# Enable Accelerated and Accuracy-Guaranteed Federated Edge Learning Service via Model Splitting and Device Association

Shucun Fu[1]

School of Computer Science and Engineering, Southeast University, Nanjing, China
shucunfu@seu.edu.cn

## 1 Appendix

### 1.1 Proof of the Theorem 1

Case1: $\frac{\alpha_m(1-\beta)|D_m|F_w}{c_m} \geq \frac{(1-\alpha_m)|D_m|F_w}{C_s}$, there is

$$\alpha_m(1-\beta)|D_m|F_w \cdot C_s \geq (1-\alpha_m)|D_m|F_w \cdot c_m$$

Hence, we have

$$\frac{\alpha_m}{(1-\alpha_m)} \geq \frac{c_m}{(1-\beta)C_s} \Rightarrow \alpha_m \geq \frac{c_m}{c_m+(1-\beta)C_s}$$

Formula (6) can be converted to

$$\begin{aligned} T_m^{SL}(\alpha_m, \mathbf{a}) &= \tfrac{\alpha_m\beta|D_m|F_w}{c_m} + \tfrac{|D_m|Ir(\cdot)}{R_m} + \tfrac{\alpha_m(1-\beta)|D_m|F_w}{c_m} \\ &= \alpha_m \cdot \tfrac{|D_m|F_w}{c_m} + \tfrac{|D_m|Ir(\cdot)}{R_m} \end{aligned}$$

It's easy to observe that Formula (6) is an increasing function of $\alpha_m$. Hence, we have

$$\alpha_m^*{=}\frac{c_m}{c_m+(1-\beta)C_s},\ pr_{\alpha_m} \leq \overline{pr}_m$$

Case2: $\frac{\alpha_m(1-\beta)|D_m|F_w}{c_m} < \frac{(1-\alpha_m)|D_m|F_w}{C_s}$, there is

$$\alpha_m < \frac{c_m}{c_m+(1-\beta)C_s}$$

Formula (6) can be converted to

$$\begin{aligned} T_m^{SL}(\alpha_m, \mathbf{a}) &= \tfrac{\alpha_m\beta|D_m|F_w}{c_m} + \tfrac{|D_m|Ir(\cdot)}{R_m} + \tfrac{(1-\alpha_m)|D_m|F_w}{C_s} \\ &= \alpha_m(\tfrac{\beta|D_m|F_w}{c_m} - \tfrac{|D_m|F_w}{C_s}) + \tfrac{|D_m|Ir(\cdot)}{R_m} + \tfrac{|D_m|F_w}{C_s} \end{aligned}$$

*(i)* if $\frac{\beta|D_m|F_w}{c_m} > \frac{|D_m|F_w}{C_s}$, we have $C_s > \frac{c_m}{\beta}$. It's easy to observe that (6) is an increasing function of $\alpha_m$.

*(ii)* if $\frac{\beta|D_m|F_w}{c_m} \leq \frac{|D_m|F_w}{C_s}$, we have $C_s \leq \frac{c_m}{\beta}$. Formula (6) is an decreasing function of $\alpha_m$. Hence, we have

$$\alpha_m^*=\frac{c_m}{c_m+(1-\beta)C_s},\ pr_{\alpha_m} \leq \overline{pr}_m$$

Combined with the above analysis, we can get

$$\alpha_m^* = \begin{cases} \frac{c_m}{c_m+(1-\beta)C_s}, \alpha_m \geq \frac{c_m}{c_m+(1-\beta)C_s} \\ \frac{c_m}{c_m+(1-\beta)C_s}, \alpha_m < \frac{c_m}{c_m+(1-\beta)C_s} \cap c_m \geq \beta C_s \\ \arg\min \alpha_m, \alpha_m < \frac{c_m}{c_m+(1-\beta)C_s} \cap c_m < \beta C_s \end{cases}$$

This proves the theorem 1.

### 1.2 Proof of the Lemma 2

The training time of split learning should not exceed the time of vanilla FL, i.e.,

$$T_m^{FP}(\alpha_m,\mathbf{a})+T_m^{com}(\alpha_m,\mathbf{a})+\max\{T_m^{BP}(\alpha_m,\mathbf{a}),T_m^{cmp\_s}(\alpha_m,\mathbf{a})\} \leq T_m^{cmp\_e}(L,0,\mathbf{a}_{-m})$$

Let

$$\Delta(\alpha_m,\mathbf{a}) \triangleq T_m^{cmp\_e}(L,0,\mathbf{a}_{-m})-(T_m^{FP}(\alpha_m,\mathbf{a})+\max\{T_m^{BP}(\alpha_m,\mathbf{a}),T_m^{cmp\_s}(\alpha_m,\mathbf{a})\})$$

We have $R_m \geq (|D_m|q(\alpha_m))/\Delta(\alpha_m,\mathbf{a})$. Then, according to Formula (1), we have

$$\frac{\rho_m\sigma_{m,s}}{\omega_0+\sum_{i\in\mathcal{M}\setminus\{m\}:a_i=a_m}\rho_i\sigma_{i,s}} \geq 2^{\frac{Ir(\cdot)|D_m|}{\Delta(\alpha_m,\mathrm{a})W^{UL}}}-1$$

Thus,

$$\sum_{i\in\mathcal{M}\setminus\{m\}:a_i=a_m}\rho_i\cdot\sigma_{i,s} \leq \frac{\rho_m\cdot\sigma_{m,s}}{2^{\frac{Ir(\cdot)|D_m|}{\Delta(\alpha_m,\mathrm{a})W^{UL}}}-1}-\omega_0 \triangleq \delta_m$$

Lemma 2 holds.

### 1.3 Proof of the Theorem 3

It is assumed that changing the association decision of device $i$ from $a_i$ to ${a'}_i$ can reduce the training time, i.e., $C_i(s_i,a_i) > C_i(s_i,{a'}_i)$. We will consider the following two cases: 1) $a_i=0$ and ${a'}_i>0$; 2) $a_i>0$ and ${a'}_i=0$.

For case 1): Given $a_i=0$ and ${a'}_i>0$, there is

$$\Phi(a_i,\mathbf{a}_{-i})-\Phi({a'}_i,\mathbf{a}_{-i}) = \rho_i\sigma_{i,s}\delta_i-\rho_i\sigma_{i,s}\sum_{j\in\mathcal{M}\setminus\{i\}}\rho_j\sigma_{j,s}\cdot I_{\{{a'}_i=a_j\}}$$

If device $i$ chooses to change policy to ${a'}_i$, then Formula (10) implies that

$$\sum_{j\in\mathcal{M}\setminus\{i\}} \rho_j\sigma_{j,s}\cdot I_{\{a_i=a_j\}} \leq \delta_i$$

Hence,

$$\begin{aligned}\Phi(a_i,\mathbf{a}_{-i})-\Phi({a'}_i,\mathbf{a}_{-i}) =&\rho_i\sigma_{i,s}(\delta_i-\textstyle\sum_{j\in\mathcal{M}\setminus\{i\}}\rho_j\sigma_{j,s}\cdot I_{\{{a'}_i=a_j\}})>\\ &\rho_i\sigma_{i,s}(\textstyle\sum_{j\in\mathcal{M}\setminus\{i\}}\rho_j\sigma_{j,s}\cdot I_{\{a_i=a_j\}}-\rho_j\sigma_{j,s}\cdot I_{\{{a'}_i=a_j\}})\end{aligned}$$

Because

$$\sum_{j\in\mathcal{M}\setminus\{i\}} \rho_j\sigma_{j,s}\cdot I_{\{a_i=a_j\}} > \sum_{j\in\mathcal{M}\setminus\{i\}} \rho_j\sigma_{j,s}\cdot I_{\{{a'}_i=a_j\}}$$

Hence

$$\Phi(a_i,\mathbf{a}_{-i})-\Phi({a'}_i,\mathbf{a}_{-i})>0$$

For case 2): Given $a_i>0$ and ${a'}_i=0$, Similar to the proof in case 1), there is $\Phi(a_i,\mathbf{a}_{-i})-\Phi({a'}_i,\mathbf{a}_{-i})>0$

Therefore, the original device association issue can be modeled as a potential game to describe the competition of devices. This proves the theorem 3.

### 1.4 Proof of the Theorem 4

According to the definition of the potential function in (11)

$$0\leq\Phi(\mathrm{a})\leq\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\gamma_{\max}^2+\sum_{i=1}^{M}\gamma_{\max}\delta_{\max}=\frac{1}{2}\gamma_{\max}^2M^2+\gamma_{\max}\delta_{\max}M$$

It is assumed that changing the association decision from $a_i$ to ${a'}_i$ can reduce the training time, i.e., $C_i(a_i,\mathbf{a}_{-i})>C_i({a'}_i,\mathbf{a}_{-i})$.

According to the nature of the potential game, the potential increases by at least $\gamma_{\min}$

$$\Phi(a_i,\mathbf{a}_{-i})\geq\Phi({a'}_i,\mathbf{a}_{-i})+\gamma_{\min}$$

We will consider the following three cases: 1) $a_i=0$ and ${a'}_i>0$; 2) $a_i>0$ and ${a'}_i=0$.

For case 1): Given $a_i=0$ and ${a'}_i>0$, there is

$$\Phi(a_i,\mathbf{a}_{-i})-\Phi({a'}_i,\mathbf{a}_{-i})=\gamma_i(\delta_i-\sum_{k\neq i}\gamma_k I_{\{a_k=1\}})>0$$

Since $\gamma_i>0$ is an integer for any $i\in\mathcal{M}$, there is

$$\delta_i\geq\sum_{k\neq i}\gamma_k I_{\{a_k=1\}}+1$$

Hence, we have

$$\Phi(a_i,\mathbf{a}_{-i})\geq\Phi({a'}_i,\mathbf{a}_{-i})+\gamma_i\geq\Phi({a'}_i,\mathbf{a}_{-i})+\gamma_{\min}$$

For case 2): Given $a_i > 0$ and $a'_i = 0$, Similar to the proof in case 1), there is

$$\Phi(a_i, \mathbf{a}_{-i}) \geq \Phi(a'_i, \mathbf{a}_{-i}) + \gamma_{\min}$$

Combined with the above analysis, we can get:

$$R \leq (\frac{1}{2}\gamma_{\max}^2 M^2 + \gamma_{\max}\delta_{\max} M)/\gamma_{\min}$$

Therefore, Theorem 4 holds.

### 1.5 Proof of the Theorem 5

Case 1): the system cost corresponding to the optimal solution a$^*$ must be the lowest, Thus we have

$$PoA_{EDA-G} \geq 1$$

Case 2): for any $m \in \mathcal{M}$ and $\widehat{\mathbf{a}} \in \Omega$, there is $C_m(\widehat{\mathbf{a}}) \leq C_m^{FL}(\mathbf{a})$, we can get:

$$\sum_{m \in M} C_m(\widehat{\mathrm{a}}) \leq \sum_{m \in M} C_m^{FL}(\mathrm{a})$$

For any $a_m$, if $a_m = 0$, we have $C_m(\mathrm{a}) = C_m^{FL}(\mathrm{a})$. Otherwise,

$$R_m(\mathrm{a}) = W^{UL}\log_2(1 + \frac{\rho_m \cdot \sigma_{m,s}}{\omega_0 + \sum_{i \in \mathcal{M}\setminus\{m\}: a_i \neq 0} \rho_i \cdot \sigma_{i,s}}) \leq W^{UL}\log_2(1 + \frac{\rho_m \cdot \sigma_{m,s}}{\omega_0})$$

According to the above analysis

$$\begin{gathered} C_m^{SL}(\mathrm{a}) = T_m^{FP}(\mathrm{a}) + T_m^{com}(\mathrm{a}) + \max\{T_m^{BP}(\mathrm{a}), T_m^{cmp\_s}(\mathrm{a})\} = \\ T_m^{FP}(\mathrm{a}) + \frac{Ir(\cdot)|D_m|}{R_m} + \max\{T_m^{BP}(\mathrm{a}), T_m^{cmp\_s}(\mathrm{a})\} \geq \\ T_m^{FP}(\mathrm{a}) + \frac{Ir(\cdot)|D_m|}{W^{UL}\log_2(1+\frac{\rho_m \cdot \sigma_{m,s}}{\omega_0})} + \max\{T_m^{BP}(\mathrm{a}), T_m^{cmp\_s}(\mathrm{a})\} \triangleq C_m^{\min}(\mathrm{a}) \end{gathered}$$

Further, it can be obtained that

$$C_m(\mathrm{a}) \geq \min\{C_m^{FL}(\mathrm{a}), C_m^{\min}(\mathrm{a})\}$$

and

$$\sum_{m \in \mathcal{M}} C_m(\mathrm{a}) \geq \sum_{m \in \mathcal{M}} \min\{C_m^{FL}(\mathrm{a}), C_m^{\min}(\mathrm{a})\}$$

Thus we have

$$1 \leq PoA_{EDA-G} = \frac{\max\limits_{\widehat{\mathrm{a}} \in \Omega} \sum\limits_{m \in \mathcal{M}} C_m(\widehat{\mathrm{a}})}{\sum\limits_{m \in \mathcal{M}} C_m(\mathrm{a}^*)} \leq \frac{\sum\limits_{m \in \mathcal{M}} C_m^{FL}(\mathrm{a})}{\sum\limits_{m \in \mathcal{M}} \min\{C_m^{FL}(\mathrm{a}), C_m^{\min}(\mathrm{a})\}}$$

This proves the theorem 5.