APPENDIX

A. Proof of Theorem 1

For the sake of convenience, we first establish some notations:

Let $\mathbf{w}_t^i$ represent the model of the i-th client at the t-th iteration. Let $\mathcal{I}_\tau$ denote the global aggregation step, which is defined as $\mathcal{I}_\tau = \{n\tau | n = 1, 2, \dots\}$. Based on Assumption 3, the gradient descent equation can be expressed as:

$$\mathbf{v}_{t+1}^i = \mathbf{w}_t^i - \frac{\eta}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i} \nabla F_i(\mathbf{w}_t^i, s) + \sigma C \mathcal{N}(0, \mathbf{I}_d)\right) \tag{17}$$

$$\mathbf{w}_{t+1}^i = \begin{cases} \mathbf{v}_{t+1}^i & \text{if } t+1 \notin \mathcal{I}_\tau, \\ \sum_{i=1}^N p_i \mathbf{v}_{t+1}^i & \text{if } t+1 \in \mathcal{I}_\tau. \end{cases} \tag{18}$$

Here, we introduce $\mathbf{v}_{t+1}^i$ to represent the intermediate result of $\mathbf{w}_t^i$ after one iteration of gradient descent.

Inspired by [39], we introduce two virtual sequences in our analysis: $\bar{\mathbf{v}}_t = \sum_{i=1}^N p_i \mathbf{v}_t^i$ and $\bar{\mathbf{w}}_t = \sum_{i=1}^N p_i \mathbf{v}_t^i$. When $t+1 \notin \mathcal{I}_\tau$, $\bar{\mathbf{v}}_t = \bar{\mathbf{w}}_t$, but the server cannot access these two sequences. When $t+1 \in \mathcal{I}_\tau$, the server can access $\bar{\mathbf{w}}_t$.

Therefore, we have:

$$\bar{\mathbf{v}}_{t+1} = \bar{\mathbf{w}}_t - \eta \sum_{i=1}^N \frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i} \nabla F_i(\mathbf{w}_t^i, s) + \sigma C \mathcal{N}(0, \mathbf{I}_d)\right) \tag{19}$$

1) Key lemma:

Lemma 4: (Results of one iteration). Assume Assumption 1-3, we have:

$$\mathbb{E}\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2 \leq (1-\eta\mu)\,\mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + \eta^2 A$$

since $\Delta_t = \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2$, we have:

$$\Delta_{t+1} \leq (1-\eta\mu)\,\Delta_t + \eta^2 A$$

where $A = (\tau-1)^2(2+\eta\mu)\left(C^2 + \frac{\sigma^2 C^2 d}{\hat{B}}\right) + 3C^2 + \frac{2\Gamma}{\eta} + \frac{\sigma^2 C^2 d}{\hat{B}^2}$, $\Gamma = F^* - \sum_{i=1}^N p_i F_i^*$.

We use mathematical induction to obtain $\Delta_t \leq \frac{v}{t}$ where $v = \max\{\frac{\beta^2+A}{\beta\mu-1}, \tau\Delta_1\}$.

STEP 1. When $t = 1$, the equation $\Delta_1 \leq v$ holds obviously.

STEP 2. We assume $\Delta_t \leq \frac{v}{t}$ holds.

STEP 3. on the condition that $\eta = \frac{\beta}{t}$, $\beta > \frac{1}{\mu}$, we know

$$\begin{aligned}\Delta_{t+1} &\leq (1-\eta\mu)\,\Delta_t + \eta^2 A \\ &\leq \left(1 - \frac{\beta\mu}{t}\right)\frac{v}{t} + \frac{\beta^2 + A}{t^2} \\ &= \frac{(t-1)v}{t^2} + \left(\frac{\beta^2+A}{t^2} - \frac{\beta\mu - 1}{t^2}v\right) \\ &\leq \frac{t-1}{t^2}v \leq \frac{v}{t+1}\end{aligned}$$

Therefore, $\Delta_{t+1} \leq \frac{v}{t+1}$ holds, completing the proof by mathematical induction. Hence, $\Delta_t \leq \frac{v}{t}$ holds.

Then by the L-smoothness of $F(\cdot)$, and set $\beta = \frac{2}{\mu}$, $t \leftarrow T$

$$\begin{aligned}&\mathbb{E}[F(\bar{\mathbf{w}}_T)] - F^* \leq \frac{L}{2}\Delta_T \\ &\leq \frac{L}{2T}\left(\frac{4}{\mu^2} + A + \tau\Delta_1\right) \\ &\leq \frac{L}{2T}(\frac{4}{\mu^2} + \tau\Delta_1 + (\tau-1)^2(2+\eta\mu)\left(C^2 + \frac{\sigma^2C^2d}{\hat{B}^2}\right) \\ &\quad + 3C^2 + \frac{2\Gamma}{\eta} + \frac{\sigma^2C^2d}{\hat{B}^2}) \\ &= \frac{L(2+\eta\mu)\left(C^2 + \frac{\sigma^2C^2d}{\hat{B}^2}\right)}{2T}(\tau-1)^2 \\ &\quad + \frac{L}{2T}\left(\frac{4}{\mu^2} + \tau\Delta_1 + 3C^2 + \frac{2\Gamma}{\eta} + \frac{\sigma^2C^2d}{\hat{B}^2}\right) \\ &= h(\tau)\end{aligned} \tag{20}$$

where

$$\begin{aligned}h(\tau) \triangleq{}& \frac{L(2+\eta\mu)\left(C^2 + \frac{\sigma^2C^2d}{\hat{B}^2}\right)}{2T}\tau^2 \\ &+ \frac{L\Delta_1 - 2L(2+\eta\mu)\left(C^2 + \frac{\sigma^2C^2d}{\hat{B}^2}\right)}{2T}\tau + \\ &\frac{L\left((2+\eta\mu)\left(C^2 + \frac{\sigma^2C^2d}{\hat{B}^2}\right) + \frac{4}{\mu^2} + 3C^2 + \frac{2\Gamma}{\eta} + \frac{\sigma^2C^2d}{\hat{B}^2}\right)}{2T}\end{aligned} \tag{21}$$

2) Proof of Lemma 4:

According to (17),

$$
\begin{aligned}
&\|\bar{\mathbf{v}}_{t+1}-\mathbf{w}^*\|^2\\
&=\|\bar{\mathbf{w}}_t-\mathbf{w}^*\\
&\quad-\eta\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\|^2\\
&=\|\bar{\mathbf{w}}_t-\mathbf{w}^*\|^2\\
&+2\eta\underbrace{\langle\mathbf{w}^*-\bar{\mathbf{w}}_t,\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\rangle}_{\mathcal{A}_1}\\
&+\eta^2\underbrace{\|\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\|^2}_{\mathcal{A}_2}
\end{aligned}
\tag{22}
$$

By the convexity of $\|\cdot\|^2$,

$$
\begin{aligned}
\mathcal{A}_2&=\left\|\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\right\|^2\\
&\le\sum_{i=1}^{N}p_i\left\|\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\right\|^2\\
&\le\sum_{i=1}^{N}p_i\left\|\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\right\|^2\\
&\quad+\sum_{i=1}^{N}p_i\left\|\frac{1}{|\mathcal{B}_t^i|}\sigma C\mathcal{N}(0,\mathbf{I}_d)\right\|^2
\end{aligned}
$$

(From the Cauchy Schwartz inequality.)

Take the expectation of $\mathcal{A}_2$:

$$
\mathbb{E}\mathcal{A}_2\le C^2+\frac{\sigma^2C^2d}{\hat{B}^2}
$$

where $\frac{1}{\hat{B}}=\max_{i,t}\mathbb{E}\frac{1}{|\mathcal{B}_t^i|}$.

$$
\begin{aligned}
\mathcal{A}_1&=-\langle\bar{\mathbf{w}}_t-\mathbf{w}^*,\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)+\sigma C\mathcal{N}(0,\mathbf{I}_d)\right)\rangle\\
&=\underbrace{-\langle\bar{\mathbf{w}}_t-\mathbf{w}^*,\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\rangle}_{\mathcal{B}_1}\\
&\quad\underbrace{-\langle\bar{\mathbf{w}}_t-\mathbf{w}^*,\sum_{i=1}^{N}p_i\frac{1}{|\mathcal{B}_t^i|}\sigma C\mathcal{N}(0,\mathbf{I}_d)\rangle}_{\mathcal{B}_2}
\end{aligned}
$$

Because the noise has a mean of 0:

$$
\mathbb{E}\mathcal{B}_2=0 \tag{23}
$$

Add a zero term to $\mathcal{B}_1$:

$$
\begin{aligned}
\mathcal{B}_1&=-\sum_{i=1}^{N}p_i\langle\bar{\mathbf{w}}_t-\mathbf{w}_t^i+\mathbf{w}_t^i-\mathbf{w}^*,\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\rangle\\
&=\underbrace{-\sum_{i=1}^{N}p_i\langle\bar{\mathbf{w}}_t-\mathbf{w}_t^i,\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\rangle}_{\mathcal{C}_1}\\
&\quad\underbrace{-\sum_{i=1}^{N}p_i\langle\mathbf{w}_t^i-\mathbf{w}^*,\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\rangle}_{\mathcal{C}_2}
\end{aligned}
$$

By Cauchy-Schwarz inequality and AM-GM inequality, we have

$$
\begin{aligned}
\mathcal{C}_1&\le\frac{1}{\eta}\sum_{i=1}^{N}p_i\|\bar{\mathbf{w}}_t-\mathbf{w}_t^i\|^2+\eta\sum_{i=1}^{N}p_i\left\|\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i}\nabla F_i(\mathbf{w}_t^i,s)\right\|^2\\
&\le\frac{1}{\eta}\sum_{i=1}^{N}p_i\|\bar{\mathbf{w}}_t-\mathbf{w}_t^i\|^2+\eta C^2
\end{aligned}
$$

According to Assumption 2, we get

$$
\begin{aligned}
\mathcal{C}_2 &\le -\sum_{i=1}^{N} p_i \langle \mathbf{w}_t^i - \mathbf{w}^*, \frac{1}{|\mathcal{B}_t^i|} \sum_{s\in\mathcal{B}_t^i} \nabla F_i(\mathbf{w}_t^i, s)\rangle \\
&\le -\sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F_i(\mathbf{w}^*)\right) - \sum_{i=1}^{N} p_i \frac{\mu}{2} \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 \\
&= -\sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F(\mathbf{w}^*) + F(\mathbf{w}^*) - F_i(\mathbf{w}^*)\right) \\
&\quad - \frac{\mu}{2}\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 \\
&= -\sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F(\mathbf{w}^*)\right) + \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}^*) - F(\mathbf{w}^*)\right) \\
&\quad - \frac{\mu}{2}\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 \\
&\le -\underbrace{\sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F(\mathbf{w}^*)\right)}_{\mathcal{D}_1} - \frac{\mu}{2}\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 + \Gamma
\end{aligned}
$$

where $\Gamma \triangleq \sum_{i=1}^{N} p_i \left(F(\mathbf{w}^*) - F_i(\mathbf{w}^*)\right)$.

Process the $\mathcal{D}_1$

$$
\begin{aligned}
\mathcal{D}_1 &= \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F(\mathbf{w}^*)\right) \\
&= \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F_i(\bar{\mathbf{w}}_t)\right) + \sum_{i=1}^{N} p_i \left(F_i(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) \\
&\ge \sum_{i=1}^{N} p_i \langle \nabla F_i(\bar{\mathbf{w}}_t), \mathbf{w}_t^i - \bar{\mathbf{w}}_t\rangle + \left(F(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) \\
&\quad \text{(from the Assumption 2)} \\
&\ge -\frac{1}{2}\sum_{i=1}^{N} p_i \left[\eta \|F_i(\bar{\mathbf{w}}_t)\|^2 + \frac{1}{\eta}\|\mathbf{w}_t^i - \bar{\mathbf{w}}_t\|^2\right] \\
&\quad + \left(F(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) \\
&\quad \text{(from the AM-GM inequality)} \\
&\ge -\sum_{i=1}^{N} p_i \left[\eta L \left(F_i(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) + \frac{1}{2\eta}\|\mathbf{w}_t^i - \bar{\mathbf{w}}_t\|^2\right] \\
&\quad + \left(F(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) \\
&\quad \text{(from the Assumption 1)} \\
&= (1 - \eta L)\left(F(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) + \frac{1}{2\eta}\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \bar{\mathbf{w}}_t\|^2
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
\mathcal{B}_1 &\le \frac{1}{\eta}\sum_{i=1}^{N} p_i \|\bar{\mathbf{w}}_t - \mathbf{w}_t^i\|^2 + \eta C^2 \\
&\quad + (\eta L - 1)\left(F(\bar{\mathbf{w}}_t) - F(\mathbf{w}^*)\right) - \frac{1}{2\eta}\|\mathbf{w}_t^i - \bar{\mathbf{w}}_t\|^2 \\
&\quad - \frac{\mu}{2}\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 + \Gamma \\
&\le -\frac{\mu}{2}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + (\frac{1}{\eta} + \frac{\mu}{2})\sum_{i=1}^{N} p_i \|\bar{\mathbf{w}}_t - \mathbf{w}_t^i\|^2 + \eta C^2 + \Gamma
\end{aligned}
$$

Comprehensive equation $\mathcal{A} - \mathcal{D}$, take exception to (22):

$$
\begin{aligned}
&\mathbb{E}\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2 = \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + 2\eta \mathbb{E}\mathcal{A}_1 + \eta^2 \mathbb{E}\mathcal{A}_2 \\
&\le \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + \eta^2 \left(C^2 + \frac{\sigma^2 C^2 d}{\hat{B}^2}\right) \\
&\quad + 2\eta(-\frac{\mu}{2}\mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 \\
&\quad + \left(\frac{1}{\eta} + \frac{\mu}{2}\right)\sum_{i=1}^{N} p_i \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}_t^i\|^2 + \eta C^2 + \Gamma) \\
&= (1 - \eta\mu)\mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + (2 + \eta\mu)\sum_{i=1}^{N} p_i \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}_t^i\|^2 \\
&\quad + 3\eta^2 C^2 + 2\eta\Gamma + \frac{\eta^2 \sigma^2 C^2 d}{\hat{B}^2}
\end{aligned} \tag{24}
$$

Then we do such process:

$$
\begin{aligned}
&\sum_{i=1}^{N} p_i \mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}_t^i\|^2 \\
&= \sum_{i=1}^{N} p_i \mathbb{E}\|(\mathbf{w}_t^i - \bar{\mathbf{w}}_{t_0}) - (\bar{\mathbf{w}}_t - \bar{\mathbf{w}}_{t_0})\|^2 \\
&\le \sum_{i=1}^{N} p_i \mathbb{E}\sum_{t=t_0}^{t-1} (\tau - 1)\eta^2 \left\|\frac{1}{|\mathcal{B}_t^i|}\left(\sum_{s\in\mathcal{B}_t^i} \nabla F_i(\mathbf{w}_t^i, s) + \sigma C \mathcal{N}(0, \mathbf{I}_d)\right)\right\|^2 \\
&\le \sum_{i=1}^{N} p_i \eta^2 (\tau - 1)^2 \left(\mathbb{E}\left\|\frac{1}{|\mathcal{B}_t^i|}\sum_{s\in\mathcal{B}_t^i} \nabla F_i(\mathbf{w}_t^i, s) +\right\|^2 \frac{\sigma^2 C^2 d}{\hat{B}^2}\right) \\
&\le \eta^2 (\tau - 1)^2 \left(C^2 + \frac{\sigma^2 C^2 d}{\hat{B}^2}\right)
\end{aligned} \tag{25}
$$

Combine equation (22) and equation (25):

$$
\mathbb{E}\|\bar{\mathbf{v}}_{t+1} - \mathbf{w}^*\|^2 \le (1 - \eta\mu)\mathbb{E}\|\bar{\mathbf{w}}_t - \mathbf{w}^*\|^2 + \eta A \tag{26}
$$

where

$$
A = (\tau - 1)^2 (2 + \eta\mu)\left(C^2 + \frac{\sigma^2 C^2 d}{\hat{B}^2} + 3C^2 + \frac{2\Gamma}{\eta} + \frac{\sigma^2 C^2 d}{\hat{B}^2}\right) \tag{27}
$$