Convergence Analysis of SFL-ISCC

Xiangwang Hou, Jingjing Wang, Jiacheng Wang, Jun Du, Zekai Zhang, Chunxiao Jiang, Yong Ren

Before establishing the convergence rate, some key Lemmas are presented as follows.

Lemma 1. *According to **Assumption 1**, we can derive*

$$\mathbb{E}\left[\left\|\boldsymbol{w}_c^t-\boldsymbol{w}_{c,m}^t\right\|^2\right] \leq 4\eta^2 I^2 \sum_{l=1}^{L_c} G_l^2. \tag{1}$$

Proof. See APPENDIX A. □

Lemma 2. *Under **Assumption 1** and **Lemma 1**, we have*

$$\begin{aligned}&\mathbb{E}\left[\left\langle\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right),\boldsymbol{w}^t-\boldsymbol{w}^{t-1}\right\rangle\right] \leq \sum_{m=1}^{M}\eta\Lambda_m^2\\&+4M\beta^2\eta^3 I^2\sum_{l=1}^{L_c}G_l^2-\frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right\|^2\right].\end{aligned} \tag{2}$$

Proof. See APPENDIX B. □

Lemma 3. *According to **Assumption 3** and **Assumption 4**, it holds that*

$$\begin{aligned}&\mathbb{E}\left[\|\boldsymbol{w}^t-\boldsymbol{w}^{t-1}\|^2\right] < \frac{2\eta^2\sum\limits_{l=1}^{L}\sigma_l^2}{M}+\frac{4\eta^2}{M^2}\sum_{m=1}^{M}\Lambda_m^2\\&+4\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2\\&+32\beta^2\eta^4 I^2\sum_{l=1}^{L_c}G_l^2+8\eta^2\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right].\end{aligned} \tag{3}$$

Proof. See APPENDIX C. □

Theorem 1. *We consider the learning rate η of proposed SFLSCC satisfies that*

$$0<\eta<\frac{\left(\sum\limits_{m=1}^{M}\varphi_m\right)^2}{8\beta}. \tag{4}$$

*Under **Assumption** 1-4 and **Lemma** 1-3, we can obtain*

$$\begin{aligned}&\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right] \leq \frac{2}{\eta\left(\sum\limits_{m=1}^{M}\varphi_m\right)^2-8\beta\eta^2}\cdot\\&\left(\frac{\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^0\right)\right]-\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^*\right)\right]}{N}+\frac{\beta\eta^2\sum\limits_{l=1}^{L}\sigma_l^2}{M}\right.\\&+\beta^2\eta^3 I^2\left(16\beta\eta+4M\right)\sum_{l=1}^{L_c}G_l^2+\left(\frac{2\beta\eta^2}{M^2}+\eta\right)\sum_{m=1}^{M}\Lambda_m^2\\&\left.+2\beta\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2\right).\end{aligned} \tag{5}$$

Proof. See APPENDIX D. □

Eq. (28) *indicates that when the UAVs have varying target sensing probabilities, the adverse effects resulting from data heterogeneity $2\beta\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-q_s)^2+q_s^2\right)\Lambda_m^2$ will be amplified. For alleviating this adverse effects, we assume the UAVs have the uniform target sensing probabilities, thus $q_{s,m}=q_s$.*

Then, under the constraint of learning rate η as shown in Eq. (4)*, we can obtain the expression of φ_m and κ_m, which can be given by*

$$\varphi_m=\frac{q_s}{\left[1-(1-q_s)^M\right]M}, \tag{6}$$

and

$$\kappa_m=\frac{2}{Mq_s^M}. \tag{7}$$

The proof of φ_m can be seen at APPENDIX E and the proof of κ_m can be seen at APPENDIX B of [1].

APPENDIX A

Let's fix the training round at $t\geq 1$. Identify the largest $t_0\leq t$ and t_0 is a multiple of I (i.e. t_0 mod $I=0$). It should be noted that such a t_0 definitely exists and the difference $t-t_0$ is at most I. Recalling $\boldsymbol{w}_{c,m}^{t+1}=\boldsymbol{w}_{c,m}^t-\eta\alpha_m^t\mathbf{g}\left(\boldsymbol{w}_{c,m}^t\right)$ and $\boldsymbol{w}_c^{t+1}=\frac{1}{M}\sum\limits_{m=1}^{M}\boldsymbol{w}_{c,m}^{t+1}$ for client-side model updating and aggregation, we have

$$\boldsymbol{w}_{c,m}^t=\boldsymbol{w}_c^{t_0}-\eta\sum_{\tau=t_0}^{t-1}\alpha_m^\tau\mathbf{g}_{c,m}^\tau \tag{8}$$

and

$$\boldsymbol{w}_c^t=\boldsymbol{w}_c^{t_0}-\eta\sum_{\tau=t_0}^{t-1}\frac{1}{M}\sum_{m=1}^{M}\alpha_m^\tau\mathbf{g}_{c,m}^\tau. \tag{9}$$

Xiangwang Hou, Jun Du and Yong Ren are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (E-mail: xiangwanghou@163.com; jundu@tsinghua.edu.cn; reny@tsinghua.edu.cn).

Jingjing Wang is with the School of Cyber Science and Technology, Beihang University, Beijing 100191, China. (Email: drwangjj@buaa.edu.cn).

Jiacheng Wang is with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798. (E-mail: jcwang_cq@foxmail.com).

Zekai Zhang is with Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China. (E-mail: zhangzekai_2000@163.com).

Chunxiao Jiang is with the Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, 100084, China. (E-mail: jchx@tsinghua.edu.cn).

Thus, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\boldsymbol{w}_c^t-\boldsymbol{w}_{c,m}^t\right\|^2\right]\\
&=\eta^2\mathbb{E}\left[\left\|\sum_{\tau=t_0}^{t-1}\frac{1}{M}\sum_{m=1}^{M}\alpha_m^\tau\mathbf{g}_{c,m}^\tau-\sum_{\tau=t_0}^{t-1}\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2\right]\\
&\overset{(a)}{\leq} 2\eta^2\mathbb{E}\left[\left\|\sum_{\tau=t_0}^{t-1}\frac{1}{M}\sum_{m=1}^{M}\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2+\left\|\sum_{\tau=t_0}^{t-1}\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2\right]\\
&\overset{(b)}{\leq} 2\eta^2\left(t-t_0\right)\mathbb{E}\left[\sum_{\tau=t_0}^{t-1}\left\|\frac{\sum_{m=1}^{M}\alpha_m^\tau\mathbf{g}_{c,m}^\tau}{M}\right\|^2+\sum_{\tau=t_0}^{t-1}\left\|\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2\right]\\
&\overset{(c)}{\leq} 2\eta^2\left(t-t_0\right)\mathbb{E}\left[\frac{\sum_{\tau=t_0}^{t-1}\sum_{m=1}^{M}\left\|\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2}{M}+\sum_{\tau=t_0}^{t-1}\left\|\alpha_m^\tau\mathbf{g}_{c,m}^\tau\right\|^2\right]\\
&\overset{(d)}{\leq} 4\eta^2\left(t-t_0\right)^2\sum_{l=1}^{L_c}G_l^2\\
&\leq 4\eta^2 I^2\sum_{l=1}^{L_c}G_l^2,
\end{aligned}
\tag{10}
$$

where inequality $(a)-(c)$ follows from $\|\sum_{m=1}^{n}\boldsymbol{x}_m\|^2 \leq n\sum_{m=1}^{n}\|\boldsymbol{x}_m\|^2$, and inequality (d) is due to **Assumption 2**.

APPENDIX B

To begin with, we derive the expectation expression of $\boldsymbol{w}^t-\boldsymbol{w}^{t-1}$.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{m=1}^{M}\alpha_m^{t-1}\mathbf{g}_m^{t-1}}{M}\middle|\sum_{m=1}^{M}\alpha_m^{t-1}\neq 0\right]\\
&=\mathbb{E}\Bigg[\sum_{l=1}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}Pr\Bigg(\alpha_{m_1}^{t-1}=1\forall m_1\in\mathcal{M}_1,\alpha_{m_2}^{t-1}=0\\
&\qquad\forall m_2\in\mathcal{M}_2\middle|\sum_{m=1}^{M}\alpha_m^{t-1}\neq 0\Bigg)\frac{1}{M}\sum_{m_1\in\mathcal{M}_1}\mathbf{g}_{m_1}^{t-1}\Bigg]\\
&=\mathbb{E}\Bigg[\sum_{l=1}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\frac{1}{M}\sum_{m_1\in\mathcal{M}_1}\mathbf{g}_{m_1}^{t-1}\\
&\qquad\cdot\frac{\prod_{m_1\in\mathcal{M}_1}q_{s,m_1}\prod_{m_2\in\mathcal{M}_2}\left(1-q_{s,m_2}\right)}{1-\prod_{m\in\mathcal{M}}\left(1-q_{s,m}\right)}\Bigg]\\
&\triangleq\sum_{m=1}^{M}\varphi_m\mathbb{E}\left[\mathbf{g}_m^{t-1}\right],
\end{aligned}
\tag{11}
$$

where $\mathcal{M}_1$ represents the set of UAVs which succeed in sensing the target, while $\mathcal{M}_2$ denotes the set of UAVs which fail to do that. Besides, $\varphi_m,\forall m\in\mathcal{M}$ is related to $q_{s,m}$ and $\sum_{m=1}^{M}\varphi_m\leq 1$, this can be observed by setting $\mathbf{g}_m^{t-1}=1$.

Next, we prove **Lemma 2** as follows

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right),\boldsymbol{w}^t-\boldsymbol{w}^{t-1}\right\rangle\right]\\
&=-\eta\mathbb{E}\left[\left\langle\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right),\frac{\sum_{m=1}^{M}\alpha_m^{t-1}\mathbf{g}_m^{t-1}}{M}\right\rangle\right]\\
&\overset{(a)}{=}-\eta\mathbb{E}\left[\left\langle\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right),\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right\rangle\right]\\
&\leq-\eta\mathbb{E}\left[\left\langle\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right),\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right\rangle\right]\\
&\overset{(b)}{=}\frac{\eta}{2}\mathbb{E}\Bigg[\left\|\sum_{m=1}^{M}\varphi_m\left(\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right)\right\|^2\\
&\qquad-\left\|\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right\|^2-\left\|\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right\|^2\Bigg]\\
&\leq\underbrace{\frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{m=1}^{M}\varphi_m\left(\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right)\right\|^2\right]}_{\triangleq X_B}\\
&\qquad-\frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right\|^2\right],
\end{aligned}
\tag{12}
$$

where equality (a) is based on Eq.(11) and **Assumption** 4, equality (b) follows from $\langle\boldsymbol{x},\boldsymbol{y}\rangle=\frac{1}{2}(\|\boldsymbol{x}\|^2+\|\boldsymbol{y}\|^2-\|\boldsymbol{x}-\boldsymbol{y}\|^2)$. Then, we derive the upper bound of X_B.

$$
\begin{aligned}
X_B&=\frac{\eta}{2}\mathbb{E}\Bigg[\Bigg\|\sum_{m=1}^{M}\varphi_m\left(\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)\right.\\
&\qquad\left.+\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right)\Bigg\|^2\Bigg]\\
&\overset{(b)}{\leq}\eta\mathbb{E}\Bigg[\left\|\sum_{m=1}^{M}\varphi_m\left(\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)\right)\right\|^2\\
&\qquad+\left\|\sum_{m=1}^{M}\varphi_m\left(\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right)\right\|^2\Bigg]\\
&\overset{(c)}{\leq}\eta\sum_{m=1}^{M}\varphi_m^2\mathbb{E}\Bigg[\sum_{m=1}^{M}\left\|\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)\right\|^2\\
&\qquad+\sum_{m=1}^{M}\left\|\nabla\mathcal{L}_m\left(\boldsymbol{w}^{t-1}\right)-\nabla\mathcal{L}_m\left(\boldsymbol{w}_m^{t-1}\right)\right\|^2\Bigg]
\end{aligned}
\tag{13}
$$

$$\overset{(d)}{\leq} \eta \left(\sum_{m=1}^{M} \Lambda_m^2 + \beta^2 \sum_{m=1}^{M} \mathbb{E}\left[\left\| \boldsymbol{w}^{t-1} - \boldsymbol{w}_m^{t-1} \right\|^2 \right] \right)$$
$$\leq \eta \sum_{m=1}^{M} \Lambda_m^2 + \eta\beta^2 \sum_{m=1}^{M} \mathbb{E}\left[\|\boldsymbol{w}_s^{t-1} - \boldsymbol{w}_{s,m}^{t-1}\|^2 + \|\boldsymbol{w}_c^{t-1} - \boldsymbol{w}_{c,m}^{t-1}\|^2 \right]$$
$$\leq \sum_{m=1}^{M} \eta\Lambda_m^2 + 4M\beta^2\eta^3 I^2 \sum_{l=1}^{L_c} G_l^2,$$

where inequality (b) is due to $\|\boldsymbol{x}+\boldsymbol{y}\|^2 \leq 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$ and inequality (c) stems from $\|\sum_{m=1}^{M} \boldsymbol{x}_m \boldsymbol{y}_m\|^2 \leq \sum_{m=1}^{M} \|\boldsymbol{x}_m\|^2 \sum_{m=1}^{M} \|\boldsymbol{y}_m\|^2$, inequality (d) results from **Assumption** 1 and 3.

Thus, we can derive that

$$\mathbb{E}\left[\langle \nabla \mathcal{L}\left(\boldsymbol{w}^{t-1}\right), \boldsymbol{w}^t - \boldsymbol{w}^{t-1} \rangle \right] \leq \sum_{m=1}^{M} \eta\Lambda_m^2 + 4M\beta^2\eta^3 I^2 \sum_{l=1}^{L_c} G_l^2 - \frac{\eta}{2}\mathbb{E}\left[\left\| \sum_{m=1}^{M} \varphi_m \nabla \mathcal{L}\left(\boldsymbol{w}^{t-1}\right) \right\|^2 \right]. \tag{14}$$

APPENDIX C

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{w}^t - \boldsymbol{w}^{t-1}\|^2\right] &= \eta^2 \mathbb{E}\left[\left\| \frac{\sum_{m=1}^{M} \alpha_m^{t-1} \mathbf{g}_m^{t-1}}{M} \right\|^2 \right] \\
&= \eta^2 \mathbb{E}\left[\left\| \frac{\sum_{M=1}^{m} \alpha_m^{t-1}\left(\mathbf{g}_m^{t-1} - \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})\right)}{M} + \frac{\sum_{m=1}^{M} \alpha_m^{t-1} \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})}{M} \right\|^2 \right] \\
&\leq 2\eta^2 \underbrace{\mathbb{E}\left[\left\| \frac{\sum_{m=1}^{M} \alpha_m^{t-1}\left(\mathbf{g}_m^{t-1} - \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})\right)}{M} \right\|^2 \right]}_{\triangleq X} \\
&+ 2\eta^2 \underbrace{\mathbb{E}\left[\left\| \frac{\sum_{m=1}^{M} \alpha_m^{t-1} \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})}{M} \right\|^2 \right]}_{\triangleq Y},
\end{aligned} \tag{15}$$

Then, we derive the upper bound of X and Y, respectively.

1) Bound of X:

$$\begin{aligned}
X &\overset{(a)}{=} \mathbb{E}\left[\frac{\sum_{m=1}^{M} \alpha_m^{t-1} \left\| \left(\mathbf{g}_m^{t-1} - \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})\right) \right\|^2}{M^2} \right] \\
&+ \mathbb{E}\left[\frac{\sum_{m_1=1}^{M} \alpha_{m_1}^{t-1} \sum_{m_2=1, m_2 \neq m_1}^{M} \alpha_{m_2}^{t-1} \triangle_{m_1} \triangle_{m_2}}{M^2} \right] \\
&\leq \mathbb{E}\left[\frac{\sum_{m=1}^{M} \left\| \left(\mathbf{g}_m^{t-1} - \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})\right) \right\|^2 + \sum_{m_1=1}^{M} \sum_{\substack{m_2=1 \\ m_2 \neq m_1}}^{M} \triangle_{m_1} \triangle_{m_2}}{M^2} \right] \\
&\overset{(b)}{=} \frac{1}{M^2} \sum_{m=1}^{M} \mathbb{E}\left[\left\| \left(\mathbf{g}_m^{t-1} - \nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1})\right) \right\|^2 \right] \\
&\overset{(c)}{\leq} \frac{1}{M} \sum_{l=1}^{L} \sigma_l^2,
\end{aligned} \tag{16}$$

where (a) is due to $\|\boldsymbol{x}_1 + \boldsymbol{x}_2 + \cdots + \boldsymbol{x}_M\|^2 = \|\boldsymbol{x}_1\|^2 + \|\boldsymbol{x}_2\|^2 + \cdots + \|\boldsymbol{x}_M\|^2 + \sum_{m_1=1}^{M} \sum_{m_2=1, m_2 \neq m_1}^{M} \boldsymbol{x}_{m_1} \boldsymbol{x}_{m_2}$ and $\triangle_{m_1} = \mathbf{g}_{m_1}^{t-1} - \nabla \mathcal{L}_{m_1}(\boldsymbol{w}_{m_1}^{t-1})$, inequality (b) stems from **Assumption 4** and inequality (c) stems from **Assumption 2**.

2) Bound of Y:

$$\begin{aligned}
Y &\overset{(a)}{\leq} 2\mathbb{E}\left[\frac{\left\| \sum_{m=1}^{M} \alpha_m^{t-1} \left(\nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1}) - \nabla \mathcal{L}(\boldsymbol{w}_m^{t-1}) \right) \right\|^2}{M^2} \right] \\
&+ 2\mathbb{E}\left[\frac{\left\| \sum_{m=1}^{M} \alpha_m^{t-1} \nabla \mathcal{L}(\boldsymbol{w}_m^{t-1}) \right\|^2}{M^2} \right] \\
&\overset{(b)}{=} 2 \underbrace{\mathbb{E}\left[\frac{\sum_{m=1}^{M} \alpha_m^{t-1} \left\| \left(\nabla \mathcal{L}_m(\boldsymbol{w}_m^{t-1}) - \nabla \mathcal{L}(\boldsymbol{w}_m^{t-1}) \right) \right\|^2}{M^2} \right]}_{Y_1} \\
&+ 2 \underbrace{\mathbb{E}\left[\frac{\sum_{m_1=1}^{M} \sum_{m_2=1, m_2 \neq m_1}^{M} \alpha_{m_1}^{t-1} \alpha_{m_2}^{t-1} \Phi_{m_1} \Phi_{m_2}}{M^2} \right]^2}_{Y_2}
\end{aligned} \tag{17}$$

$$+2\,\mathbb{E}\underbrace{\left[\frac{\left\|\sum_{m=1}^{M}\alpha_m^{t-1}\left(\nabla\mathcal{L}(\boldsymbol{w}_m^{t-1})\right)\right\|^2}{M^2}\right]}_{Y_3},$$

where (a) follows from $\|\boldsymbol{x}+\boldsymbol{y}\|^2 \le 2\|\boldsymbol{x}\|^2+2\|\boldsymbol{y}\|^2$, equality (b) is similar to Eq. 16(a) and $\Phi_{m_1} = \nabla\mathcal{L}_{m_1}(\boldsymbol{w}_{m_1}^{t-1}) - \nabla\mathcal{L}(\boldsymbol{w}_{m_1}^{t-1})$. Next, we find the upper bound of Y_1, Y_2 and Y_3 as follows.

Firstly, according to **Assumption 3**, we have

$$Y_1 \le \mathbb{E}\left[\frac{\sum_{m=1}^{M}\left\|\left(\nabla\mathcal{L}_m(\boldsymbol{w}_m^{t-1})-\nabla\mathcal{L}(\boldsymbol{w}_m^{t-1})\right)\right\|^2}{M^2}\right] \le \frac{1}{M^2}\sum_{m=1}^{M}\Lambda_m^2. \tag{18}$$

Secondly, the upper bound of Y_2 can be derived as follows

$$\begin{aligned}
Y_2 = \mathbb{E}\Bigg[&\sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}} Pr\Bigg(\alpha_{m_1}^{t-1}=1\forall m_1\in\mathcal{M}_1,\alpha_{m_2}^{t-1}=0\\
&\forall m_2\in\mathcal{M}_2\Bigg|\sum_{m=1}^{M}\alpha_m^{t-1}\ge 2\Bigg)\frac{1}{M^2}\sum_{m_1\in\mathcal{M}_1}\sum_{\substack{m_1'\in\mathcal{M}_1\\m_1'\neq m_1}}\Phi_{m_1}\Phi_{m_1'}\Bigg]\\
\le \mathbb{E}\Bigg[&\sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}} Pr\Bigg(\alpha_{m_1}^{t-1}=1\forall m_1\in\mathcal{M}_1,\alpha_{m_2}^{t-1}=0\\
&\forall m_2\in\mathcal{M}_2\Bigg|\sum_{m=1}^{M}\alpha_m^{t-1}\ge 2\Bigg)\frac{1}{l^2}\sum_{\substack{m_1'\in\mathcal{M}_1\\m_1'\neq m_1}}\Phi_{m_1}\Phi_{m_1'}\Bigg]
\end{aligned} \tag{19}$$

$$\begin{aligned}
&= \mathbb{E}\Bigg[\sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\frac{1}{l^2}\sum_{\substack{m_1'\in\mathcal{M}_1\\m_1'\neq m_1}}\Phi_{m_1}\Phi_{m_1'}\\
&\quad\cdot\frac{\prod_{m_1\in\mathcal{M}_1}q_{s,m_1}\prod_{m_2\in\mathcal{M}_2}(1-q_{s,m_2})}{1-\prod_{m\in\mathcal{M}}(1-q_{s,m})-\sum_{m\in\mathcal{M}}q_{s,m}\prod_{\substack{m'\in M\\m'\neq m}}(1-q_{s,m'})}\Bigg]\\
&\le \sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\underbrace{\mathbb{E}\Bigg[\sum_{\substack{m_1'=1\\m_1'\neq m_1}}^{M_1}q_{s,m_1}\Phi_{m_1}q_{s,m\prime}\Phi_{m\prime}\Bigg]}_{Y_{21}}\cdot\frac{1}{l^2}\\
&\quad\cdot\frac{\prod_{m_2\in\mathcal{M}_2}(1-q_{s,m_2})}{1-\prod_{m\in\mathcal{M}}(1-q_{s,m})-\sum_{m\in\mathcal{M}}q_{s,m}\prod_{\substack{m'\in M\\m'\neq m}}(1-q_{s,m'})}
\end{aligned}$$

According to [1], We formally define the mean successful sensing probability as $\bar{q}_s = \frac{1}{M}\sum_{m\in\mathcal{M}}q_{s,m}$ and have

$$Y_{21} \le 2\,(l-1)\sum_{m_1\in\mathcal{M}_1}\left((q_{s,m_1}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_{m_1}^2. \tag{20}$$

Then, we substitute E.q. (20) into E.q. (19) and we can obtain

$$\begin{aligned}
Y_2 &\le \sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\sum_{m_1\in\mathcal{M}_1}\left((q_{s,m_1}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_{m_1}^2\frac{2\,(l-1)}{l^2}\\
&\quad\cdot\frac{\prod_{m_2\in\mathcal{M}_2}(1-q_{s,m_2})}{1-\prod_{m\in\mathcal{M}}(1-q_{s,m})-\sum_{m\in\mathcal{M}}q_{s,m}\prod_{\substack{m'\in M\\m'\neq m}}(1-q_{s,m'})}\\
&< \sum_{l=2}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\sum_{m_1\in\mathcal{M}_1}\left((q_{s,m_1}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_{m_1}^2\cdot\frac{2}{l}\\
&\quad\cdot\frac{\prod_{m_2\in\mathcal{M}_2}(1-q_{s,m_2})}{1-\prod_{m\in\mathcal{M}}(1-q_{s,m})-\sum_{m\in\mathcal{M}}q_{s,m}\prod_{\substack{m'\in\mathcal{M}\\m'\neq m}}(1-q_{s,m'})}\\
&\triangleq \sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2,
\end{aligned} \tag{21}$$

where inequality (a) is referred to [1]. Finally, we derive the upper bound of Y_3 as follows

$$
\begin{aligned}
Y_3 &\le 2\mathbb{E}\left[\frac{\left\|\sum_{m=1}^{M}\alpha_m^{t-1}\left(\nabla\mathcal{L}(\boldsymbol{w}_m^{t-1})-\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right)\right\|^2}{M^2}\right] \\
&\quad + 2\mathbb{E}\left[\frac{\left\|\sum_{m=1}^{M}\alpha_m^{t-1}\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right\|^2}{M^2}\right] \\
&\le 2\mathbb{E}\left[\frac{\sum_{m=1}^{M}\left(\alpha_m^{t-1}\right)^2\sum_{m=1}^{M}\left\|\left(\nabla\mathcal{L}(\boldsymbol{w}_m^{t-1})-\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right)\right\|^2}{M^2}\right] \\
&\quad + 2\mathbb{E}\left[\frac{\sum_{m=1}^{M}\left(\alpha_m^{t-1}\right)^2\sum_{m=1}^{M}\left\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right\|^2}{M^2}\right] \\
&\le 2\mathbb{E}\left[\frac{\sum_{m=1}^{M}\left\|\left(\nabla\mathcal{L}(\boldsymbol{w}_m^{t-1})-\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right)\right\|^2}{M}\right] \\
&\quad + 2\mathbb{E}\left[\frac{\sum_{m=1}^{M}\left\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\right\|^2}{M}\right] \\
&\le 2\beta^2\mathbb{E}\left[\frac{\sum_{m=1}^{M}\|\boldsymbol{w}^{t-1}-\boldsymbol{w}_m^{t-1}\|^2}{M}\right] + 2\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right] \\
&\le 8\beta^2\eta^2 I^2\sum_{l=1}^{L_c}G_l^2 + 2\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right].
\end{aligned}
\tag{22}
$$

Therefore, the upper bound of Y can be represented by

$$
\begin{aligned}
Y < \frac{2}{M^2}\sum_{m=1}^{M}\Lambda_m^2 + 2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2 \\
+ 16\beta^2\eta^2 I^2\sum_{l=1}^{L_c}G_l^2 + 4\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right].
\end{aligned}
\tag{23}
$$

Add X to Y, and we can obtain

$$
\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{w}^t-\boldsymbol{w}^{t-1}\|^2\right] < \frac{2\eta^2\sum_{l=1}^{L}\sigma_l^2}{M} + \frac{4\eta^2}{M^2}\sum_{m=1}^{M}\Lambda_m^2 \\
+ 4\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2 \\
+ 32\beta^2\eta^4 I^2\sum_{l=1}^{L_c}G_l^2 + 8\eta^2\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right].
\end{aligned}
\tag{24}
$$

APPENDIX D

Based on the smoothness of the loss function $\mathcal{L}(\cdot)$, for any training round $t \ge 0$, the second-order Taylor expansion of $\mathcal{L}(\cdot)$ can be expressed as

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^t\right)\right] \le \mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right] + \frac{\beta}{2}\mathbb{E}\left[\|\boldsymbol{w}^t-\boldsymbol{w}^{t-1}\|^2\right] \\
+ \mathbb{E}\left[\left\langle\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right), \boldsymbol{w}^t-\boldsymbol{w}^{t-1}\right\rangle\right].
\end{aligned}
\tag{25}
$$

And according to **Lemma 2** and **Lemma 3**, we have

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^t\right)\right] &\le \mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right] + \frac{\beta\eta^2\sum_{l=1}^{L}\sigma_l^2}{M} \\
&\quad + \frac{2\beta\eta^2}{M^2}\sum_{m=1}^{M}\Lambda_m^2 + 16\beta^3\eta^4 I^2\sum_{l=1}^{L}G_l^2 \\
&\quad + 2\beta\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2 \\
&\quad + \eta\left(\sum_{m=1}^{M}\Lambda_m^2 + 4M\beta^2\eta^2 I^2\sum_{l=1}^{L}G_l^2\right) \\
&\quad - \frac{\eta}{2}\mathbb{E}\left[\left\|\sum_{m=1}^{M}\varphi_m\nabla\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right\|^2\right] \\
&\quad + 4\beta\eta^2\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right].
\end{aligned}
\tag{26}
$$

Next, we rearrange Eq. (26) and divide its both sides by $\frac{\eta\left(\sum_{m=1}^{M}\varphi_m\right)^2-8\beta\eta^2}{2}$:

$$
\begin{aligned}
&\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right] \le \frac{2}{\eta\left(\sum_{m=1}^{M}\varphi_m\right)^2-8\beta\eta^2}\cdot \\
&\left(\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^{t-1}\right)\right]-\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^t\right)\right] + \frac{\beta\eta^2\sum_{l=1}^{L}\sigma_l^2}{M}\right. \\
&+ \beta^2\eta^3 I^2\left(16\beta\eta+4M\right)\sum_{l=1}^{L_c}G_l^2 + \frac{2\beta\eta^2}{M^2}\sum_{m=1}^{M}\Lambda_m^2 \\
&\left.+ 2\beta\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2 + \sum_{m=1}^{M}\eta\Lambda_m^2\right).
\end{aligned}
\tag{27}
$$

For ensuring the convergence of split federated learning process, we let $\eta\left(\sum_{m=1}^{M}\varphi_m\right)^2 - 8\beta\eta^2 > 0$, which leads to $0 < \eta < \frac{\left(\sum_{m=1}^{M}\varphi_m\right)^2}{8\beta}$. Furthermore, adding up the aforementioned terms from $t = 1$ to N and then dividing both sides by

N results in

$$\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\left[\|\nabla\mathcal{L}(\boldsymbol{w}^{t-1})\|^2\right] \leq \frac{2}{\eta\left(\sum\limits_{m=1}^{M}\varphi_m\right)^2 - 8\beta\eta^2}\cdot \left(\frac{\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^0\right)\right]-\mathbb{E}\left[\mathcal{L}\left(\boldsymbol{w}^*\right)\right]}{N} + \frac{\beta\eta^2\sum\limits_{l=1}^{L}\sigma_l^2}{M} + \beta^2\eta^3 I^2\left(16\beta\eta+4M\right)\sum_{l=1}^{L_c}G_l^2 + \frac{2\beta\eta^2}{M^2}\sum_{m=1}^{M}\Lambda_m^2 + 2\beta\eta^2\sum_{m=1}^{M}\kappa_m\left((q_{s,m}-\bar{q}_s)^2+\bar{q}_s^2\right)\Lambda_m^2 + \sum_{m=1}^{M}\eta\Lambda_m^2\right), \tag{28}$$

where $\boldsymbol{w}^*$ is the optimal model.

APPENDIX E

We assume that each UAV has an identical successful sensing probability $q_{s,m}=q_s, \forall m \in \mathcal{M}$, thus we can have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{l=1}^{M}\sum_{\substack{|\mathcal{M}_1|=l\\|\mathcal{M}_2|=M-l}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\frac{1}{M}\sum_{m_1\in\mathcal{M}_1}\mathbf{g}_{m_1}^{t-1} \cdot\frac{\prod\limits_{m_1\in\mathcal{M}_1}q_{s,m_1}\prod\limits_{m_2\in\mathcal{M}_2}(1-q_{s,m_2})}{1-\prod\limits_{m\in\mathcal{M}}(1-q_{s,m})}\right]\\
&=\frac{1}{M}\mathbb{E}\left[\sum_{m=1}^{M}\frac{q_s^m(1-q_s)^{M-m}}{1-(1-q_s)^M}\sum_{\substack{|\mathcal{M}_1|=m\\|\mathcal{M}_2|=M-m}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\sum_{m_1\in\mathcal{M}_1}\mathbf{g}_{m_1}^{t-1}\right]\\
&\overset{(a)}{=}\frac{1}{M}\mathbb{E}\left[\sum_{m=1}^{M}\frac{q_s^m(1-q_s)^{M-m}}{1-(1-q_s)^M}C_{M-1}^{m-1}\sum_{m\in\mathcal{M}}\mathbf{g}_{m}^{t-1}\right]\\
&=\frac{q_s}{[1-(1-q_s)^M]\cdot M}\mathbb{E}\left[\sum_{m=0}^{M-1}q_s^m(1-q_s)^{M-m-1} C_{M-1}^{m}\sum_{m\in\mathcal{M}}\mathbf{g}_{m}^{t-1}\right]\\
&=\frac{q_s}{[1-(1-q_s)^M]\cdot M}\sum_{m\in\mathcal{M}}\mathbb{E}\left[\mathbf{g}_m^{t-1}\right],
\end{aligned}\tag{29}$$

where (a) follows from $\sum\limits_{\substack{|\mathcal{M}_1|=m\\|\mathcal{M}_2|=M-m}}^{\mathcal{M}_1\cup\mathcal{M}_2=\mathcal{M}}\sum\limits_{m_1\in\mathcal{M}_1}\mathbf{g}_{m_1}^{t-1} = C_{M-1}^{m-1}\sum\limits_{m\in\mathcal{M}}\mathbf{g}_m^{t-1}$. Therefore, we have $\varphi_m=\frac{q_s}{[1-(1-q_s)^M]\cdot M}$.

REFERENCES

[1] Y. Tang, G. Zhu, W. Xu, M. H. Cheung, T.-M. Lok, and S. Cui, "Integrated sensing, computation, and communication for uav-assisted federated edge learning," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2025.