y. Guo, X. Fu, and M. Pan, "Dolphins first: Dolphin-aware communications in multi-hop underwater cognitive acoustic networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2043–2056, Apr. 2017.