- **Assumption 1:** $\nabla F(\boldsymbol{w})$ *is uniformly L-Lipschitz continuous with respect to* $\boldsymbol{w}$*, which can be given by*

$$\left\|\nabla F\left(\boldsymbol{w}^{t+1}\right)-\nabla F\left(\boldsymbol{w}^{t}\right)\right\| \leqslant L\left\|\boldsymbol{w}^{t+1}-\boldsymbol{w}^{t}\right\|, \quad (17)$$

where L *is Lipschitz constant depending on* $F(\cdot)$.

- **Assumption 2:** $\nabla F(\boldsymbol{w})$ *is twice-continuously differentiable. Considering both* Assumption 1 *and* Assumption 2*, the following inequality can be established:*

$$\gamma \boldsymbol{I} \leqslant \nabla^2 F(\boldsymbol{w}) \leqslant L\boldsymbol{I}, \quad (18)$$

where $\mathbf{I}$ *is an identity matrix.*

- **Assumption 3:** *The second moments of local gradient and parameters are constrained by*

$$\left\|\nabla f\left(\boldsymbol{w}^{t}, \boldsymbol{x}_{u,k}, \boldsymbol{y}_{u,k}\right)\right\|^{2} \leqslant G^{2}, \quad (19)$$

and

$$\mathbb{E}\{\|\boldsymbol{w}\|^2\} \leqslant D^2. \quad (20)$$

- **Assumption 4:** *The stochastic gradients are unbiased, which can be represented as*

$$\mathbb{E}\{g(\boldsymbol{w})\} = \nabla F(\boldsymbol{w}). \quad (21)$$

Appendix A
Proof of Lemma 2

The following two inequalities will be frequently used in the subsequent derivations.
Cauchy-Schwarz inequality

$$\sum_{k=1}^{n} {a_k}^2 \sum_{k=1}^{n} {b_k}^2 \geqslant \left(\sum_{k=1}^{n} a_k b_k\right)^2, \quad (A.1)$$

for $a_k \geqslant 0$

$$\sum_{k=1}^{n} {a_k}^2 \leqslant \left(\sum_{k=1}^{n} a_k\right)^2, \quad (A.2)$$

and the triangle inequality of the Euclidean norm

$$\|\boldsymbol{X}+\boldsymbol{Y}\| \leqslant \|\boldsymbol{X}\| + \|\boldsymbol{Y}\|. \quad (A.3)$$

Let $\hat{\boldsymbol{g}}^t$ denote the gradient reconstructed by compressed sensing. According to [**?**], the reconstruction error can be expressed as $\mathbb{E}\left\|\hat{\boldsymbol{g}}^{\boldsymbol{t}}-\tilde{\boldsymbol{g}}^{\boldsymbol{t}}\right\|^2$.Based on 1-bit quantization, feedback, and device dropout, the reconstruction error can be expressed as

$$\mathbb{E}\left\|\hat{\boldsymbol{g}}^{\boldsymbol{t}}-\tilde{\boldsymbol{g}}^{\boldsymbol{t}}\right\|^2 \leqslant \frac{C^2}{S}\mathbb{E}\left\|\bar{\boldsymbol{g}}^t+\boldsymbol{r}^t-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^t-\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^{t-1}|_{\alpha^t=0}\right)\right\|^2, \quad (A.4)$$

where C is the constant depending on the properties of $\boldsymbol{\Phi}$ and $0<\delta<1$ is a constant of the RIP condition. We can express the total quantization error as

$$\begin{aligned} &E\left\|\bar{\boldsymbol{g}}^t+\boldsymbol{r}^t-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^t-\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^{t-1}|_{\alpha^t=0}\right)\right\|^2 \\ &=\mathbb{E}\left\|\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left(\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}|_{\alpha_u^{t-1}=0}\right)}{\sum_{u=1}^{U} K_u\alpha_u^t}\right. \\ &\qquad \left.+\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left(\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t}\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t}\right)}{\sum_{u=1}^{U} K_u\alpha_u^t}\right\|^2. \end{aligned} \quad (A.5)$$

We will separately provide the upper bounds for the two aforementioned parts. First, we derive the upper bound of the gradient before CS and 1-bit quantization based on Assumption 3.

$$\begin{aligned} \mathbb{E}\|\boldsymbol{g}\|^2 &= \mathbb{E}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla f\left(\boldsymbol{w}_u;\boldsymbol{x}_k,\boldsymbol{y}_k\right)\right\|^2 \\ &\overset{(c)}{\leqslant} \frac{1}{K^2}\mathbb{E}\left\{\sum_{k=1}^{K}\left\|\nabla f\left(\boldsymbol{w}_u;\boldsymbol{x}_k,\boldsymbol{y}_k\right)\right\|^2\right\} \overset{(d)}{\leqslant} \frac{G^2}{K}. \end{aligned} \quad (A.6)$$

where (c) comes from Eq. (A.3), and (d) arises from Assumption 3. Next, we derive the upper bound of the error caused by 1-bit quantization. Measurement matrix $\boldsymbol{\Phi}$ obeys ”restricted isometry hypothesis”, which means

$$(1-\delta)\|\boldsymbol{g}\|^2 \leqslant \|\boldsymbol{\Phi}\boldsymbol{g}\|^2 \leqslant (1+\delta)\|\boldsymbol{g}\|^2. \quad (A.7)$$

In the t-th iteration, the upper bound of the error caused by 1-bit quantization of each device’s gradient is given by

$$\begin{aligned} \mathbb{E}\left\|\boldsymbol{e}_{u,q}^t\right\|^2 &= \mathbb{E}\left\|\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right\|^2 \\ &\overset{(d)}{\leqslant} 2\mathbb{E}\left(\left\|\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right)\right\|^2+\left\|\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right\|^2\right) \\ &\leqslant 2S+2(\delta+1)\frac{G^2}{K}, \end{aligned} \quad (A.8)$$

where (d) is due to the fact that $(a-b)^2 \leqslant 2\left(a^2+b^2\right)$ and the error of the accumulated gradient is given by

$$\begin{aligned} \mathbb{E}\left\|\boldsymbol{e}_{qr}^t\right\|^2 &= \mathbb{E}\left\|\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}|_{\alpha_u^{t-1}=0}\right\|^2 \\ &\leqslant 2\mathbb{E}\left(\left\|\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right)\right\|^2+\left\|\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right\|^2|_{\alpha_u^{t-1}=0}\right) \\ &\leqslant 2S\sum_{u=1}^{U} q_u+2(\delta+1)\frac{G^2}{K}\sum_{u=1}^{U} q_u. \end{aligned} \quad (A.9)$$

The error after considering device dropout can be expressed as

$$\begin{aligned} &\mathbb{E}\left\|\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left(\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^t\right)}{\sum_{u=1}^{U} K_u\alpha_u^t}\right\|^2 \\ &\overset{(e)}{\leqslant} \mathbb{E}\left\{\left(\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left\|\boldsymbol{e}_{u,q}^t\right\|}{\sum_{u=1}^{U} K_u\alpha_u^t}\right)^2\right\} \leqslant max(\mathbb{E}\left\|\boldsymbol{e}_{u,q}^t\right\|^2) \\ &\qquad \leqslant 2S+2(\delta+1)\frac{G^2}{K}, \end{aligned} \quad (A.10)$$

and

$$\begin{aligned} &\mathbb{E}\left\|\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left(\boldsymbol{sign}\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}\right)-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}_u^{t-1}|_{\alpha_u^{t-1}=0}\right)}{\sum_{u=1}^{U} K_u\alpha_u^t}\right\|^2 \\ &\leqslant \mathbb{E}\left\{\left(\frac{\sum_{u=1}^{U} K_u\alpha_u^t\left\|\boldsymbol{e}_{qr}^t\right\|}{\sum_{u=1}^{U} K_u\alpha_u^t}\right)^2\right\} \leqslant max(\mathbb{E}\left\|\boldsymbol{e}_{qr}^t\right\|^2) \\ &\qquad 2S\sum_{u=1}^{U} q_u+2(\delta+1)\frac{G^2}{K}\sum_{u=1}^{U} q_u. \end{aligned} \quad (A.11)$$

Substituting Eq. (A.10) and Eq. (A.11) into Eq. (A.5), we can obtain

$$\begin{aligned}&\mathbb{E}\left\|\hat{\boldsymbol{g}}^t-\tilde{\boldsymbol{g}}^t\right\|^2 \leqslant \frac{C^2}{S}\mathbb{E}\left\|\bar{\boldsymbol{g}}^t+\boldsymbol{r}^t-\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^t-\left(\boldsymbol{\Phi}\tilde{\boldsymbol{g}}^{t-1}|_{\alpha^t=0}\right)\right\|^2\\&\leqslant 6C^2+\frac{6(\delta+1)C^2G^2}{KS}+6C^2\sum_{u=1}^{U}q_u\\&+\frac{6(\delta+1)C^2G^2\sum_{u=1}^{U}q_u}{KS}.\end{aligned} \tag{A.12}$$

Appendix B
Proof of Theorem 1

When the BS obtains the information from local devices, it reconstructs the global gradient $\hat{\boldsymbol{g}}^t$ to update the FL model as

$$\boldsymbol{w}^{t+1}=\boldsymbol{w}^t-\eta\hat{\boldsymbol{g}}^t. \tag{B.1}$$

Performing a second-order Taylor expansion of $F\left(\boldsymbol{w}^{t+1}\right)$ at $\boldsymbol{w}^t$ and using Assumption 2, we can obtain

$$\begin{aligned}F\left(\boldsymbol{w}^{t+1}\right)&\leqslant F\left(\boldsymbol{w}^t\right)+\left(\nabla F\left(\boldsymbol{w}^t\right)\right)^\top\left(\boldsymbol{w}^{t+1}-\boldsymbol{w}^t\right)\\&\quad+\frac{\nabla^2F\left(\boldsymbol{w}^t\right)}{2}\left\|\boldsymbol{w}^{t+1}-\boldsymbol{w}^t\right\|^2\\&\leqslant F\left(\boldsymbol{w}^t\right)+\left(\nabla F\left(\boldsymbol{w}^t\right)\right)^\top\left(\boldsymbol{w}^{t+1}-\boldsymbol{w}^t\right)\\&\quad+\frac{L}{2}\left\|\boldsymbol{w}^{t+1}-\boldsymbol{w}^t\right\|^2\\&\leqslant F\left(\boldsymbol{w}^t\right)-\eta\left(\nabla F\left(\boldsymbol{w}^t\right)\right)^\top\hat{\boldsymbol{g}}^t+\frac{L\eta^2}{2}\left\|\hat{\boldsymbol{g}}^t\right\|^2,\end{aligned} \tag{B.2}$$

and

$$\begin{aligned}&-\eta\left(\nabla F\left(\boldsymbol{w}^t\right)\right)^\top\hat{\boldsymbol{g}}^t\\&=\frac{\eta}{2}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t\right\|^2-\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2-\left\|\hat{\boldsymbol{g}}^t\right\|^2\right\}\\&\leqslant\frac{\eta}{2}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t\right\|^2-\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2\right\}\\&=\frac{\eta}{2}\left\|\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t\right\|^2-\frac{\eta}{2}\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2.\end{aligned} \tag{B.3}$$

So we obtain

$$\begin{aligned}\mathbb{E}\left\{F\left(\boldsymbol{w}^{t+1}\right)\right\}\leqslant\mathbb{E}\Big\{&F\left(\boldsymbol{w}^t\right)+\frac{\eta}{2}\left\|\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t\right\|^2\\&-\frac{\eta}{2}\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2+\frac{L\eta^2}{2}\left\|\hat{\boldsymbol{g}}^t\right\|^2\Big\}.\end{aligned} \tag{B.4}$$

For the sake of simplicity in writing, we introduce an auxiliary variable as

$$\boldsymbol{\lambda}^t=\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t, \tag{B.5}$$

which can be bounded by

$$\begin{aligned}&\mathbb{E}\left\|\boldsymbol{\lambda}^t\right\|^2=\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\hat{\boldsymbol{g}}^t\right\|^2\right\}\\&=\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\tilde{\boldsymbol{g}}^t(\tilde{\boldsymbol{w}}^t)-\left(\hat{\boldsymbol{g}}^t-\tilde{\boldsymbol{g}}^t\right)\right\|^2\right\}\\&\leqslant 2\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\tilde{\boldsymbol{g}}^t(\tilde{\boldsymbol{w}}^t)\right\|^2\right\}+2\mathbb{E}\left\{\left\|\hat{\boldsymbol{g}}^t-\tilde{\boldsymbol{g}}^t\right\|^2\right\}.\end{aligned} \tag{B.6}$$

We have provided the upper bound for $\mathbb{E}\left\{\left\|\hat{\boldsymbol{g}}^t-\tilde{\boldsymbol{g}}^t\right\|^2\right\}$ in Appendix A, then we will present the upper bound for the first part. We introduce an auxiliary variable as

$$\nabla F(\boldsymbol{w}^t)'=\frac{\sum\limits_{u=1}^{U}K_u\alpha_u^t\nabla F_u\left(\boldsymbol{w}_u^t\right)}{\sum_{u=1}^{U}K_u\alpha_u^t}, \tag{B.7}$$

so we can obtain

$$\begin{aligned}&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F(\tilde{\boldsymbol{w}}^t)\right\|^2\right\}\\&=\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F\left(\boldsymbol{w}^t\right)'+\nabla F\left(\boldsymbol{w}^t\right)'-\nabla F(\tilde{\boldsymbol{w}}^t)\right\|^2\right\}\\&\leqslant\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F\left(\boldsymbol{w}^t\right)'\right\|^2+\left\|\nabla F\left(\boldsymbol{w}^t\right)'-\nabla F(\tilde{\boldsymbol{w}}^t)\right\|^2\right\}.\end{aligned} \tag{B.8}$$

Next, we will provide the upper bounds for the two parts above separately. Let U_1 represents the set of devices participating in the global iteration, and $|U_1|$ denotes the number of devices. Conversely, U_2 represents the set of devices that are not involved in the global iteration, and $|U_2|$ denotes the number of devices.

For the first part $\mathbb{E}\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F\left(\boldsymbol{w}^t\right)'\right\|^2$, we have

$$\begin{aligned}&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F(\boldsymbol{w}^t)'\right\|^2\right\}\\&=\mathbb{E}\left\{\left\|\frac{\sum\limits_{u=1}^{U}K_u\nabla F_u\left(\boldsymbol{w}_u^t\right)}{K}-\frac{\sum\limits_{u=1}^{U}K_u\alpha_u^t\nabla F_u\left(\boldsymbol{w}_u^t\right)}{\sum_{u=1}^{U}K_u\alpha_u^t}\right\|^2\right\},\end{aligned} \tag{B.9}$$

based on Eq. (8), we can obtain

$$\begin{aligned}&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)-\nabla F(\boldsymbol{w}^t)'\right\|^2\right\}\\&\overset{(f)}{\leqslant}\mathbb{E}\left\{\left\|\sum_{u=1}^{U}\left(\frac{1}{K}-\frac{\alpha_u^t}{\sum\limits_{u=1}^{U}K_u\alpha_u^t}\right)\cdot\right.\right.\\&\qquad\qquad\left.\left.\sum_{k=1}^{K_u}\nabla f\left(\boldsymbol{w}_u;\boldsymbol{x}_{u,k},\boldsymbol{y}_{u,k}\right)\right\|^2\right\}\\&\overset{(g)}{\leqslant}\mathbb{E}\left\{\sum_{u=1}^{U}\left\|\frac{1}{K}-\frac{\alpha_u^t}{\sum\limits_{u=1}^{U}K_u\alpha_u^t}\right\|^2\cdot\right.\\&\qquad\qquad\left.\sum_{u=1}^{U}\left\|\sum_{k=1}^{K_u}\nabla f\left(\boldsymbol{w}_u;\boldsymbol{x}_{u,k},\boldsymbol{y}_{u,k}\right)\right\|^2\right\}\end{aligned} \tag{B.10}$$

where (f) arises from Eq. (8), (g) comes from Eq. (A.1). We use Assumption 3 to obtain

$$
\begin{aligned}
&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\nabla F(\boldsymbol{w}^{t})'\right\|^{2}\right\} \\
&\overset{(h)}{\leqslant} \mathbb{E}\left\{\sum_{u=1}^{U}\left\|\frac{1}{K}-\frac{\alpha_{u}^{t}}{\sum\limits_{u=1}^{U} K_{u} \alpha_{u}^{t}}\right\|^{2} \sum_{u=1}^{U} K_{u} G^{2}\right\} \\
&\leqslant K G^{2} \mathbb{E}\left\{\sum_{u=1}^{U}\left(\frac{1}{K}-\frac{\alpha_{u}^{t}}{\sum\limits_{u=1}^{U} K_{u} \alpha_{u}^{t}}\right)^{2}\right\} \\
&\leqslant K G^{2} \sum_{i=1}^{U} \sum_{\substack{|U_1|=i \\ |U_2|=U-i}}\left\{\frac{1}{K}-\right. \\
&\left.\left(\prod_{u_1 \in U_1}\left(1-q_{u_1}^{t}\right) \prod_{u_2 \in U_2} q_{u_2}^{t} \frac{1}{\sum\limits_{u_1 \in U_1} K_{u_1}}\right)\right\}^{2},
\end{aligned} \tag{B.11}
$$

For the second part

$$
\begin{aligned}
&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)'-\nabla F(\tilde{\boldsymbol{w}}^{t})\right\|^{2}\right\} \\
&=\mathbb{E}\left\{\left\|\frac{\sum\limits_{u=1}^{U} K_{u} \alpha_{u}^{t}\left(\nabla F_{u}\left(\boldsymbol{w}_{u}^{t}\right)-\nabla F_{u}\left(\tilde{\boldsymbol{w}}_{u}^{t}\right)\right)}{\sum\limits_{u=1}^{U} K_{u} \alpha_{u}^{t}}\right\|^{2}\right\} \\
&\overset{(i)}{\leqslant} \mathbb{E}\left\{\frac{\left(\sum_{u=1}^{U}\left\|K_{u} \alpha_{u}^{t}\right\|^{2}\right)\left(\sum_{u=1}^{U}\left\|\boldsymbol{w}_{u}^{t}-\tilde{\boldsymbol{w}}_{u}^{t}\right\|^{2}\right)}{\left\|\sum_{u=1}^{U} K_{u} \alpha_{u}^{t}\right\|^{2}}\right\} L^{2} \\
&=\frac{\left(\sum_{u=1}^{U}\left\|K_{u} \alpha_{u}^{t}\right\|^{2}\right)\left(\sum_{u=1}^{U} \mathbb{E}\left\{\left\|\boldsymbol{w}_{u}^{t}-\tilde{\boldsymbol{w}}_{u}^{t}\right\|^{2}\right\}\right)}{\left\|\sum_{u=1}^{U} K_{u} \alpha_{u}^{t}\right\|^{2}} L^{2} \\
&\overset{(j)}{\leqslant} \frac{\left(\sum_{u=1}^{U}\left\|K_{u} \alpha_{u}^{t}\right\|^{2}\right)\left(\sum_{u=1}^{U} \rho_{u}^{t} D^{2}\right)}{\left\|\sum_{u=1}^{U} K_{u} \alpha_{u}^{t}\right\|^{2}} L^{2} \\
&\overset{(k)}{\leqslant} \frac{\left(\sum_{u=1}^{U}\left\|K_{u} \alpha_{u}^{t}\right\|^{2}\right)\left(\sum_{u=1}^{U} \rho_{u}^{t}\right)}{\sum_{u=1}^{U}\left\|K_{u} \alpha_{u}^{t}\right\|^{2}} L^{2} D^{2} \triangleq L^{2} D^{2} \Gamma^{t}.
\end{aligned} \tag{B.12}
$$

(i) is due to Assumption 1 and (k) is derived from Eq. (A.2). And we can obtain (j) from [2].

Based on the results of the two parts above, we can obtain

$$
\begin{aligned}
&\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\nabla F(\tilde{\boldsymbol{w}}^{t})\right\|^{2}\right\} \\
&=\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\nabla F\left(\boldsymbol{w}^{t}\right)'+\nabla F\left(\boldsymbol{w}^{t}\right)'-\nabla F(\tilde{\boldsymbol{w}}^{t})\right\|^{2}\right\} \\
&\leqslant \mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\nabla F\left(\boldsymbol{w}^{t}\right)'\right\|^{2}+\left\|\nabla F\left(\boldsymbol{w}^{t}\right)'-\nabla F(\tilde{\boldsymbol{w}}^{t})\right\|^{2}\right\} \\
&\leqslant L^{2} D^{2} \Gamma^{t}+K G^{2} \mathbb{E}\left\{\sum_{u=1}^{U}\left(\frac{1}{K}-\frac{\alpha_{u}^{t}}{\sum\limits_{u=1}^{U} K_{u} \alpha_{u}^{t}}\right)^{2}\right\} \\
&\leqslant L^{2} D^{2} \Gamma^{t}+K G^{2} \sum_{i=1}^{U} \sum_{\substack{|U_1|=i \\ |U_2|=U-i}}\left\{\frac{1}{K}-\right. \\
&\left.\left(\prod_{u_1 \in U_1}\left(1-q_{u_1}^{t}\right) \prod_{u_2 \in U_2} q_{u_2}^{t} \frac{1}{\sum\limits_{u_1 \in U_1} K_{u_1}}\right)\right\}^{2}.
\end{aligned} \tag{B.13}
$$

$$
\begin{aligned}
&E\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\|^{2} \\
&=E\|\nabla F(\boldsymbol{w}^{t})\|^{2}+E\left\|\tilde{\boldsymbol{g}}^{t}(\tilde{\boldsymbol{w}}^{t})\right\|^{2} \\
&\qquad -2 E\left\{\left(\nabla F\left(\boldsymbol{w}^{t}\right)\right)^{\top} \tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\} \\
&=E\left\|\nabla F\left(\boldsymbol{w}^{t}\right)\right\|^{2}+E\left\|\tilde{\boldsymbol{g}}^{t}(\tilde{\boldsymbol{w}}^{t})\right\|^{2} \\
&\qquad -2 E\left\{\nabla F\left(\tilde{\boldsymbol{w}}^{t}\right)\right\}^{\top} E\left\{\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\} \\
&\qquad -2 \operatorname{Tr}\left(\operatorname{Cov}\left(\nabla F, \tilde{\boldsymbol{g}}^{t}\right)\right) \\
&\leqslant E\left\|\nabla F\left(\boldsymbol{w}^{t}\right)\right\|^{2}+E\left\|\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\|^{2} \\
&\qquad -2 E\left\{\nabla F\left(\boldsymbol{w}^{t}\right)\right\}^{\top} E\left\{\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\} \\
&=E\left\|\nabla F\left(\boldsymbol{w}^{t}\right)\right\|^{2}+E\left\|\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\|^{2} \\
&\qquad -2 E\left\{\nabla F\left(\boldsymbol{w}^{t}\right)\right\}^{\top} E\left\{\nabla F\left(\tilde{\boldsymbol{w}}^{t}\right)\right\} \\
&=E\left\{\left\|\nabla F\left(\boldsymbol{w}^{t}\right)-\nabla F(\tilde{\boldsymbol{w}}^{t})\right\|^{2}\right\}-E\|\nabla F(\tilde{\boldsymbol{w}}^{t})\|^{2} \\
&\qquad +E\left\|\tilde{\boldsymbol{g}}^{t}\left(\tilde{\boldsymbol{w}}^{t}\right)\right\|^{2} \\
&\leqslant L^{2} D^{2} \Gamma^{t}+\frac{G^{2}}{K}-E\|\nabla F(\tilde{\boldsymbol{w}}^{t})\|^{2} \\
&\qquad +K G^{2} \sum_{i=1}^{U} \sum_{\substack{|U_1|=i \\ |U_2|=U-i}}\left\{\frac{1}{K}-\right. \\
&\left.\left(\prod_{u_1 \in U_1}\left(1-q_{u_1}^{t}\right) \prod_{u_2 \in U_2} q_{u_2}^{t} \frac{1}{\sum\limits_{u_1 \in U_1} K_{u_1}}\right)\right\}^{2}
\end{aligned} \tag{B.14}
$$

Substituting Eq. (A.12) and Eq. (B.14)into Eq. (B.6), we can obtain

$$\mathbb{E}\left\|\boldsymbol{\lambda}^t\right\|^2 \leqslant 2L^2D^2\Gamma^t + 2\frac{G^2}{K} - 2E\|\nabla F(\tilde{\boldsymbol{w}}^t)\|^2 + 12C^2\sum_{u=1}^{U} q_u^t$$
$$+ 12C^2 + \frac{12(\delta+1)C^2G^2}{SK} + \frac{12(\delta+1)C^2G^2\sum_{u=1}^{U} q_u^t}{SK}$$
$$+ 2KG^2\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2$$
$$\leqslant 2L^2D^2\Gamma^t + 2\frac{G^2}{K} + 12C^2\sum_{u=1}^{U} q_u^t$$
$$+ 12C^2 + \frac{12(\delta+1)C^2G^2}{SK} + \frac{12(\delta+1)C^2G^2\sum_{u=1}^{U} q_u^t}{SK}$$
$$+ 2KG^2\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2. \tag{B.15}$$

Substituting Eq. (A.6) and Eq. (B.15) into Eq. (B.4), we can obtain

$$\mathbb{E}\left\{F\left(\boldsymbol{w}^{t+1}\right)\right\} \leqslant \mathbb{E}\left\{F\left(\boldsymbol{w}^{t}\right)\right\} + \eta L^2D^2\Gamma^t + \frac{\eta G^2}{K} + 6\eta C^2$$
$$+ \frac{6\eta(\delta+1)C^2G^2}{SK} + 6\eta C^2\sum_{u=1}^{U} q_u^t - \frac{\eta}{2}\mathbb{E}\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2$$
$$+ \frac{6\eta(\delta+1)C^2G^2\sum_{u=1}^{U} q_u^t}{SK} + \frac{L\eta^2G^2}{2K}$$
$$+ \eta KG^2\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2, \tag{B.16}$$

and we can easily get

$$\mathbb{E}\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2 \leqslant \frac{2}{\eta}\mathbb{E}\left\{F\left(\boldsymbol{w}^t\right) - F\left(\boldsymbol{w}^{t+1}\right)\right\} + 2L^2D^2\Gamma^t$$
$$+ \frac{2G^2}{K} + 12C^2 + \frac{12(\delta+1)C^2G^2}{SK} + 12C^2\sum_{u=1}^{U} q_u^t$$
$$+ \frac{12(\delta+1)C^2G^2\sum_{u=1}^{U} q_u^t}{SK} + \frac{L\eta G^2}{K}$$
$$+ 2KG^2\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2. \tag{B.17}$$

We sum and average the above equation from $t=0$ to $\Omega-1$,

$$\frac{1}{\Omega}\sum_{t=0}^{\Omega-1}\mathbb{E}\left\{\left\|\nabla F\left(\boldsymbol{w}^t\right)\right\|^2\right\} \leqslant \frac{2}{\eta\Omega}\mathbb{E}\left\{F\left(\boldsymbol{w}^0\right) - F\left(\boldsymbol{w}^*\right)\right\}$$
$$+ \frac{2L^2D^2}{\Omega}\sum_{t=0}^{\Omega-1}\Gamma^t + \frac{12(\delta+1)C^2G^2}{SK} + \frac{12C^2}{\Omega-1}\sum_{t=1}^{\Omega-1}\sum_{u=1}^{U} q_u^t$$
$$+ 12C^2 + \frac{2G^2}{K} + \frac{12(\delta+1)C^2G^2}{SK(\Omega-1)}\sum_{t=1}^{\Omega-1}\sum_{u=1}^{U} q_u^t + \frac{L\eta G^2}{K}$$
$$+ \frac{2KG^2}{\Omega}\sum_{t=0}^{\Omega-1}\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2. \tag{B.18}$$

Therefore we can obtain the upper bound of the Euclidean norm of the gradient as

$$\frac{1}{\Omega}\mathbb{E}\left\{\left\|\nabla F\left(\tilde{\boldsymbol{w}}^t\right)\right\|^2\right\} \leqslant \frac{2}{\eta\Omega}\mathbb{E}\left\{F\left(\boldsymbol{w}^0\right) - F\left(\boldsymbol{w}^*\right)\right\}$$
$$+ \frac{2L^2D^2}{\Omega}\sum_{t=0}^{\Omega-1}\Gamma^t + \frac{12(\delta+1)C^2G^2}{SK} + \frac{12C^2}{\Omega-1}\sum_{t=1}^{\Omega}\sum_{u=1}^{U} q_u^t$$
$$+ 12C^2 + \frac{2G^2}{K} + \frac{12(\delta+1)C^2G^2}{SK\Omega}\sum_{t=1}^{\Omega}\sum_{u=1}^{U} q_u^t + \frac{L\eta G^2}{K}$$
$$+ \frac{2KG^2}{\Omega}\sum_{t=0}^{\Omega-1}\sum_{i=1}^{U}\sum_{\substack{|U_1|=i\\|U_2|=U-i}}\left\{\frac{1}{K}-\right.$$
$$\left.\left(\prod_{u_1\in U_1}\left(1-q_{u_1}^t\right)\prod_{u_2\in U_2} q_{u_2}^t \frac{1}{\sum\limits_{u_1\in U_1} K_{u_1}}\right)\right\}^2, \tag{B.19}$$

where $\boldsymbol{w}^*$ represents the final model.