APPENDIX

In this section, we provide the details of proofs. Let's start by defining a new series of notations.

A. Additional Notation

In this section, we present various notations that will be utilized in subsequent proofs. During each local round t, every client k trains the local model with the embedding $\hat{\Phi}_k^{t_0}$. This process can also be interpreted as the client k training the model θ_k^t and $\theta_j^{t_0}$ for all $j \neq k$, where t_0 is the last communication iteration when client k received the embeddings. Then we have:

$$\gamma_{k,j}^t = \begin{cases} \theta_j^t, & k = j \\ \theta_j^{t_0}, & \text{otherwise} \end{cases} \tag{14}$$

Which represents the real model that client k utilized in the round t. We define the column vector $\Gamma_k^t = \left[(\gamma_{k,0}^t)^T; \ldots; (\gamma_{k,K}^t)^T\right]$ to be the client k's view of global model in the round t. Then we define $\hat{F}(\Gamma_k^t)$ to be the loss with pruning error by client k at round t:

$$\hat{F}(\Gamma_k^t) = F\left(\theta_0^{t_0}, h_1^{t_0}(\theta_1^{t_0}) + \varphi_1^{t_0} + \epsilon_1^{t_0}, \ldots, h_k^t(\theta_k^t) + \varphi_k^{t_0}, \ldots, h_K^{t_0}(\theta_K^{t_0}) + \varphi_K^{t_0} + \epsilon_K^{t_0}\right) \tag{15}$$

It is crucial to emphasize that $\theta_0^{t_0}$ is free from computation and communication pruning errors, while the expression $h_k^t(\theta_k^t)$ is unaffected by communication pruning errors. Building on this understanding, we can now reform the definition of $\hat{\mathbf{G}}^t$ as follows::

$$\hat{\mathbf{G}}^t := \left[(\nabla_0 \hat{F}(\Gamma_0^t))^T, \ldots, (\nabla_K \hat{F}(\Gamma_K^t))^T\right]^T \tag{16}$$

In the subsequent proof, we will employ both variations of $\hat{\mathbf{G}}^t$. Following that, we will define the computation pruning error associated with each embedding utilized in client k's gradient calculation during round t:

$$E_k^{t_0} := \left[(\varphi_1^{t_0} + \epsilon_1^{t_0})^T, \ldots, (\varphi_k^{t_0})^T, \ldots, (\varphi_K^{t_0} + \epsilon_K^{t_0})^T\right]^T \tag{17}$$

Given the error, we can derive the following:

$$\nabla_k F(\Phi_k^t + E_k^{t_0}) = \nabla_k \hat{F}(\Gamma_k^t) \tag{18}$$

We apply the chain rule to $\nabla_k \hat{F}(\Gamma_k^t)$ to get:

$$\nabla_k \hat{F}(\Gamma_k^t) = \nabla_{\theta_k} h_k(\theta_k^t) \nabla_{h_k(\theta_k)} F(\Phi_k^t + E_k^{t_0}) \tag{19}$$

Using the Taylor series expansion to further expand $\nabla_{h_k(\theta_k)} F(\Phi_k^t + E_k^{t_0})$ around the point Φ_k^t:

$$\nabla_{h_k(\theta_k)} F(\Phi_k^t + E_k^{t_0}) = \nabla_{h_k(\theta_k)} F(\Phi_k^t) + \nabla_{h_k(\theta_k)}^2 F(\Phi_k^t)^T E_k^{t_0} + \ldots \tag{20}$$

To convince, we let the infinite sum of all terms from the second partial derivatives as $R_0^k(\Phi_k^t + E_k^{t_0})$. Then we can get:

$$\nabla_{h_k(\theta_k)} F(\Phi_k^t + E_k^{t_0}) = \nabla_{h_k(\theta_k)} F(\Phi_k^t) + R_0^k(\Phi_k^t + E_k^{t_0}) \tag{21}$$

Building on the definition provided above, we will now move to the theoretical proof.

B. Proof of Lemma 1

By definition, we know that

$$\begin{aligned} \nabla_k \hat{F}(\Gamma_k^t) &= \nabla_k F(\Phi_k^t + E_k^{t_0}) && (22) \\ &= \nabla_{\theta_k} h_k(\theta_k^t) \nabla_{h_k(\theta_k)} F(\Phi_k^t + E_k^{t_0}) && (23) \\ &= \nabla_{\theta_k} h_k(\theta_k^t) \left(\nabla_{h_k(\theta_k)} F(\Phi_k^t) + R_0^k(\Phi_k^t + E_k^{t_0})\right) && (24) \\ &= \nabla_k F(\Gamma_k^t) + \nabla_{\theta_k} h_k(\theta_k^t) R_0^k(\Phi_k^t + E_k^{t_0}) && (25) \end{aligned}$$

Where Eq. 24 uses the Taylor series expansion. Then we can get

$$\mathbb{E}\left\|\nabla_k \hat{F}(\Gamma_k^t) - \nabla_k F(\Gamma_k^t)\right\|^2 = \mathbb{E}\left\|\nabla_{\theta_k} h_k(\theta_k^t) R_0^k(\Phi_k^t + E_k^{t_0})\right\|^2 \tag{26}$$

Thus we need to get the bound of the right side first.

$$\left\|\nabla_{\theta_k} h_k(\theta_k^t) R_0^k(\Phi_k^t + E_k^{t_0})\right\|^2 \le \left\|\nabla_{\theta_k} h_k(\theta_k^t)\right\|_{\mathcal{F}}^2 \left\|R_0^k(\Phi_k^t + E_k^{t_0})\right\|_{\mathcal{F}}^2 \le G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{27}$$

Then applying expectation we can finally get:

$$\mathbb{E}\left\|\nabla_{\theta_k} h_k(\theta_k^t) R_0^k(\Phi_k^t + E_k^{t_0})\right\|^2 \le G_k^2 H_k^2 \mathbb{E}\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{28}$$

$$\le G_k^2 H_k^2 \mathbb{E}\left[\sum_{j\neq 0, j\neq k}^{K} \left\|\varphi_j^{t_0} + \epsilon_j^{t_0}\right\|_{\mathcal{F}}^2 + \left\|\varphi_k^{t_0}\right\|_{\mathcal{F}}^2\right] \tag{29}$$

$$\le G_k^2 H_k^2 \mathbb{E}\left[\sum_{j\neq 0, j\neq k}^{K} \left(2\left\|\varphi_j^{t_0}\right\|_{\mathcal{F}}^2 + 2\left\|\epsilon_j^{t_0}\right\|_{\mathcal{F}}^2\right) + \left\|\varphi_k^{t_0}\right\|_{\mathcal{F}}^2\right] \tag{30}$$

$$\le 2G_k^2 H_k^2 \sum_{j\neq 0}^{K} \mathbb{E}\left\|\varphi_j^{t_0}\right\|_{\mathcal{F}}^2 + 2G_k^2 H_k^2 \sum_{j\neq 0, j\neq k}^{K} \mathbb{E}\left\|\epsilon_j^{t_0}\right\|_{\mathcal{F}}^2 \tag{31}$$

$$= 2G_k^2 H_k^2 \sum_{j\neq 0}^{K} \Psi_j^{t_0} + 2G_k^2 H_k^2 \sum_{j\neq 0, j\neq k}^{K} \Omega_j^{t_0} \tag{32}$$

C. Proof of Lemma 2

In this section, we will get the bound of $\sum_{t=t_0}^{t_0+E-1} \mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t - \mathbf{G}^{t_0}\right\|^2$.

$$\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t - \mathbf{G}^{t_0}\right\|^2 = \sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^t) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{33}$$

$$= \sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^t) - \nabla_k \hat{F}(\Gamma_k^{t-1}) + \nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{34}$$

$$\le (1+n)\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^t) - \nabla_k \hat{F}(\Gamma_k^{t-1})\right\|^2\right] + (1+\frac{1}{n})\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{35}$$

$$\le 2(1+n)\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^t) - \nabla_k F(\Gamma_k^{t-1})\right\|^2\right] + (1+\frac{1}{n})\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{36}$$

$$+ 2(1+n)\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_{\theta_k} h_k(\theta_k^t) R_0^k(\Phi_k^t + E_k^{t_0}) - \nabla_{\theta_k} h_k(\theta_k^{t-1}) R_0^k(\Phi_k^{t-1} + E_k^{t-1})\right\|^2\right] \tag{37}$$

$$\le 2(1+n)\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^t) - \nabla_k F(\Gamma_k^{t-1})\right\|^2\right] + (1+\frac{1}{n})\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{38}$$

$$+ 8(1+n)\sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{39}$$

Where the last term follows the Eq. 27. Then we have:

$$\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t - \mathbf{G}^{t_0}\right\|^2$$

$$\le 2(1+n)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1})\right\|^2\right] + (1+\frac{1}{n})\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{40}$$

$$+ 8(1+n)\sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{41}$$

$$\le 2(1+n)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0}) + \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{42}$$

$$+ (1+\frac{1}{n})\sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] + 8(1+n)\sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{43}$$

$$\leq 4(1+n)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{44}$$

$$+ 4(1+n)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{45}$$

$$+ (1+\frac{1}{n}) \sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] + 8(1+n) \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{46}$$

$$\leq \sum_{k=0}^{K} \left(4(1+n)(\eta^{t_0})^2 (L_k)^2 + (1+\frac{1}{n})\right) \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{47}$$

$$+ 4(1+n)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^{t_0})\right\|^2\right] + 8(1+n) \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{48}$$

Where Eq. 40 follows the assumption 1 and update rule $\Gamma_k^t - \Gamma_k^{t-1} = \eta^{t_0} \nabla_k \hat{F}(\Gamma_k^{t-1})$.
Then by set $n = E$ and $\eta^{t_0} \leq \frac{1}{4E \max_k L_k^a}$, we can get:

$$\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t - \mathbf{G}^{t_0}\right\|^2 \tag{49}$$

$$\leq \sum_{k=0}^{K} \left(\frac{(1+E)}{4E^2} + (1+\frac{1}{E})\right) \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{50}$$

$$+ 4(1+E)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^{t_0})\right\|^2\right] + 8(1+n) \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{51}$$

$$\leq \sum_{k=0}^{K} \left(1+\frac{2}{E}\right) \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t-1}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{52}$$

$$+ 4(1+E)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^{t_0})\right\|^2\right] + 8(1+n) \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{53}$$

Here, we set the new notations for each term as follows:

$$A^t = \sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{54}$$

$$B_0 = 4(1+E)(\eta^{t_0})^2 \sum_{k=0}^{K} (L_k)^2 \mathbb{E}_{t_0}\left[\left\|\nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \tag{55}$$

$$B_1 = 8(1+n) \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{56}$$

$$C = \left(1+\frac{2}{E}\right) \tag{57}$$

We can get $A^t \leq CA^{t-1} + B_0 + B_1$. By induction, we can obtain $A^t \leq C^{t-t_0-1}A^{t_0} + (B_0+B_1)\frac{C^{t-t_0-1}-1}{C-1}$.

$$A^{t_0} = \sum_{k=0}^{K} \mathbb{E}_{t_0}\left[\left\|\nabla_k \hat{F}(\Gamma_k^{t_0}) - \nabla_k F(\Gamma_k^{t_0})\right\|^2\right] \leq \sum_{k=0}^{K} G_k^2 H_k^2 \left\|E_k^{t_0}\right\|^2 \tag{58}$$

Then we get the bounds of C^{t-t_0-1} and $\frac{C^{t-t_0-1}-1}{C-1}$ with summing over the set of local iterations t_0 to t_0+E-1 respectively:

$$\sum_{t=t_0}^{t_0+E-1} C^{t-t_0-1} = \frac{C^E - 1}{C-1} \tag{59}$$

$$= \frac{\left(1+\frac{2}{E}\right)^E - 1}{\left(1+\frac{2}{E}\right) - 1} \leq \frac{\boldsymbol{e}^2 - 1}{2} E \leq 4E \tag{60}$$

We can also get

$$\sum_{t=t_0}^{t_0+E-1}\frac{C^{t-t_0-1}-1}{C-1}=\frac{1}{C-1}\left(\sum_{t=t_0}^{t_0+E-1}C^{t-t_0-1}-E\right) \tag{61}$$

$$=\frac{1}{C-1}\left(\frac{C^E-1}{C-1}-E\right)=\frac{E}{2}\left(\frac{E\left[(1+\frac{2}{E})^E-1\right]}{2}-E\right) \tag{62}$$

$$=\frac{E^2}{2}\left(\frac{\left[(1+\frac{2}{E})^E-1\right]}{2}-1\right)\leq\frac{E^2}{2}\left(\frac{\boldsymbol{e}^2-1}{2}-1\right)\leq 2E^2 \tag{63}$$

Plugging all values into, we can get:

$$\sum_{t=t_0}^{t_0+E-1}A^t\leq\sum_{t=t_0}^{t_0+E-1}C^{t-t_0-1}A^{t_0}+(B_0+B_1)\sum_{t=t_0}^{t_0+E-1}\frac{C^{t-t_0-1}-1}{C-1} \tag{64}$$

$$\leq 4E\sum_{k=0}^{K}G_k^2H_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{65}$$

$$+8E^2(1+E)(\eta^{t_0})^2\sum_{k=0}^{K}(L_k)^2\mathbb{E}_{t_0}\left[\left\|\nabla_kF(\Theta_k^{t_0})\right\|^2\right] \tag{66}$$

$$+16E^2(1+E)\sum_{k=0}^{K}G_k^2H_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{67}$$

$$\leq 16E^3(\eta^{t_0})^2\sum_{k=0}^{K}(L_k)^2\mathbb{E}_{t_0}\left[\left\|\nabla_kF(\Theta^{t_0})\right\|^2\right] \tag{68}$$

$$+4(4E^2(1+E)+E)\sum_{k=0}^{K}G_k^2H_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{69}$$

$$\leq 16E^3(\eta^{t_0})^2\sum_{k=0}^{K}(L_k)^2\mathbb{E}_{t_0}\left[\left\|\nabla_kF(\Theta^{t_0})\right\|^2\right] \tag{70}$$

$$+36E^3\sum_{k=0}^{K}G_k^2H_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{71}$$

D. *Proof of Theorem 1*

Here we will derive Theorem 1, first, we set $t_0^+=t_0+E-1$, then we get:

$$F(\Theta^{t_0^+})-F(\Theta^{t_0})\leq\left\langle\nabla F(\Theta^{t_0}),\Theta^{t_0^+}-\Theta^{t_0}\right\rangle+\frac{L}{2}\left\|\Theta^{t_0^+}-\Theta^{t_0}\right\|^2 \tag{72}$$

$$=-\left\langle\nabla F(\Theta^{t_0}),\sum_{t=t_0}^{t_0^+}\eta^{t_0}\hat{\mathbf{G}}^t\right\rangle+\frac{L}{2}\left\|\sum_{t=t_0}^{t_0^+}\eta^{t_0}\hat{\mathbf{G}}^t\right\|^2 \tag{73}$$

$$\leq-\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\langle\nabla F(\Theta^{t_0}),\hat{\mathbf{G}}^t\right\rangle+\frac{LE}{2}\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\left\|\hat{\mathbf{G}}^t\right\|^2 \tag{74}$$

$$\leq\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\langle-\nabla F(\Theta^{t_0}),\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\rangle-\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\langle\nabla F(\Theta^{t_0}),\mathbf{G}^{t_0}\right\rangle \tag{75}$$

$$+LE\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2+LE\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\left\|\mathbf{G}^{t_0}\right\|^2 \tag{76}$$

$$\leq\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\|\nabla F(\Theta^{t_0})\right\|^2+\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2-\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\langle\nabla F(\Theta^{t_0}),\mathbf{G}^{t_0}\right\rangle \tag{77}$$

$$+LE\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2+LE\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\left\|\mathbf{G}^{t_0}\right\|^2 \tag{78}$$

where the last term follows $A\cdot B=\frac{1}{2}A^2+\frac{1}{2}B^2-\frac{1}{2}(A-B)^2$. Then we apply the expectation $\mathbb{E}_{t_0}$ to both sides of the last term.

$$\mathbb{E}_{t_0}[F(\Theta^{t_0^+})]-F(\Theta^{t_0}) \tag{79}$$

$$\leq -\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}\left\|\nabla F(\Theta^{t_0})\right\|^2+\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}(1+2LE\eta^{t_0})\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2+LE\sum_{t=t_0}^{t_0^+}(\eta^{t_0})^2\mathbb{E}_{t_0}\left\|\mathbf{G}^{t_0}\right\|^2 \tag{80}$$

$$\leq -\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}(1-2LE\eta^{t_0})\left\|\nabla F(\Theta^{t_0})\right\|^2+\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}(1+2LE\eta^{t_0})\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2 \tag{81}$$

$$= -\frac{E}{2}\eta^{t_0}(1-2LE\eta^{t_0})\left\|\nabla F(\Theta^{t_0})\right\|^2+\frac{1}{2}\sum_{t=t_0}^{t_0^+}\eta^{t_0}(1+2LE\eta^{t_0})\mathbb{E}_{t_0}\left\|\hat{\mathbf{G}}^t-\mathbf{G}^{t_0}\right\|^2 \tag{82}$$

where the Eq.80 follows $\mathbf{G}^{t_0}=\nabla F(\Theta^{t_0})$ since we consider the full batch participation in each training round. If mini-batch is considered here there will have additional terms. Then with Lemma 2, we have:

$$\mathbb{E}_{t_0}[F(\Theta^{t_0^+})]-F(\Theta^{t_0}) \tag{83}$$

$$\leq -\frac{E}{2}\eta^{t_0}(1-2LE\eta^{t_0})\left\|\nabla F(\Theta^{t_0})\right\|^2+8E^3(\eta^{t_0})^3(1+2LE\eta^{t_0})\sum_{k=0}^{K}(L_k)^2\left\|\nabla_k F(\Theta_k^{t_0})\right\|^2 \tag{84}$$

$$+18E^3\eta^{t_0}(1+2LE\eta^{t_0})\sum_{k=0}^{K}H_k^2G_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{85}$$

Then by let $\eta^{t_0}\leq\frac{1}{16E\max\{L,\max_k L_k\}}$, we can get the further bound:

$$\mathbb{E}_{t_0}[F(\Theta^{t_0^+})]-F(\Theta^{t_0})\leq -\frac{E}{2}\sum_{k=0}^{K}\eta^{t_0}(1-2LE\eta^{t_0}-16E^2(L_k)^2(\eta^{t_0})^2-16E^3(L_k)^2L(\eta^{t_0})^3)\left\|\nabla_k F(\Theta_k^{t_0})\right\|^2$$

$$+18E^3\eta^{t_0}(1+2L^aE\eta^{t_0})\sum_{k=0}^{K}H_k^2G_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{86}$$

$$\leq -\frac{3E}{8}\eta^{t_0}\left\|\nabla F(\Theta^{t_0})\right\|^2+18E^3\eta^{t_0}(1+2LE\eta^{t_0})\sum_{k=0}^{K}H_k^2G_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{87}$$

Then we rearranged the terms:

$$\eta^{t_0}\left\|\nabla F(\Theta^{t_0})\right\|^2\leq\frac{3\left[F(\Theta^{t_0})-\mathbb{E}_{t_0}[F(\Theta^{t_0^+})]\right]}{E}+48E^2\eta^{t_0}(1+2LE\eta^{t_0})\sum_{k=0}^{K}H_k^2G_k^2\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2 \tag{88}$$

summing over all global round $t_0=0,\ldots,R-1$ with expectation:

$$\sum_{t_0=0}^{R-1}\eta^{t_0}\mathbb{E}[\left\|\nabla F(\Theta^{t_0})\right\|^2]\leq\frac{3\left[F(\Theta^0)-\mathbb{E}_{t_0}[F(\Theta^T)]\right]}{E}+48E^2\eta^{t_0}(1+2LE\eta^{t_0})\sum_{t_0=0}^{R-1}\sum_{k=0}^{K}H_k^2G_k^2\mathbb{E}[\left\|E_k^{t_0}\right\|_{\mathcal{F}}^2] \tag{89}$$

Then with $\eta^{t_0}=\eta$ and averaging over R global rounds, we get:

$$\frac{1}{R}\sum_{t_0=0}^{R-1}\mathbb{E}[\left\|\nabla F(\Theta^{t_0})\right\|^2]\leq\frac{3\left[F(\Theta^0)-\mathbb{E}_{t_0}[F(\Theta^T)]\right]}{\eta RE} \tag{90}$$

$$+\frac{96E^2}{R}(1+2LE\eta)\sum_{t_0=0}^{R-1}\sum_{k=0}^{K}H_k^2G_k^2\sum_{j\neq 0}^{K}\Psi_j^{t_0}+\frac{96E^2}{R}(1+2LE\eta)\sum_{t_0=0}^{R-1}\sum_{k=0}^{K}G_k^2H_k^2\sum_{j\neq 0,j\neq k}^{K}\Omega_j^{t_0} \tag{91}$$

$$\leq\frac{3\left[F(\Theta^0)-\mathbb{E}_{t_0}[F(\Theta^T)]\right]}{\eta RE}+\frac{108E^2}{R}\sum_{t_0=0}^{R-1}\sum_{k=0}^{K}H_k^2G_k^2\sum_{j\neq 0}^{K}\Psi_j^{t_0}+\frac{108E^2}{R}\sum_{t_0=0}^{R-1}\sum_{k=0}^{K}G_k^2H_k^2\sum_{j\neq 0,j\neq k}^{K}\Omega_j^{t_0} \tag{92}$$

where the last term is from $\eta=\eta^{t_0}\leq\frac{1}{16E\max\{L,\max_k L_k\}}$. This finishes the proof of Theorem 1.

E. *Proof of Corollary 1*

To get the corollary 1, we need further bound the Ω_k^t and Ψ_k^t respectively. We begin by giving the bound of Ω_k^t.

$$\Omega_k^t = \mathbb{E}\|h_k^t(\hat{\theta}_k^t) - \hat{h}_k^t(\hat{\theta}_k^t)\|^2 \tag{93}$$
$$= \mathbb{E}\|h_k^t(\hat{\theta}_k^t) \odot \mathbf{1} - h_k^t(\hat{\theta}_k^t) \odot \mathbf{l}_k^t\|^2 = \mathbb{E}\|h_k^t(\hat{\theta}_k^t) \odot (\mathbf{1} - \mathbf{l}_k^t)\|^2 \tag{94}$$
$$\leq \mathbb{E}\left\|\left\langle h_k^t(\hat{\theta}_k^t), (\mathbf{1} - \mathbf{l}_k^t)\right\rangle\right\|^2 \leq \mathbb{E}\left\|h_k^t(\hat{\theta}_k^t)\right\|^2 \mathbb{E}\left\|(\mathbf{1} - \mathbf{l}_k^t)\right\|^2 \tag{95}$$
$$\leq \delta^2(\beta_k^t)^2 \leq \delta^2\beta_k^t \tag{96}$$

The last term is derived under the Assumption 4 and β_k^t are valued within the $(0, 1)$. Then we will show the bound of Ψ_k^t as:

$$\Psi_k^t = \mathbb{E}\|h_k^t(\theta_k^t) - h_k^t(\hat{\theta}_k^t)\|^2 \tag{97}$$
$$\leq M_k^2\mathbb{E}\|\theta_k^t - \hat{\theta}_k^t\|^2 \leq M_k^2\mathbb{E}\|\theta_k^t \odot \mathbf{1} - \theta_k^t \odot \mathbf{m}_k^t\|^2 \tag{98}$$
$$\leq M_k^2\mathbb{E}\|\theta_k^t \odot (\mathbf{1} - \mathbf{m}_k^t)\|^2 \leq M_k^2\mathbb{E}\left\|\langle\theta_k^t, (\mathbf{1} - \mathbf{m}_k^t)\rangle\right\|^2 \tag{99}$$
$$\leq M_k^2\mathbb{E}\left\|\theta_k^t\right\|^2 \mathbb{E}\left\|(\mathbf{1} - \mathbf{m}_k^t)\right\|^2 \tag{100}$$
$$\leq M_k^2\mu^2(\alpha_k^t)^2 \leq M_k^2\mu^2\alpha_k^t \tag{101}$$

Where Eq. 98 is due to Assumption 5. And the last term is also derived under the Assumption 4 and α_k^t are valued within the $(0, 1)$. This finishes the proof of Corollary 1.