Supplementary Material for “A Decentralized Federated Learning Framework via Committee Mechanism with Convergence Guarantee”

Chunjiang Che, Xiaoli Li, *Student Member, IEEE,*
Chuan Chen, *Member, IEEE,* Xiaoyu He, and Zibin Zheng, *Senior Member, IEEE*

Abstract

This supplement provides technical proofs of the theoretical results in the main manuscripts.

Index Terms

Decentralized Federated Learning, Committee Mechanism, Byzantine Robustness, Theoretical Convergence Analysis.

A Proof of the Theorem 1

A.1 Definition and Lemma

In our analysis, the (t) is defined as the index of the total local SGD iterations, where $(i) = t\tau + i$. We iterpret $\mathbf{w}_k^{(i)}$ as the local model of k-th client at iteration (t). In such a setting, the aggregation client set and the commitee client set are expressed as $S_a^{(t)}$ and $S_c^{(t)}$. Note that the aggregation clients and committee clients perform τ iterations of local SGD, the aggregation client set $S_a^{(t)}$ and the committee client set $S_c^{(t)}$ remain the same for every τ iterations. That is, $S_a^{(t)} = S_a^{(t+1)} = ... = S_a^{(t+\tau-1)}$ and $S_c^{(t)} = S_c^{(t+1)} = ... = S_c^{(t+\tau-1)}$, where $t\%\tau = 0$. The local gradient of k-th client is expressed as $g_k(\mathbf{w}_k^{(t)}, \mathcal{B}_k^{(t)})$, which is used for the local SGD training:

$$\mathbf{w}_k^{(t+1)} = \mathbf{w}_k^{(t)} - \eta_t g_k(\mathbf{w}_k^{(t)}, \mathcal{B}_k^{(t)}), \tag{1}$$

where $\mathbf{w}_k^{(t)}$ denotes the local model of the k-th client at iteration t. Actually we only update the global model $\overline{\mathbf{w}}^{(t)}$ after every τ rounds, but for the need of convergence proof and analysis we assume the $\overline{\mathbf{w}}^{(t)}$ is updated at each iteration as follows:

$$\overline{\mathbf{w}}^{(t+1)} = \overline{\mathbf{w}}^{(t)} - \eta_t \overline{g}^{(t)} = \overline{\mathbf{w}}^{(t)} - \eta_t (\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} g_k(\mathbf{w}_k^{(t)}, \mathcal{B}_k^{(t)})), \tag{2}$$

where η_t is the learning rate at iteration (t). At iteration (t) which satisfies $t\%\tau = 0$, the k-th client downloads the $\overline{\mathbf{w}}^{(t)}$ to the local as the new local model $\mathbf{w}_k^{(t)}$, as a result, the whole updating method of the local model can be written as follows:

$$\mathbf{w}_k^{(t+1)} = \begin{cases} \mathbf{w}_k^{(t)} - \eta_t g_k(\mathbf{w}_k^{(t)}, \mathcal{B}_k^{(t)}), (t+1)\%\tau \neq 0 \\ \overline{\mathbf{w}}^{(t+1)}, (t+1)\%\tau = 0. \end{cases} \tag{3}$$

Then we introduce our proposed lemmas.

Lemma 1. ***(Global-Local Gradient Product).*** *According to the definitions of $\nabla F_k(\mathbf{w}_k^{(t)})$, we have that*

$$\begin{aligned} &-\langle \overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ \leq &\frac{1}{2\eta_t} \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \frac{\eta_t}{2} \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\nabla F_k(\mathbf{w}_k^{(t)})||^2 - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k(\mathbf{w}^*)) \\ &- \frac{\mu}{2} \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2. \end{aligned} \tag{4}$$

Chunjiang Che, Xiaoli Li, Chuan Chen, Xiaoyu He and Zibin Zheng are with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. e-mail: {chechj, lixli27}@mail2.sysu.edu.cn, {chenchuan, hexy73, zhzibin}@mail.sysu.edu.cn. Corresponding author: Chuan Chen.

Lemma 2. ***(Local Parameter-Optimal Gap).*** *Let $v_t = 2\eta_t - 4L\eta_t^2 = 2\eta_t(1-2L\eta_t)$, we have that*

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)}) - F_k^*) \\ \leq& -(1-\eta_t L)\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)}) - F^*) + \frac{1}{v_t}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)} - \overline{\mathbf{w}}^{(t)}||^2. \end{aligned} \tag{5}$$

Lemma 3. ***(Client Heterogeneous Bound).*** *According to the Assumption 4, we have that*

$$\mathbb{E}[\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2] \leq 16\eta_t^2\tau^2G^2. \tag{6}$$

A.2 Proof of Lemma

In this section we will prove Lemma 1 to 3.

A.2.1 Proof of Lemma 1

Proof. We introduce $\mathbf{w}_k^{(t)}$ into the formula:

$$\begin{aligned} &-\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ =& -\sum_{k\in S_a^{(t)}}\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, p_{k,S_a^{(t)}}\nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ =& -\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)} + \mathbf{w}_k^{(t)} - \mathbf{w}^*, \nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ =& -\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}, \nabla F_k(\mathbf{w}_k^{(t)})\rangle - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\langle\mathbf{w}_k^{(t)} - \mathbf{w}^*, \nabla F_k(\mathbf{w}_k^{(t)})\rangle. \end{aligned} \tag{7}$$

According to the Cauchy-Schwarz inequality and AM-GM inequality, we have

$$\begin{aligned} &-\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ \leq&\frac{1}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(\frac{1}{\eta_t}||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \eta_t||\nabla F_k(\mathbf{w}_k^{(t)})||^2) - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\langle\mathbf{w}_k^{(t)} - \mathbf{w}^*, \nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ \leq&\frac{1}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(\frac{1}{\eta_t}||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \eta_t||\nabla F_k(\mathbf{w}_k^{(t)})||^2) - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(\mathbf{w}_k^{(t)} - \mathbf{w}^*)^T\nabla F_k(\mathbf{w}_k^{(t)}). \end{aligned} \tag{8}$$

Due to the Assumption 2, we expand the above formula and complete the proof:

$$\begin{aligned} &-\langle\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}\nabla F_k(\mathbf{w}_k^{(t)})\rangle \\ \leq&\frac{1}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(\frac{1}{\eta_t}||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \eta_t||\nabla F_k(\mathbf{w}_k^{(t)})||^2) - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)} - F_k(\mathbf{w}^*)) - \frac{\mu}{2}||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2) \\ =&\frac{1}{2\eta_t}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \frac{\eta_t}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\nabla F_k(\mathbf{w}_k^{(t)})||^2 - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)}) - F_k(\mathbf{w}^*)) \\ &-\frac{\mu}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2. \end{aligned} \tag{9}$$

□

A.2.2 Proof of Lemma 2

Proof. The original formula can be written as:

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)}) - F_k^*) \\ =& -\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)}) - F_k(\overline{\mathbf{w}}^{(t)}) + F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*) \\ \leq& -\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)}) - F_k(\overline{\mathbf{w}}^{(t)})) - \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*). \end{aligned} \tag{10}$$

Due to the assumption 2 we have

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)})-F_k^*) \\ &\leq \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}[\frac{\eta_t}{2}||\nabla F_k(\overline{\mathbf{w}}^{(t)})||^2+\frac{1}{2\eta_t}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2-\frac{\mu}{2}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2-(F_k(\overline{\mathbf{w}}^{(t)})-F_k^*)]. \end{aligned} \tag{11}$$

We continue to expand the above formula as follows:

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)})-F_k^*) \\ =&\frac{\eta_t}{2}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\nabla F_k(\overline{\mathbf{w}}^{(t)})||^2-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)})-F_k^*)+(\frac{1}{2\eta_t}-\frac{\mu}{2})\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2. \end{aligned} \tag{12}$$

Due to the Assumption 1, we have

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)})-F_k^*) \\ \leq&\eta_t L\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)})-F_k^*)-\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)})-F_k^*)+\frac{1-\eta_t\mu}{2\eta_t}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2 \\ =&-(1-\eta_t L)\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)})-F^*)+\frac{1-\eta_t\mu}{2\eta_t}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2. \end{aligned} \tag{13}$$

As $\frac{1-\eta_t L}{2\eta_t}\leq\frac{1}{v_t}$, we continue to expand the formula and complete the proof:

$$\begin{aligned} &-\sum_{k\in S_a^{(t)}}(F_k(\mathbf{w}_k^{(t)})-F_k^*) \\ \leq&-(1-\eta_t L)\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}(F_k(\overline{\mathbf{w}}^{(t)})-F^*)+\frac{1}{v_t}\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\mathbf{w}_k^{(t)}-\overline{\mathbf{w}}^{(t)}||^2. \end{aligned} \tag{14}$$

□

A.2.3 Proof of Lemma 3

Proof. According to the update rule, for k and k' which are in the same set $S_a^{(t)}$, the term $||\mathbf{w}_{k'}^{(t)}-\mathbf{w}_k^{(t)}||^2$ will be zero when $k=k'$. As a result we have

$$\begin{aligned} &\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\overline{\mathbf{w}}^{(t)}-\mathbf{w}_k^{(t)}||^2 \\ =&\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\sum_{k'\in S_a^{(t)}} p_{k',S_a^{(t)}}\mathbf{w}_{k'}^{(t)}-\mathbf{w}_k^{(t)}||^2 \\ =&\sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}}||\sum_{k'\in S_a^{(t)}}(p_{k',S_a^{(t)}}\mathbf{w}_{k'}^{(t)}-p_{k',S_a^{(t)}}\mathbf{w}_k^{(t)})||^2 \\ =&\sum_{\substack{k\neq k' \\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}}p_{k',S_a^{(t)}}||\mathbf{w}_{k'}^{(t)}-\mathbf{w}_k^{(t)}||^2. \end{aligned} \tag{15}$$

Since the selected local models are udpated with the global model at every τ, for any t there is a t_0 satisfies that $0\leq t-t_0\leq\tau$ and $\mathbf{w}_{k'}^{(t_0)}=\mathbf{w}_k^{(t_0)}=\overline{\mathbf{w}}^{(t)}$. Therefore for any (t) the term $||\mathbf{w}_{k'}^{(t)}-\mathbf{w}_k^{(t)}||^2$ can be bound by τ epochs. With non-increasing η_t

over (t) and $\eta_{t_0} \leq 2\eta_t$, Eq. 15 can be further bound as

$$\begin{aligned}
&\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} ||\mathbf{w}_{k'}^{(t)} - \mathbf{w}_k^{(t)}||^2 \\
\leq &\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} ||\sum_{i=t_0}^{t_0+\tau-1} \eta_i (g_{k'}(\mathbf{w}_{k'}^{(i)}, B_{k'}^{(i)}) - g_k^{(i)}(\mathbf{w}_k^{(t)}, B_k^{(i)}))||^2 \\
\leq &\eta_{t_0}^2 \tau \sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} \sum_{i=t_0}^{t_0+\tau-1} ||g_{k'}(\mathbf{w}_{k'}^{(i)}, B_{k'}^{(i)}) - g_k^{(i)}(\mathbf{w}_k^{(t)}, B_k^{(i)})||^2 \\
\leq &\eta_{t_0}^2 \tau \sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} \sum_{i=t_0}^{t_0+\tau-1} [2||g_{k'}(\mathbf{w}_{k'}^{i}, B_{k'}^{(i)})||^2 + 2||g_k(\mathbf{w}_k^{(i)}, B_k^{(i)})||^2].
\end{aligned} \tag{16}$$

According to the Assumption 4, the expectation over Eq.16 can be written as

$$\begin{aligned}
&\mathbb{E}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} ||\mathbf{w}_{k'}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
=&2\eta_0^2 \tau \mathbb{E}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} \sum_{i=t_0}^{t_0+\tau-1} (||g_{k'}(\mathbf{w}_{k'}^{i}, B_{k'}^{(i)})||^2 + ||g_k(\mathbf{w}_k^{(i)}, B_k^{(i)})||^2)] \\
\leq &2\eta_0^2 \tau \mathbb{E}_{S_a^{(t)}}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} \sum_{i=t_0}^{t_0+\tau-1} 2G^2] \\
=&2\eta_0^2 \tau \mathbb{E}_{S_a^{(t)}}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} 2p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} \tau G^2] \\
\leq &2\eta_0^2 \tau \mathbb{E}_{S_a^{(t)}}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} 2\tau G^2].
\end{aligned} \tag{17}$$

Since there are at most $m(m-1)$ pairs such that $k \neq k'$ in $S_a^{(t)}$, we have

$$\begin{aligned}
&\mathbb{E}[\sum_{\substack{k\neq k'\\ k,k'\in S_a^{(t)}}} p_{k,S_a^{(t)}} p_{k',S_a^{(t)}} ||\mathbf{w}_{k'}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
\leq &\frac{16\eta_t^2 (m-1)\tau^2 G^2}{m} \\
\leq &16\eta_t^2 \tau^2 G^2.
\end{aligned} \tag{18}$$

□

A.3 Proof of Theorem 1

Proof. According to Eq. (2) we define $\mathcal{H}(\mathbf{w}, t+1)$ as

$$\mathcal{H}(\mathbf{w}, t+1) = ||\overline{\mathbf{w}}^{(t+1)} - \mathbf{w}^*||^2 = ||\overline{\mathbf{w}}^{(t)} - \eta_t \overline{g}^{(t)} - \mathbf{w}^*||^2. \tag{19}$$

The $\mathcal{H}(\mathbf{w}, t+1)$ can be written as:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
=&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^* - \eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) + \eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2.
\end{aligned} \tag{20}$$

We expand the above formula as follows:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
=&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^* - \eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)})||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 \\
&\underbrace{+ 2\eta_t \langle \overline{\mathbf{w}}^{(t)} - \mathbf{w}^* - \eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}), \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \overline{g}^{(t)} \rangle}_{A_1}.
\end{aligned} \tag{21}$$

Due to the Assumption 3, we have $\mathbb{E}[A_1] = 0$ and expand the rest of the formula further:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
=&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^* - \eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)})||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1 \\
=&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)})||^2 - 2\eta_t \langle \overline{\mathbf{w}}^{(t)} - \mathbf{w}^*, \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) \rangle \\
&+ ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1.
\end{aligned} \tag{22}$$

Due to the Lemma 1, we have that:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
\leq&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)})||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 \\
&+ \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 + \eta_t^2 p_{k,S_a^{(t)}} ||\nabla F_k(\mathbf{w}_k^{(t)})||^2 - 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k(\mathbf{w}^*)) \\
&- \eta_t \mu \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2 + A_1 \\
=&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 + 2\eta_t^2 \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\nabla F_k(\mathbf{w}_k^{(t)})||^2 + \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 - 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k(\mathbf{w}^*)) \\
&- \eta_t \mu \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1 \\
\leq&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 + \underbrace{4L\eta_t^2 \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k^*)}_{A_2} - \underbrace{2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k(\mathbf{w}^*))}_{A_3} \\
&+ \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 - \eta_t \mu \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1.
\end{aligned} \tag{23}$$

The difference of A_2 and A_3 can be written as

$$\begin{aligned}
&A_2 - A_3 \\
=&(4L\eta_t^2 - 2\eta_t) \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}_k^{(t)}) - 4L\eta_t^2 \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^* + 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}^*) \\
=&(4L\eta_t^2 - 2\eta_t) \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}_k^{(t)}) - (4L\eta_t^2 - 2\eta_t) \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^* + 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}^*) - 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^* \\
=&(4L\eta_t^2 - 2\eta_t) \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k^*) + 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*).
\end{aligned} \tag{24}$$

Let $v_t = 2\eta_t - 4L\eta_t^2 = 2\eta_t(1 - 2L\eta_t)$, we have that:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
\leq&||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 - v_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}_k^{(t)}) - F_k^*) + 2\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*) + \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 \\
&- \eta_t \mu \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2 + ||\eta_t \sum_{k\in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1.
\end{aligned} \tag{25}$$

Due to the Lemma 2, we have that:

$$\begin{aligned}
&\mathcal{H}(\mathbf{w}, t+1) \\
\leq &||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2 + 2\eta_t \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*) + 2 \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 \\
&- v_t(1-\eta_t L) \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*) - \eta_t \mu \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2 \\
&+ ||\eta_t \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2 + A_1.
\end{aligned} \tag{26}$$

We next solve the expectation over $\mathcal{H}(\mathbf{w}, t+1)$:

$$\begin{aligned}
&\mathbb{E}[\mathcal{H}(\mathbf{w}, t+1)] = \mathbb{E}[||\overline{\mathbf{w}}^{(t+1)} - \mathbf{w}^*||^2] \\
\leq &\mathbb{E}[||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2] + \mathbb{E}[2\eta_t \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*)] + \mathbb{E}[2 \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
&- \mathbb{E}[v_t(1-\eta_t L) \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*)] - \mathbb{E}[\eta_t \mu \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2] \\
&+ \mathbb{E}[||\eta_t \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} \nabla F_k(\mathbf{w}_k^{(t)}) - \eta_t \overline{g}^{(t)}||^2] + \mathbb{E}[A_1].
\end{aligned} \tag{27}$$

Due to Assumption 3 and $\mathbb{E}[A_1] = 0$, we have

$$\begin{aligned}
&\mathbb{E}[||\overline{\mathbf{w}}^{(t+1)} - \mathbf{w}^*||^2] \\
\leq &\mathbb{E}[||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2] + 2\eta_t \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*)] + 2\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
&- v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*)] - \eta_t \mu \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\mathbf{w}_k^{(t)} - \mathbf{w}^*||^2] + \eta_t^2 \sum_{k=1}^{K} p_k^2 \sigma_k^2. \\
= &(1-\eta_t \mu)\mathbb{E}[||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2] + 2\eta_t \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*)] + 2\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
&- v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*)] + \eta_t^2 \sum_{k=1}^{K} p_k^2 \sigma_k^2 \\
= &(1-\eta_t \mu)\mathbb{E}[||\overline{\mathbf{w}}^{(t)} - \mathbf{w}^*||^2] + \mathbb{E}[\mathcal{Q}(\mathbf{w}, k, t)] + \eta_t^2 \sum_{k=1}^{K} p_k^2 \sigma_k^2,
\end{aligned} \tag{28}$$

where $\mathcal{Q}(\mathbf{w}, k, t)$ are defined as follows:

$$\begin{aligned}
&\mathcal{Q}(\mathbf{w}, k, t) \\
= &2\eta_t \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*) + 2 \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2 - v_t(1-\eta_t L) \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*).
\end{aligned} \tag{29}$$

Note that $S_c^* = \arg\min_{S_c} \sum_{k \in S_c} p_{k,S_c} F_k^*$. Due to the Lemma 3, the expectation of the $\mathcal{Q}(\mathbf{w}, k, t)$ can be written as:

$$
\begin{aligned}
&\mathbb{E}[\mathcal{Q}(\mathbf{w}, k, t)] \\
=& - v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*)] + 2\eta_t \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*] + 2\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} ||\overline{\mathbf{w}}^{(t)} - \mathbf{w}_k^{(t)}||^2] \\
=& - v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - F_k^*)] + 2\eta_t \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\mathbf{w}^*) - F_k^*)] + 32\eta_t^2\tau^2 G^2 \\
=& - v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} (F_k(\overline{\mathbf{w}}^{(t)}) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* + \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* - \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^*] \\
&+ 2\eta_t \mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}^*) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* + \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* - \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^*] + 32\eta_t^2\tau^2 G^2 \\
=& - v_t(1-\eta_t L)(\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\overline{\mathbf{w}}^{(t)}) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*] + \mathbb{E}[\sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* - \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^*]) \\
&+ 2\eta_t (\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}^*) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*] + \mathbb{E}[\sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^* - \sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^*]) + 32\eta_t^2\tau^2 G^2 \\
=& - v_t(1-\eta_t L)\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\overline{\mathbf{w}}^{(t)}) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*] + 2\eta_t (\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k(\mathbf{w}^*) - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*] \\
&- (2\eta_t - v_t(1-\eta_t L))\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^* - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*] + 32\eta_t^2\tau^2 G^2.
\end{aligned}
\tag{30}
$$

According to the Assumption 5 and Definition 1 and 2, we have

$$
\begin{aligned}
&\mathbb{E}[\mathcal{Q}(\mathbf{w}, k, t)] \\
\leq& - v_t(1-\eta_t L)\mathbb{E}[\varphi(S_a^{(t)}, \overline{\mathbf{w}})(F(\overline{\mathbf{w}}) - \sum_{k=1}^K p_k F_k^*)] + 2\eta_t \mathbb{E}[\varphi(S_a^{(t)}, \mathbf{w}^*)(F * - \sum_{k=1}^K p_k F_k^*)] \\
&+ (2\eta_t - v_t(1-\eta_t L))||\mathbb{E}[\sum_{k \in S_a^{(t)}} p_{k,S_a^{(t)}} F_k^* - \sum_{k' \in S_c^*} p_{k',S_c^*} F_{k'}^*]|| + 32\eta_t^2\tau^2 G^2 \\
\leq& - v_t(1-\eta_t L)\mathbb{E}[\varphi(S_a^{(t)}, \overline{\mathbf{w}})(F(\overline{\mathbf{w}}) - \sum_{k=1}^K p_k F_k^*)] + 2\eta_t \mathbb{E}[\varphi(S_a^{(t)}, \mathbf{w}^*)(F^* - \sum_{k=1}^K p_k F_k^*)] \\
&+ (2\eta_t - v_t(1-\eta_t L))\kappa^2 + 32\eta_t^2\tau^2 G^2 \\
\leq& - v_t(1-\eta_t L)\varphi_{min}\mathbb{E}[(F(\overline{\mathbf{w}}) - \sum_{k=1}^K p_k F_k^*)] + 2\eta_t \varphi_{max}\mathbb{E}[(F(\mathbf{w}^*) - \sum_{k=1}^K p_k F_k^*)] + (2\eta_t - v_t(1-\eta_t L))\kappa^2 \\
&+ 32\eta_t^2\tau^2 G^2 \\
\leq& \underbrace{-v_t(1-\eta_t L)\varphi_{min}\mathbb{E}[(F(\overline{\mathbf{w}}) - \sum_{k=1}^K p_k F_k^*)]}_{A_4} + 2\eta_t \varphi_{max}\Gamma + 6L\eta_t^2\kappa^2 + 32\eta_t^2\tau^2 G^2.
\end{aligned}
\tag{31}
$$

We can expand the A_4 as

$$
\begin{aligned}
&A_4 \\
=& - v_t(1-\eta_t L)\varphi_{min}\mathbb{E}[(F(\overline{\mathbf{w}}) - \sum_{k=1}^K p_k F_k^*)] \\
=& - v_t(1-\eta_t L)\varphi_{min}\sum_{k=1}^K p_k(\mathbb{E}[F(\overline{\mathbf{w}})] - F^* + F^* - F_k^*) \\
=& - v_t(1-\eta_t L)\varphi_{min}\sum_{k=1}^K p_k(\mathbb{E}[F_k(\overline{\mathbf{w}}^{(t)})] - F^*) - v_t(1-\eta_t L)\varphi_{min}\sum_{k=1}^K p_k(F^* - F_k^*) \\
=& - v_t(1-\eta_t L)\varphi_{min}(\mathbb{E}[F(\overline{\mathbf{w}}^{(t)})] - F^*) - v_t(1-\eta_t L)\varphi_{min}\Gamma
\end{aligned}
\tag{32}
$$

$$\begin{aligned}
&\leq -\frac{v_t(1-\eta_t L)\mu\varphi_{min}}{2}\mathbb{E}[||\overline{\mathbf{w}}^{(t)}-\mathbf{w}^*||^2] - v_t(1-\eta_t L)\varphi_{min}\Gamma \\
&\leq -\frac{3\eta_t\mu\varphi_{min}}{8}\mathbb{E}[||\overline{\mathbf{w}}^{(t)}-\mathbf{w}^*||^2] - 2\eta_t(1-2L\eta_t)(1-\eta_t L)\varphi_{min}\Gamma \\
&\leq -\frac{3\eta_t\mu\varphi_{min}}{8}\mathbb{E}[||\overline{\mathbf{w}}^{(t)}-\mathbf{w}^*||^2] - 2\eta_t\varphi_{min}\Gamma + 6\eta_t^2\varphi_{min}L\Gamma.
\end{aligned} \tag{33}$$

So we have

$$\begin{aligned}
&\mathbb{E}[\mathcal{Q}(\mathbf{w},k,t)] \\
=&-\frac{3\eta_t\mu\varphi_{min}}{8}\mathbb{E}[||\overline{\mathbf{w}}^{(t)}-\mathbf{w}^*||^2] + 2\eta_t\Gamma(\varphi_{max}-\varphi_{min}) + \eta_t^2(6\varphi_{min}L\Gamma + 32\tau^2G^2 + 6L\kappa^2).
\end{aligned} \tag{34}$$

As a result, we have

$$\begin{aligned}
&\mathbb{E}[||\overline{\mathbf{w}}^{(t+1)}-\mathbf{w}^*||] \\
\leq&[1-\eta_t\mu(1+\frac{3\varphi_{min}}{8})]\mathbb{E}[||\overline{\mathbf{w}}^{(t)}-\mathbf{w}^*||^2] + 2\eta_t\Gamma(\varphi_{max}-\varphi_{min}) + \eta_t^2(6\varphi_{min}L\Gamma + 32\tau^2G^2 + 6L\kappa^2 + \sum_{k=1}^{K}p_k\sigma_k^2).
\end{aligned} \tag{35}$$

By defining $\Delta_{t+1}=\mathbb{E}[||\overline{\mathbf{w}}^{(t+1)}-\mathbf{w}^*||]$, $B=1+\frac{3\varphi_{min}}{8}$, $C=6\varphi_{min}L\Gamma+32\tau^2G^2+6L\kappa^2+\sum_{k=1}^{K}p_k\sigma_k^2$, $D=\Gamma(\varphi_{max}-\varphi_{min})$, we have

$$\Delta_{t+1}\leq(1-\eta_t\mu B)\Delta_t+\eta_t^2C+\eta_t D. \tag{36}$$

If we set $\Delta_t\leq\frac{\psi}{t+\gamma}$, $\eta_t=\frac{\beta}{t+\gamma}$ and $\beta>\frac{1}{\mu B}$, $\gamma>0$ by induction, we have

$$\psi=\max\left\{\gamma||\overline{\mathbf{w}}^1-\mathbf{w}^*||^2,\frac{1}{\beta\mu B-1}(\beta^2C+D\beta(t+\gamma))\right\}. \tag{37}$$

Then by the L-smoothness of $F(\cdot)$,

$$\mathbb{E}[F(\overline{\mathbf{w}}^{(t)})]-F^*\leq\frac{L}{2}\Delta_t\leq\frac{L}{2}\frac{\psi}{\gamma+t}. \tag{38}$$

Finally, we complete the proof of Theorem 1:

$$\begin{aligned}
&\mathbb{E}[F(\overline{\mathbf{w}}^T)]-F^* \\
\leq&\frac{1}{T+\gamma}\left[\frac{4L(32\tau^2G^2+\sum_{k=1}^{K}p_k\sigma_k^2)+24L^2\kappa^2}{3\mu^2\varphi_{min}}+\frac{8L^2\Gamma}{\mu^2}+\frac{L\gamma||\overline{\mathbf{w}}^1-\mathbf{w}^*||^2}{2}\right]+\frac{8L\Gamma}{3\mu}\left(\frac{\varphi_{max}}{\varphi_{min}}-1\right),
\end{aligned} \tag{39}$$

where the T means the maximal communication rounds, which satisfies $T=i\tau$ for $i=1,2,...$ in realistic scenarios. □