I. APPENDIX

A. Proof of Lemma 1

Lemma 1: When historical gradient compensation is applied, the weight for each client in each round is bounded by

$$p_{k,r} \leq \frac{1}{1+(K-1)(2\tau_{max}+K)^{-\alpha_s}} \tag{1}$$

Proof: Given the application of historical gradient compensation, for a client not in S_r, its staleness is calculated as the current trained model's staleness ($\leq \tau_{\max}$) plus the staleness of its last uploaded model ($\leq \tau_{\max}$), along with the round difference between its two consecutive updates.

Because satellites arrive sequentially, after a satellite's last update, if it does not receive the latest model during its communication window, it must wait until the next visible window. During the interval between two windows, other satellites may upload their models and update the global model. Thus, there can be at most $K-1$ global updates between two communication windows for a given satellite. Since the staleness of models trained in round r is constrained, the staleness of the historical gradient will not exceed $2\tau_{\max}+K-1$, *i.e.*, $0 \leq \tau_{k,r} \leq 2\tau_{\max}+K-1$. Therefore, we have

$$\begin{aligned} p_{k,r} &= \frac{q_{k,r}}{q_{k,r}+\sum_{j\neq k} q_{j,r}} \\ &\leq \frac{q_{k,r}}{q_{k,r}+(K-1)(2\tau_{max}+K)^{-\alpha_s}} \\ &\leq \frac{1}{1+(K-1)(2\tau_{max}+K)^{-\alpha_s}} \end{aligned} \tag{2}$$

For brevity of notation, the subsequent analysis denotes $p_{k,r} \leq C$, where $C = \frac{1}{1+(K-1)(2\tau_{max}+K)^{-\alpha_s}}$.

B. Proof of Lemma 2

Lemma 2: If Assumption 1, 2 and 4 are satisfied, the deviation of local mode from the global model after i iterations is bounded by

$$\mathbb{E}\|\mathbf{w}_{k,r}^i - \mathbf{w}_r\|^2 \leq 4\eta^2(I-1)(2IG^2+\varsigma^2) \tag{3}$$

Proof: Note that Lemma 1 is satisfied when $i=0$ due to $\mathbf{w}_{k,r}^0 = \mathbf{w}_r$. For the case of $i \geq 1$, we have

$$\begin{aligned} \mathbb{E}\|\mathbf{w}_{k,r}^i - \mathbf{w}_r\|^2 &= \mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \eta\nabla\tilde{F}_k(\mathbf{w}_{k,r}^{i-1}) - \mathbf{w}_r\|^2 \\ &= \mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \eta\nabla F_k(\mathbf{w}_{k,r}^{i-1}) - \mathbf{w}_r\|^2 \\ &\quad + \eta^2\mathbb{E}\|\nabla\tilde{F}_k(\mathbf{w}_{k,r}^{i-1}) - \nabla F_k(\mathbf{w}_{k,r}^{i-1})\|^2 \\ &\leq \mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \eta\nabla F_k(\mathbf{w}_{k,r}^{i-1}) - \mathbf{w}_r\|^2 + \eta^2\varsigma^2 \end{aligned} \tag{4}$$

where the second equation is due to the unbiased stochastic gradient assumption. The last inequation is due to Assumption 2. For the first term of the RHS of (4), we have

$$\begin{aligned} &\mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \eta\nabla F_k(\mathbf{w}_{k,r}^{i-1}) - \mathbf{w}_r\|^2 \\ &= \mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r\|^2 + \eta^2\mathbb{E}\|\nabla F_k(\mathbf{w}_{k,r}^{i-1})\|^2 \\ &\quad - 2\mathbb{E}\langle \frac{1}{\sqrt{I-1}}(\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r), \sqrt{I-1}\eta\nabla F_k(\mathbf{w}_{k,r}^{i-1})\rangle \\ &\leq (1+\frac{1}{I-1})\mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r\|^2 + \eta^2 I\mathbb{E}\|\nabla F_k(\mathbf{w}_{k,r}^{i-1})\|^2 \\ &\leq (1+\frac{1}{I-1})\mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r\|^2 + 2\eta^2 I\mathbb{E}\|\nabla F_k(\mathbf{w}_r)\|^2 \\ &\quad + 2\eta^2 I\mathbb{E}\|\nabla F_k(\mathbf{w}_{k,r}^{i-1}) - \nabla F_k(\mathbf{w}_r)\|^2 \\ &\leq (1+\frac{1}{I-1}+2\eta^2 IL^2)\mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r\|^2 + 2\eta^2 IG^2 \end{aligned} \tag{5}$$

where the first inequation holds by the Young's Inequality and the second holds by triangle inequality. The last inequation comes from Assumption 1 and 4. Substituting (5) into (4) obtains

$$\begin{aligned} \mathbb{E}\|\mathbf{w}_{k,r}^i - \mathbf{w}_r\|^2 \leq (1+\frac{1}{I-1}+2\eta^2 IL^2)\mathbb{E}\|\mathbf{w}_{k,r}^{i-1} - \mathbf{w}_r\|^2 \\ + 2\eta^2 IG^2 + \eta^2\varsigma^2 \end{aligned} \tag{6}$$

Telescoping the above inequation obtains

$$\begin{aligned} &\mathbb{E}\|\mathbf{w}_{k,r}^i - \mathbf{w}_r\|^2 \\ &\leq (2\eta^2 IG^2 + \eta^2\varsigma^2)\frac{(1+\frac{1}{I-1}+\frac{1}{2K^2 I})^{I-1}-1}{\frac{1}{I-1}+\frac{1}{2K^2 I}} \end{aligned} \tag{7}$$

where the first inequation holds by $K\eta \leq \frac{1}{2IL}$. Because $a = \frac{1}{I-1}+\frac{1}{2K^2 I} \geq \frac{1}{I-1}$, $(1+a)^{I-1} = e^{(I-1)ln(1+a)} \leq e^{(I-1)a}$. Then, we have

$$\begin{aligned} (I-1)a &= (I-1)(\frac{1}{I-1}+\frac{1}{2K^2 I}) \\ &= 1+\frac{I-1}{2K^2 I} \\ &\leq 1+\frac{1}{2K^2} \leq \frac{3}{2} \end{aligned} \tag{8}$$

Therefore, we have

$$\frac{(a+1)^{I-1}-1}{a} \leq \frac{e^{\frac{3}{2}}-1}{a} \leq (e^{\frac{3}{2}}-1)(I-1) \leq 4(I-1) \tag{9}$$

Substituting (9) into (7) obtains

$$\mathbb{E}\|\mathbf{w}_{k,r}^i - \mathbf{w}_r\|^2 \leq 4\eta^2(I-1)(2IG^2+\varsigma^2) \tag{10}$$

C. Proof of Lemma 3

Lemma 3: If Assumptions 1, 2, and 4 is satisfied, the drift between the global models in two rounds, *i.e.*, r_1 and r_2 ($r_1 > r_2$), is bounded by

$$\mathbb{E}\|\mathbf{w}_{r_1} - \mathbf{w}_{r_2}\|^2 \leq 3\eta^2(r_1-r_2)^2 C^2 D_1 \tag{11}$$

where $D_1 = (I^2K^2+I-1)\varsigma^2 + (I^2K^2+2I(I-1))G^2$.

Proof: For the two different rounds, *i.e.*, r_1 and r_2 ($r_1 > r_2$), we have

$$\begin{aligned}
&\mathbb{E}\|\mathbf{w}_{r_1} - \mathbf{w}_{r_2}\|^2 \\
&= \mathbb{E}\|\sum_{r=r_2}^{r_1-1}(\mathbf{w}_{r+1} - \mathbf{w}_r)\|^2 \\
&= \eta^2\mathbb{E}\|\sum_{r=r_2}^{r_1-1}\sum_{k=1}^{K}\sum_{i=0}^{I-1} p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\leq \eta^2(r_1-r_2)KI\sum_{r=r_2}^{r_1-1}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\leq \eta^2(r_1-r_2)KIC^2\sum_{r=r_2}^{r_1-1}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2
\end{aligned} \tag{12}$$

where the first inequation comes from using Cauchy-Schwarz Inequality and , and the last inequation is due to Lemma 1.

$$\begin{aligned}
&\mathbb{E}\|\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&= \mathbb{E}\|\underbrace{\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})}_{A_1} \\
&\quad + \underbrace{\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}_{r-\tau_{k,r}})}_{A_2} \\
&\quad + \underbrace{\nabla F_k(\mathbf{w}_{r-\tau_{k,r}})}_{A_3}\|^2 \\
&\leq 3\mathbb{E}\|A_1\|^2 + 3\mathbb{E}\|A_2\|^2 + 3\mathbb{E}\|A_3\|^2 \\
&\leq 3\varsigma^2 + 12\eta^2(I-1)(2IG^2+\varsigma^2)L^2 + 3G^2
\end{aligned} \tag{13}$$

where the first inequation comes from using the triangle inequality, and the second inequation is due to Assumption 1 and Lemma 2. Substituting (13) into (12), considering $K\eta \leq \frac{1}{2IL}$, we have

$$\begin{aligned}
\mathbb{E}\|\mathbf{w}_{r_1} - \mathbf{w}_{r_2}\|^2 &\leq \eta^2(r_1-r_2)^2K^2I^2C^2 \\
&\quad (3\varsigma^2 + 12\eta^2(I-1)(2IG^2+\varsigma^2)L^2 + 3G^2) \\
&\leq 3\eta^2(r_1-r_2)^2C^2D_1
\end{aligned} \tag{14}$$

where $D_1 = (I^2K^2 + I - 1)\varsigma^2 + (I^2K^2 + 2I(I-1))G^2$.

D. Proof of Theorem 1

Theorem 1: If Assumptions 1-4 is satisfied, and $K\eta \leq \frac{1}{2IL}$, the R-round convergency bound of SC-SFL is

$$\begin{aligned}
\mathbb{E}[F(\mathbf{w}_R) - F^*] &\leq c_1^R\mathbb{E}[F(\mathbf{w}_0) - F^*] + c_2\sum_{r=0}^{R-1}c_1^r \\
&\quad + c_3\sum_{r=0}^{R-2}c_1^{R-r-2}\sum_{k=1}^{K}(1-s_{k,r})(\tau_{k,r}+1)^2
\end{aligned} \tag{15}$$

where $c_1 = (1 - \eta I\mu + 2CKI^2L\mu\eta^2)$, $c_2 = 12\eta^3L^2KI(I-1)(2IG^2+\varsigma^2)(\frac{1}{2}C + CIL\eta) + \frac{3}{2}CK\eta I\Gamma^2 + \frac{1}{2}\eta KC^2\varsigma^2 + \frac{1}{2}CKI^2\eta^2L\varsigma^2 + 3CKI^2\eta^2L\Gamma^2$ and $c_3 = 3\eta^3C^2D_1IL^2(\frac{3}{2}C + \frac{7}{2} + 3CI\eta L)$.

Proof: According to Assumption 1, we have

$$\begin{aligned}
&\mathbb{E}[F(\mathbf{w}_{r+1}) - F(\mathbf{w}_r)] \\
&\leq \mathbb{E}\langle\nabla F(\mathbf{w}_r), \mathbf{w}_{r+1} - \mathbf{w}_r\rangle + \frac{L}{2}\mathbb{E}||\mathbf{w}_{r+1} - \mathbf{w}_r||^2 \\
&= -\eta\mathbb{E}\langle\nabla F(\mathbf{w}_r), \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\rangle \\
&\quad + \frac{\eta^2L}{2}\mathbb{E}||\sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})||^2
\end{aligned} \tag{16}$$

For the first term of (15), we have

$$\begin{aligned}
&\mathbb{E}\langle\nabla F(\mathbf{w}_r), \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\rangle \\
&= \mathbb{E}\langle\nabla F(\mathbf{w}_r), \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) \\
&\quad - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) + \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\rangle \\
&= \mathbb{E}\langle\nabla F(\mathbf{w}_r), \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\rangle \\
&\quad + \mathbb{E}\langle\nabla F(\mathbf{w}_r), b_1\rangle
\end{aligned} \tag{17}$$

where $b_1 = \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]$. Substituting (17) into (16), we have

$$\begin{aligned}
&\mathbb{E}[F(\mathbf{w}_{r+1}) - F(\mathbf{w}_r)] \\
&\underbrace{\leq -\eta\mathbb{E}\langle\nabla F(\mathbf{w}_r), \sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\rangle}_{B_1} \\
&\underbrace{-\eta\mathbb{E}\langle\nabla F(\mathbf{w}_r), b_1\rangle}_{B_2} + \frac{\eta^2L}{2}\underbrace{\mathbb{E}\|\sum_{k=1}^{K}\sum_{i=0}^{I-1}p_{k,r}\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2}_{B_3}
\end{aligned} \tag{18}$$

For term B_1, we have

$$\begin{aligned}
B_1 &= -\eta I\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 \\
&\quad + \eta I\mathbb{E}\langle\nabla F(\mathbf{w}_r), \nabla F(\mathbf{w}_r) - \sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\rangle \\
&\leq -\eta I\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 + \frac{\eta I}{2}\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 \\
&\quad + \frac{\eta I}{2}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&= -\frac{\eta I}{2}\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 \\
&\quad + \frac{\eta I}{2}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2
\end{aligned} \tag{19}$$

where the first equation is due to the unbiased stochastic gradient assumption and the first inequation holds by triangle inequality. For the last term on the RHS of (19), we have

$$
\begin{aligned}
&\mathbb{E}\|\nabla F(\mathbf{w}_r) - \sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&= \mathbb{E}\|\sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}[\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\|^2 \\
&\le \mathbb{E}[\sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{p_{k,r}}{I}\|\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2] \\
&\le \mathbb{E}[\sum_{k=1}^{K}\sum_{i=0}^{I-1}\frac{C}{I}\|\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2] \\
&\le \frac{C}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2
\end{aligned}
\tag{20}
$$

where the first inequation holds by Jensen's inequality and the second inequation holds by Lemma 1. According to Assumption 1 and 3, we have

$$
\begin{aligned}
&\frac{C}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\le \frac{3C}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \nabla F_k(\mathbf{w}_r)\|^2 \\
&+ \frac{3C}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F_k(\mathbf{w}_r) - \nabla F_k(\mathbf{w}_{r-\tau_{k,r}})\|^2 \\
&+ \frac{3C}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F_k(\mathbf{w}_{r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\le 3CK\Gamma^2 + 3CL^2\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_r - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
&+ \frac{3CL^2}{I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\mathbf{w}_{r-\tau_{k,r}} - \mathbf{w}^i_{k,r-\tau_{k,r}}\|^2
\end{aligned}
\tag{21}
$$

Substituting (20) and (21) into (19), we have

$$
\begin{aligned}
B_1 &\le -\frac{\eta I}{2}\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 + \frac{3}{2}CK\eta I\Gamma^2 \\
&+ \frac{3}{2}C\eta IL^2\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_r - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
&+ \frac{3}{2}C\eta L^2\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\mathbf{w}_{r-\tau_{k,r}} - \mathbf{w}^i_{k,r-\tau_{k,r}}\|^2
\end{aligned}
\tag{22}
$$

For the term B_2, we have

$$
\begin{aligned}
B_2 &= -\eta I\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\langle\nabla F(\mathbf{w}_r), \\
&\qquad \frac{p_{k,r}}{I}[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\rangle \\
&= -\eta I\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\langle\nabla F(\mathbf{w}_r) - \nabla F(\mathbf{w}_{r-\tau_{k,r}}), \\
&\qquad \frac{p_{k,r}}{I}[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\rangle \\
&\le \frac{\eta I^2}{2}\sum_{k=1}^{K}\mathbb{E}\|\nabla F(\mathbf{w}_r) - \nabla F(\mathbf{w}_{r-\tau_{k,r}})\|^2 \\
&+ \frac{\eta I}{2}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\frac{p_{k,r}}{I}[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\|^2 \\
&\le \frac{\eta I^2L^2}{2}\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_r - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
&+ \frac{\eta C^2}{2I}\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|[\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})]\|^2 \\
&\le \frac{\eta I^2L^2}{2}\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_r - \mathbf{w}_{r-\tau_{k,r}}\|^2 + \frac{1}{2}\eta KC^2\varsigma^2
\end{aligned}
\tag{23}
$$

where the first inequation holds by Young's inequality and the last two inequation is due to Assumption 1 and 2. For the term B_3, we have

$$
\begin{aligned}
B_3 &\le CI\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla\tilde{F}_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\le CKI^2\varsigma^2 + CI\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})\|^2 \\
&\le CKI^2\varsigma^2 \\
&+ \underbrace{2CI\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}}) - \nabla F(\mathbf{w}_r)\|^2}_{B_4} \\
&+ 2CKI^2\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2
\end{aligned}
\tag{24}
$$

where the first inequation holds by Jensen's inequality and Lemma 1, and the second inequation is due to Assumption 2. The third inequation is by adding and subtracting $\nabla F(\mathbf{w}_r)$ and $\nabla F_k(\mathbf{w}^i_{k,r-\tau_{k,r}})$. For the term B_4, we have

$$
\begin{aligned}
B_4 &\le 6CIL^2\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\mathbf{w}^i_{k,r-\tau_{k,r}} - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
&+ 6CI^2L^2\mathbb{E}\|\mathbf{w}_{r-\tau_{k,r}} - \mathbf{w}_r\|^2 \\
&+ 6CI^2\sum_{k=1}^{K}\mathbb{E}\|\nabla F_k(\mathbf{w}_r) - \nabla F(\mathbf{w}_r)\|^2 \\
&\le 6CIL^2\sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\mathbf{w}^i_{k,r-\tau_{k,r}} - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
&+ 6CI^2L^2\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_{r-\tau_{k,r}} - \mathbf{w}_r\|^2 \\
&+ 6CKI^2\Gamma^2
\end{aligned}
\tag{25}
$$

where the first inequation is by adding and subtracting $\nabla F_k(\mathbf{w}_{r-\tau_{k,r}})$ and $\nabla F_k(\mathbf{w}_r)$ and using Assumption 1. The second inequation is due to Assumption 3. Substituting (24) into (25), we have

$$
\begin{aligned}
B_3 \leq\ & CKI^2\varsigma^2 + +6CKI^2\Gamma^2 + 2CKI^2\mathbb{E}\|\nabla F(\mathbf{w}_r)\|^2 \\
& + 6CIL^2 \sum_{k=1}^{K}\sum_{i=0}^{I-1}\mathbb{E}\|\mathbf{w}_{k,r-\tau_{k,r}}^{i} - \mathbf{w}_{r-\tau_{k,r}}\|^2 \\
& + 6CI^2L^2\sum_{k=1}^{K}\mathbb{E}\|\mathbf{w}_{r-\tau_{k,r}} - \mathbf{w}_r\|^2
\end{aligned}
\tag{26}
$$

Substituting (22), (23) and (26) into (18), and using Assumption 1, Lemma 2 and Lemma 3, we have

$$
\begin{aligned}
& \mathbb{E}[F(\mathbf{w}_{r+1}) - F(\mathbf{w}_r)] \\
& \leq (-\eta I\mu + 2CKI^2L\mu\eta^2)\mathbb{E}[F(\mathbf{w}_r) - F^*] \\
& \quad + c_2 + c_3\sum_{k=1}^{K}\tau_{k,r}^2
\end{aligned}
\tag{27}
$$

where $c_2 = 12\eta^3L^2KI(I-1)(2IG^2+\varsigma^2)(\frac{1}{2}C + CIL\eta) + \frac{3}{2}CK\eta I\Gamma^2 + \frac{1}{2}\eta KC^2\varsigma^2 + \frac{1}{2}CKI^2\eta^2L\varsigma^2 + 3CKI^2\eta^2L\Gamma^2$ and $c_3 = 3\eta^3C^2D_1IL^2(\frac{3}{2}C + \frac{I}{2} + 3CI\eta L)$.

Subtracting F^* to $F(\mathbf{w}_{r+1})$ and $F(\mathbf{w}_r)$, we have

$$
\begin{aligned}
& \mathbb{E}[F(\mathbf{w}_{r+1}) - F^*] \\
& \leq (1-\eta I\mu + 2CKI^2L\mu\eta^2)\mathbb{E}[F(\mathbf{w}_r) - F^*] \\
& \quad + c_2 + c_3 bV\sum_{k=1}^{K}\tau_{k,r}^2
\end{aligned}
\tag{28}
$$

Since the staleness of each client k evolves as

$$
\tau_{k,r} = \begin{cases} \tau_{k,r-1}+1, & \text{if } s_{k,r-1} = 0 \\ 0, & \text{else} \end{cases},
\tag{29}
$$

we have $\tau_{k,r}^2 = (1-s_{k,r-1})^2(\tau_{k,r-1}+1)^2 = (1-s_{k,r-1})(\tau_{k,r-1}+1)^2$. Noticing that the $\tau_{k,0} = 0$, and telescoping the above inequation, we have

$$
\begin{aligned}
& \mathbb{E}[F(\mathbf{w}_R) - F^*] \\
& \leq c_1^R\mathbb{E}[F(\mathbf{w}_0) - F^*] + c_2\sum_{r=0}^{R-1}c_1^r \\
& \quad + c_3\sum_{r=1}^{R-1}c_1^{R-r-1}\sum_{k=1}^{K}(1-s_{k,r-1})(\tau_{k,r-1}+1)^2 \\
& = c_1^R\mathbb{E}[F(\mathbf{w}_0) - F^*] + c_2\sum_{r=0}^{R-1}c_1^r \\
& \quad + c_3\sum_{r=0}^{R-2}c_1^{R-r-2}\sum_{k=1}^{K}(1-s_{k,r})(\tau_{k,r}+1)^2
\end{aligned}
\tag{30}
$$

where $c_1 = (1-\eta I\mu + 2CKI^2L\mu\eta^2)$.