Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Peixi LIU, Jiamo JIANG, Guangxu ZHU, Lei CHENG, Wei JIANG, Wu LUO, Ying DU, Zhiqin WANG, 2022. Training time minimization for federated edge learning with optimized gradient quantization and bandwidth allocation. *Front Inform Technol Electron Eng*, 23(8):1247-1263. https://doi.org/10.1631/FITEE.2100538

Proof S1 Proof of Theorem 1

We first give one necessary lemma before proving Theorem 1, and also provide the proof of this lemma after the proof of Theorem 1.

Lemma S1 Under the same conditions as in Theorem 1, we have

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq (1-\mu\eta_n)\,\mathbb{E}\left[\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2\right]+\alpha\Gamma\eta_n^2,$$

where $\Gamma = 2LF_\delta + \frac{1}{K}\sum_{k=1}^{K}\delta_k^2$.

Proof The proof is provided later in this section.

Next, with Lemma S1 and decaying learning rate $\eta_n = \frac{\beta}{\gamma+n}$, we prove that $\mathbb{E}\left[\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2\right] \leq \frac{\nu}{\gamma+n}$ by induction, where

$$\nu = \max\left\{(\gamma+1)\left\|\boldsymbol{w}_0-\boldsymbol{w}_*\right\|^2, \frac{\alpha\Gamma\beta^2}{\mu\beta-1}\right\}.$$

First, it holds for $n=1$ by the definition of ν. Then, assuming that it holds for some $n>1$, it follows from Lemma S1 that

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq \frac{(\gamma+n-\mu\beta)\nu}{(\gamma+n)^2}+\frac{\alpha\Gamma\beta^2}{(\gamma+n)^2} = \frac{(\gamma+n-1)\nu}{(\gamma+n)^2}+\frac{\alpha\Gamma\beta^2-(\mu\beta-1)\nu}{(\gamma+n)^2}.$$

By the definition of ν, we have $\alpha\Gamma\beta^2-(\mu\beta-1)\nu \leq 0$. Then, it follows that

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq \frac{(\gamma+n-1)\nu}{(\gamma+n)^2} \leq \frac{\nu}{\gamma+n+1}.$$

Specifically, we choose $\beta = \frac{2}{\mu}$ and $\gamma = \frac{2\alpha L}{\mu}-1$. Using $\max\{x,y\} \leq x+y$, we have $\nu \leq \frac{2\alpha L}{\mu}\left\|\boldsymbol{w}_0-\boldsymbol{w}_*\right\|^2 + \frac{4\alpha\Gamma}{\mu^2}$. Therefore, we have

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2\right] \leq \frac{\alpha/\mu}{n+2\alpha L/\mu-1}\left(2L\left\|\boldsymbol{w}_0-\boldsymbol{w}_*\right\|^2+\frac{4\Gamma}{\mu}\right).$$

Then, by the L-smoothness of $F(\boldsymbol{w})$, it holds that

$$\mathbb{E}\left[F\left(\boldsymbol{w}^{(n)}\right)\right]-F(\boldsymbol{w}_*) \leq \frac{L}{2}\mathbb{E}\left[\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2\right].$$

It follows that

$$\mathbb{E}\left[F\left(\boldsymbol{w}^{(n)}\right)\right]-F(\boldsymbol{w}_*) \leq \frac{\alpha L/\mu}{n+2\alpha L/\mu-1}\left(L\left\|\boldsymbol{w}_0-\boldsymbol{w}_*\right\|^2+\frac{2\Gamma}{\mu}\right).$$

We complete the proof of Theorem 1 by setting $n=N$.

Proof Proof of Lemma S1 is as follows:

Notice that $\boldsymbol{w}^{(n+1)} = \boldsymbol{w}^{(n)} - \dfrac{\eta_n}{K}\sum_{k=1}^{K}\mathcal{Q}\left(\boldsymbol{g}_k^{(n)}\right)$. Then we have

$$\left\|\boldsymbol{w}^{(n+1)} - \boldsymbol{w}_*\right\|^2 = \left\|\boldsymbol{w}^{(n)} - \frac{\eta_n}{K}\sum_{k=1}^{K}\mathcal{Q}\left(\boldsymbol{g}_k^{(n)}\right) - \boldsymbol{w}_*\right\|^2 = \|\boldsymbol{a}_1 - \boldsymbol{a}_2\|^2 = \|\boldsymbol{a}_1\|^2 + \|\boldsymbol{a}_2\|^2 - 2\left\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \right\rangle,$$

where $\boldsymbol{a}_1 = \boldsymbol{w}^{(n)} - \boldsymbol{w}_* - \frac{\eta_n}{K}\sum_{k=1}^{K}\boldsymbol{g}_k^{(n)}$ and $\boldsymbol{a}_2 = \frac{\eta_n}{K}\sum_{k=1}^{K}\left(\mathcal{Q}\left(\boldsymbol{g}_k^{(n)}\right) - \boldsymbol{g}_k^{(n)}\right)$. Due to $\mathbb{E}_{\mathcal{Q}}\left[\mathcal{Q}\left(\boldsymbol{g}_k^{(n)}\right)\right] = \boldsymbol{g}_k^{(n)}$, we have $\mathbb{E}_{\mathcal{Q}}[\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle] = 0$, which leads to

$$\left\|\boldsymbol{w}^{(n+1)} - \boldsymbol{w}_*\right\|^2 = \|\boldsymbol{a}_1\|^2 + \|\boldsymbol{a}_2\|^2. \tag{S1}$$

Next, we first obtain the upper bounds of A_1 and A_2; taking these bounds into Eq. (S1), then we find the connection between $\left\|\boldsymbol{w}^{(n+1)} - \boldsymbol{w}_*\right\|^2$ and $\left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_*\right\|^2$ after some proper manipulations.

1. Bound of $\|\boldsymbol{a}_1\|^2$: To bound $\|\boldsymbol{a}_1\|^2$, we break $\|\boldsymbol{a}_1\|^2$ as

$$\|\boldsymbol{a}_1\|^2 = \left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_* - \frac{\eta_n}{K}\sum_{k=1}^{K}\boldsymbol{g}_k^{(n)}\right\|^2 = \left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_*\right\|^2 + \underbrace{\left\|\frac{\eta_n}{K}\sum_{k=1}^{K}\boldsymbol{g}_k^{(n)}\right\|^2}_{B_1} + \underbrace{2\left\langle \boldsymbol{w}_* - \boldsymbol{w}^{(n)}, \frac{\eta_n}{K}\sum_{k=1}^{K}\boldsymbol{g}_k^{(n)}\right\rangle}_{B_2}.$$

To bound B_1, we use $\left\|\sum_{k=1}^{K}\boldsymbol{a}_k\right\|^2 \le K\sum_{k=1}^{K}\|\boldsymbol{a}_k\|^2$. This gives

$$B_1 \le \frac{\eta_n^2}{K}\sum_{k=1}^{K}\left\|\boldsymbol{g}_k^{(n)}\right\|^2.$$

By the μ-strong convexity of $F_k(\boldsymbol{w})$, it follows that

$$\left\langle \boldsymbol{w}_* - \boldsymbol{w}^{(n)}, \boldsymbol{g}_k^{(n)} \right\rangle \le F_k(\boldsymbol{w}_*) - F_k\left(\boldsymbol{w}^{(n)}\right) - \frac{\mu}{2}\left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_*\right\|^2.$$

Hence, B_2 can be bounded by

$$B_2 \le 2\frac{\eta_n}{K}\sum_{k=1}^{K}\left(F_k(\boldsymbol{w}_*) - F_k\left(\boldsymbol{w}^{(n)}\right)\right) - \mu\eta_n\left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_*\right\|^2.$$

2. Bound of $\|\boldsymbol{a}_2\|^2$: Since $\boldsymbol{g}_k^{(n)}$'s are independent and $\mathbb{E}_{\mathcal{Q}}\left[\left\|\mathcal{Q}\left(\boldsymbol{g}_k^{(n)}\right) - \boldsymbol{g}_k^{(n)}\right\|\right]^2 \le \dfrac{\sqrt{d}}{q}\left\|\boldsymbol{g}_k^{(n)}\right\|^2$ holds, it follows that

$$\mathbb{E}_{\mathcal{Q}}\left[\|\boldsymbol{a}_2\|^2\right] \le \frac{\sqrt{d}\eta_n^2}{qK^2}\sum_{k=1}^{K}\left\|\boldsymbol{g}_k^{(n)}\right\|^2.$$

With these bounds at hand, and taking expectation of Eq. (S1) over the stochastic quantizer $\mathcal{Q}$ and stochastic gradient at round n, we have

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)} - \boldsymbol{w}_*\right\|^2\right] \le (1 - \mu\eta_n)\left\|\boldsymbol{w}^{(n)} - \boldsymbol{w}_*\right\|^2 + 2\frac{\eta_n}{K}\sum_{k=1}^{K}\left(F_k(\boldsymbol{w}_*) - F_k\left(\boldsymbol{w}^{(n)}\right)\right) + \frac{\alpha\eta_n^2}{K}\sum_{k=1}^{K}\mathbb{E}\left\|\boldsymbol{g}_k^{(n)}\right\|^2. \tag{S2}$$

Recall that $\alpha = \frac{\sqrt{d}}{qK} + 1$. From Assumption 3, we have

$$\mathbb{E}\left\|\boldsymbol{g}_k^{(n)}\right\|^2 \le \delta_k^2 + \left\|\nabla F_k(\boldsymbol{w}^{(n)})\right\|^2. \tag{S3}$$

Substituting inequality (S3) into inequality (S2) yields

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq (1-\mu\eta_n)\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2 + 2\frac{\eta_n}{K}\sum_{k=1}^{K}\left(F_k(\boldsymbol{w}_*)-F_k\left(\boldsymbol{w}^{(n)}\right)\right) + \frac{\alpha\eta_n^2}{K}\sum_{k=1}^{K}\left(\delta_k^2+\left\|\nabla F_k\left(\boldsymbol{w}^{(n)}\right)\right\|^2\right).$$

The L-smoothness of $F_k(\boldsymbol{w})$ gives

$$\left\|\nabla F_k\left(\boldsymbol{w}^{(n)}\right)\right\|^2 \leq 2L\left(F_k\left(\boldsymbol{w}^{(n)}\right)-F_k^*\right).$$

It follows that

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq (1-\mu\eta_n)\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2 + \frac{\alpha\eta_n^2}{K}\sum_{k=1}^{K}\delta_k^2 + \underbrace{2\frac{\eta_n}{K}\sum_{k=1}^{K}\left(F_k\left(\boldsymbol{w}_*\right)-F_k\left(\boldsymbol{w}^{(n)}\right)\right)}_{C_1} + \underbrace{\frac{2L\eta_n^2\alpha}{K}\sum_{k=1}^{K}\left(F_k\left(\boldsymbol{w}^{(n)}\right)-F_k^*\right)}_{C_2}.$$

After rearranging C_1+C_2, we have

$$C_1+C_2 = 2\eta_n\left(\alpha L\eta_n-1\right)\left(F\left(\boldsymbol{w}^{(n)}\right)-F(\boldsymbol{w}_*)\right)+2\alpha L\eta_n^2F_\delta,$$

where $F_\delta \coloneqq F(\boldsymbol{w}_*)-\frac{1}{K}\sum_{k=1}^{K}F_k^*$.

It can be verified that $\eta_n \leq \frac{1}{\alpha L}$, and from $F(\boldsymbol{w}^{(n)}) \geq F(\boldsymbol{w}_*)$, we have

$$C_1 \leq 2\alpha L\eta_n^2F_\delta.$$

Taking the total expectation of Eq. (S1) yields

$$\mathbb{E}\left[\left\|\boldsymbol{w}^{(n+1)}-\boldsymbol{w}_*\right\|^2\right] \leq (1-\mu\eta_n)\,\mathbb{E}\left[\left\|\boldsymbol{w}^{(n)}-\boldsymbol{w}_*\right\|^2\right]+\alpha\eta_n^2\left(2LF_\delta+\frac{1}{K}\sum_{k=1}^{K}\delta_k^2\right),$$

which completes the proof.

Proof S2 Proof of Lemma 1

If $T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}<T_{\mathrm{d}}, \forall k\in[K]$, T_{d} can be reduced until $k\in[K]$ satisfies $T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}=T_{\mathrm{d}}$. Denote $\mathcal{K}=\{k\in[K]|T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}<T_{\mathrm{d}}\}$ and $\bar{\mathcal{K}}=\{k\in[K]|T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}=T_{\mathrm{d}}\}$. Obviously, $\mathcal{K}+\bar{\mathcal{K}}=[K]$. Since T_k^{comm} is an decreasing function of b_k, we can enforce $T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}=T_{\mathrm{d}}$ by decreasing b_k for all $k\in\mathcal{K}$. Then, $\mathcal{K}=\varnothing$ and $\bar{\mathcal{K}}=[K]$. In this case, if $\sum_{k=1}^{K}b_k<B_0$, we can properly increase each b_k, without violating $T_k^{\mathrm{comp}}+T_k^{\mathrm{comm}}=T_{\mathrm{d}}$, $k\in[K]$, until $\sum_{k=1}^{K}b_k=B_0$, and T_{d} will decrease as well.

Table S1 Simulation parameters

Parameter	Value
Number of edge devices (K)	6
Transmit power of edge devices (p_k)	1 dBm
CPU frequency (f_k)	100 MHz–1 GHz
Number of CPU cycles for one batch (ν)	10^8 (simulation 1); 2.5×10^{10} (simulation 2)
Variance of shadow fading (σ_η^2)	8 dB
Noise power spectral density (N_0)	-174 dBm/Hz
Total bandwidth (B_0)	10 kHz

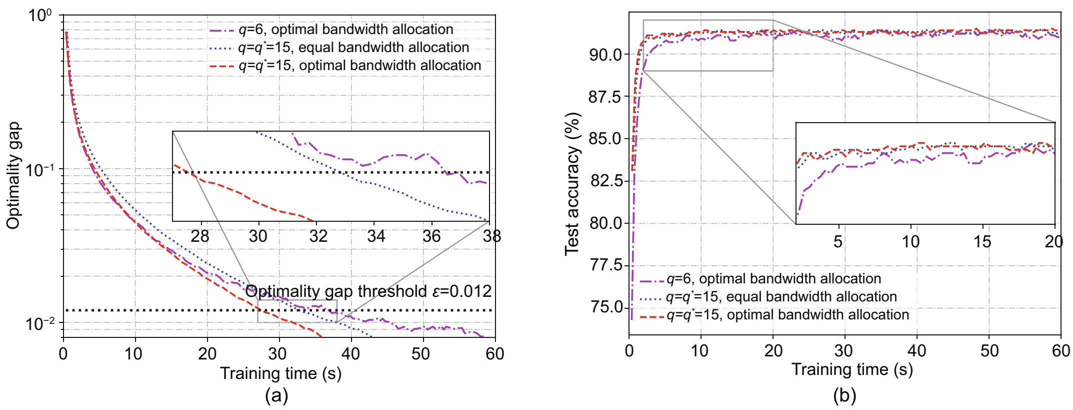

Fig. S1 Optimality gap (a) and test accuracy (b) in simulation 1

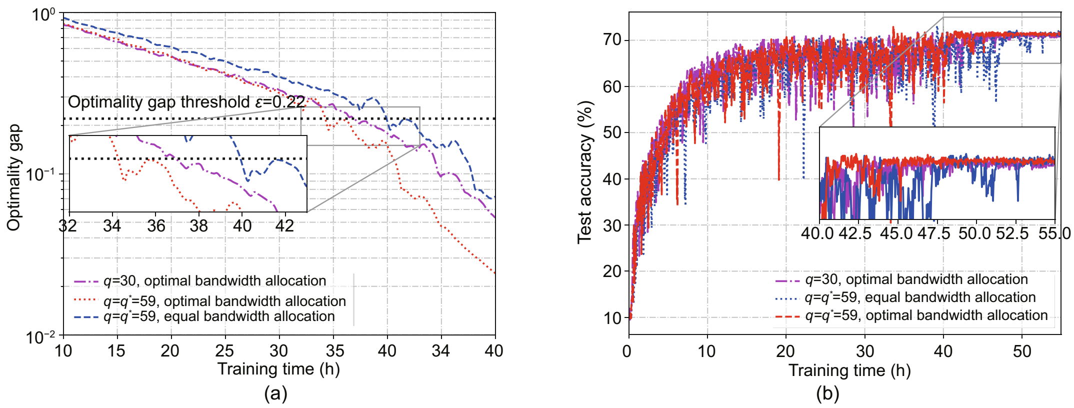

Fig. S2 Optimality gap (a) and test accuracy (b) in simulation 2