Online Appendix

1 proof sketch of theorems and lemmas

1.1 Differential Privacy Theorem Analysis

Lemma 1 (Adaptive Estimation of Actual Selected Times of the Clients)**.** *This Lemma is the Lemma 1 in the paper. Given the total sampling times $T = KR$, $\mathbf{q} = \{q_1, \ldots, q_N\}$, the probability that client i is actually sampled more than R_i times is not greater than δ''' if*

$$R_i \leqslant Tq_i + \ln\left(\frac{1-\delta'''}{\delta'''}\right)\frac{\sqrt{Tq_i(1-q_i)}}{k}. \tag{1}$$

Proof. To solve for r with violation probability δ''', we need to solve $B(r, T, q_i) \leqslant 1 - \delta'''$, where $B(r, T, q_i)$ is the cumulative function of Bernoulli distribution. Since T is quite bigger and client sampling rate q_i is smaller enough, we can approximate this binomial distribution to normal distribution $t \sim \mathcal{N}(\mu = Tq_i, \sigma^2 = Tq_i(1-q_i)) = \sqrt{Tq_i(1-q_i)}\mathcal{Z} + Tq_i$, where $\mathcal{Z}$ is the standard Normal distribution. So we can transform the equation r as follows,

$$\begin{aligned} 1-\delta''' &= \sum_{t=0}^{r} q^t(1-q)^{T-t}\binom{T}{t} && (2)\\ &\approx \frac{1}{\sqrt{2\pi Tq_i(1-q_i)}}\int_{-\infty}^{t} e^{-\frac{(t-Tq_i)^2}{2Tq_i(1-q_i)}}\,dt && (3) \end{aligned}$$

Let $\Phi(t)$ denote the cdf of $\mathcal{N}(0,1)$, by applying the sigmoid approximation $\Phi(x) = \frac{1}{1+exp(-kx)}$ proposed by Waissi and Rossin, we can derive the actually selected times R_i to be used in privacy loss analysis through solving

$$\Phi(\frac{R_i - Tq_i}{\sqrt{Tq_i(1-q_i)}}) = \frac{1}{1+exp(-k(\frac{R_i-Tq_i}{\sqrt{Tq_i(1-q_i)}}))} = 1-\delta''' \tag{4}$$

Theorem 1 (Privacy Guarantee for Output Perturbation)**.** *This Theorem does not appear in the paper but it is important for the analysis of differnetial privacy. Let $T = KR$ big enough, given the personalized privacy budget $\epsilon = \{\epsilon_1, \ldots, \epsilon_N\}$,* *tolerant probability of exceeding privacy budget* $0 < \delta''' \ll 1$*, Algorithm 1 is* (ϵ_i, δ_i)*-DP towards a third party for client* i *if we choose*

$$\sigma_i = \Omega\left(s_i\sqrt{T_i^* log(1/\delta')}/\epsilon_i\right),$$

where $\delta = \delta' + \delta_i''' - \delta_i'\delta_i'''$, $T_i^* = Tq_i + \ln\left(\frac{1-\delta'''}{\delta'''}\right)\frac{\sqrt{Tq_i(1-q_i)}}{k}$

Proof Sketch.

Theorem 2 (Privacy Guarantee for Gradient Perturbation)**.** *This Theorem is the Theorem 1 in the paper. Let* $T = KR$ *big enough, given the personalized privacy budget* $\epsilon = \{\epsilon_1, \ldots, \epsilon_N\}$*, tolerant probability of exceeding privacy budget* $0 < \delta''' \ll 1$*, Algorithm 1 is* (ϵ_i, δ_i)*-DP towards a third party for client* i *if*

$$\sigma_{i,g} = \Omega(s_i\sqrt{T^* E \log(1/\delta')\log(1/\delta'')}/\epsilon_i),$$

where $\delta = T\delta' + \delta'' + \delta''' - (T\delta' + \delta'')\delta'''$.

Proof Sketch. The proof basically follows Maxence's analysis for T rounds' composition in global DP. Since we consider the case in LDP, we substitute the term $\sqrt{lT}$ with the actual participating rounds $\sqrt{T_i^*}$ and get the upper bound. In order to obtain a more concise convergence result, we consider further bounding T_i^* to obtain a more concise expression.

Corollary 1. *Let* $T = KR$ *big enough, given the personalized privacy budget* $\epsilon = \{\epsilon_1, \ldots, \epsilon_N\}$*, tolerant probability of exceeding privacy budget* $0 < \delta''' \ll 1$*, if* $R < (log(1/\delta'''))^2/(q_iKk^2)$*, Algorithm 1 is* (ϵ_i, δ_i)*-DP towards a third party for client* i *if*

$$\sigma_{i,g} = \Omega(\sqrt{E\sqrt{q_iKR}}B_i),$$

where $B_i = s_i\sqrt{\log(1/\delta')\log(1/\delta'')log(1/\delta''')}/\epsilon_i$

Proof. Noticing

$$T_i^* = Tq_i + \ln\left(\frac{1-\delta'''}{\delta'''}\right)\frac{\sqrt{Tq_i(1-q_i)}}{k} \tag{5}$$

$$\leqslant Tq_i + \ln\left(\frac{1}{\delta'''}\right)\frac{\sqrt{Tq_i}}{k} \tag{6}$$

$$\tag{7}$$

Hence when $T \leqslant (log(1/\delta'''))^2/(q_ik^2)$, $\ln\left(\frac{1}{\delta'''}\right)\frac{\sqrt{Tq_i}}{k}$ *is greater or equal to the first term* Tq_i*, so the second term* $\ln\left(\frac{1}{\delta'''}\right)\frac{\sqrt{Tq_i}}{k}$ *is the dominant term, and we have* $T_i^* = \Omega(log(1/\delta''')\sqrt{q_iKR})$*. Substituting this upper bound, we obtain the corallary.*

1.2 Convergence Analysis

For the part of convergence analysis, we first recall the three assumptions we made in our paper. These assumptions are general in other papers like [**?**], and our main theorems are basing on it too.

Assumption 1. *For each client $i \in \mathcal{N}$, $F_i(w)$ is L-Smooth*

$$\|\nabla F_i(v) - \nabla F_i(w)\| \leqslant L \|v - w\|, \tag{8}$$

for any v and w and some $L > 0$.

Assumption 2. *Local stochastic gradients are unbiased: for any $\mathbf{w}$, we have*

$$\mathbb{E}[\mathbf{g}_n(\mathbf{w})|\mathbf{w}] = \nabla f_n(\mathbf{w}). \tag{9}$$

Assumption 3 (Bounded stochastic gradients)**.**

$$\mathbb{E}\left[\|\mathbf{g}_n(\mathbf{y})\|^2\right] \leq G_n^2, \forall \mathbf{y}, n \tag{10}$$

for some $G_n > 0$.

Theorem 3 (Non-convex Convergence Upper Bound With Gradient Perturbation)**.** *In the paper, we just show the corollary of this Theorem, but we think it isessentialt to post the detailed proof process here. Let Assumptions 1, 2, and 3 hold with γ, T, I, N, d, q_n^t, $\sigma_{t,n}$, defined as above, we have*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f(\mathbf{x}_t)\|^2\right] \leq \Phi + \frac{\gamma LNS^2}{K}\sum_{n=1}^{N}\theta_n p_n^2\sigma_n^2 \tag{11}$$

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f(\mathbf{x}_t)\|^2\right] = \Omega(\Phi + \eta LNS^2\sum_{n=1}^{N}q_n\sigma_n^2) \tag{12}$$

where

$$\begin{aligned}\Phi =& \frac{2\left(f\left(\mathbf{x}_0\right) - f^*\right)}{\gamma TI} \\ &+ \frac{\gamma^2L^2N(I-1)}{IT}\sum_{t=0}^{T-1}\sum_{n=1}^{N}\sum_{i=0}^{I-1}p_n^2\sum_{j=0}^{i-1}\mathbb{E}\left[\left\|\mathbf{g}_n\left(\mathbf{y}_{t,j}^n\right)\right\|^2\right] \\ &+ \frac{\gamma LN}{KT}\sum_{t=0}^{T-1}\sum_{n=1}^{N}p_n^2\theta_n\sum_{i=0}^{I-1}\mathbb{E}\left[\left\|\mathbf{g}_n\left(\mathbf{y}_{t,j}^n\right)\right\|^2\right],\end{aligned}$$

$S = \frac{2\sum_g C}{m}$, *and $\theta_n = K - 1 + \frac{1}{q_n}$.*

Applying Assumption 3 and Theorem 2, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f(\mathbf{x}_t)\|^2\right] = \mathcal{O}\Omega(\phi + \gamma\ln\left(\frac{1}{\delta'''}\right)LNE\sqrt{R}\sum_{n=1}^{N}B_n\left(\sqrt{q_n} + \frac{1}{\sqrt{Kq_n}}\right)\Bigg) \tag{13}$$

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla f\left(\mathbf{x}_t\right)\|^2\right]=\mathcal{O}\Omega(\phi+\gamma\ln\left(\frac{1}{\delta'''}\right)LNE\sqrt{R}\sum_{n=1}^{N}B_n\left(\sqrt{q_n}+\frac{1}{\sqrt{Kq_n}}\right)\right) \tag{14}$$

*Proof Sketch Follow the main steps of [**?**] Recall that*

$$x_{t+1}-x_t=\frac{1}{N}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}(y_{t,I}^n+\sum_{i=0}^{I-1}\gamma z-y_{t,0}^n) \tag{15}$$

$$=-\frac{\gamma}{k}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}g_n(y_{t,i}^n)+\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}z \tag{16}$$

Then from L-smooth, we have

$$\mathbb{E}[f(x_{t+1}|x_t)]\leqslant f(x_t)+\langle\nabla f(x_t),\mathbb{E}(x_{t+1}-x_t|x_t)\rangle+\frac{L}{2}\mathbb{E}[\|\ x_{t+1}-x_t\ \|^2\ |x_t] \tag{17}$$

$$=f(x) \tag{18}$$

$$+\underbrace{\left\langle\nabla f(x_t),\mathbb{E}[-\frac{\gamma}{k}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}g_n(y_{t,i}^n)+\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}z]\right\rangle}_{A} \tag{19}$$

$$+\underbrace{\frac{L}{2}\mathbb{E}[\|-\frac{\gamma}{k}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}g_n(y_{t,i}^n)+\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}z\ \|^2]}_{B} \tag{20}$$

For A, $\left\langle\nabla f(x_t),\mathbb{E}[\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^tP_n}{q_n}\sum_{i=0}^{I-1}z|x_t]\right\rangle=0$*, and because of* $\mathbb{E}[\mathbb{C}_n^t|x_t]=$

Kq_n, *the rest part of A can be bounded.*

$$\left\langle \nabla f(x_t), \mathbb{E}[-\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n)] \right\rangle \tag{21}$$

$$= -\gamma\sum_{i=0}^{I-1}\mathbb{E}[\left\langle \nabla f(x_t), \sum_{n=1}^{N} P_n \nabla f_n(y_{t,i}^n) - \nabla f(x_t) + \nabla f(x_t) \right\rangle] \tag{22}$$

$$\leqslant \frac{\gamma I}{2}\mathbb{E} \parallel \nabla f(x_t) \parallel^2 + \frac{\gamma}{2}\sum_{i=0}^{I-1}\mathbb{E}[\sum_{n=1}^{N} P_n(\nabla f_n(x_t) - \nabla f_n(y_{t,i}^n))] \tag{23}$$

$$- \gamma I \mathbb{E}[\parallel \nabla f(x_t) \parallel^2] \tag{24}$$

$$\leqslant \frac{\gamma L^2 I N}{2}\sum_{n=1}^{N}\mathbb{E}[P_n^2 \parallel x_t - y_{t,i}^n \parallel^2] - \frac{\gamma I}{2}\mathbb{E}[\parallel \nabla f(x_t) \parallel^2] \tag{25}$$

$$\leqslant \frac{\gamma L^2 N I(I-1)}{2}\sum_{n=1}^{N} P_n^2 \sum_{j=0}^{i-1}\mathbb{E}[\parallel g_n(y_{t,j}^n) \parallel^2] - \frac{\gamma I}{2}\mathbb{E}[\parallel \nabla f(x_t) \parallel^2] \tag{26}$$

For B, we denote that $E_1 = -\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n)$, $E_2 = \frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} z$, *Then* $B = \frac{L}{2}(\parallel E_1 \parallel^2 + \parallel E_2 \parallel^2 + 2\langle E_1, E_2\rangle)$. *For* $\parallel E_1 \parallel^2$, *we have*

$$\parallel E_1 \parallel^2 \leqslant \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N}\mathbb{E}[\parallel \frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n) \parallel^2] \tag{27}$$

$$\leqslant \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N}\frac{\mathbb{E}[\mathbb{C}_n^2] P_n^2}{q_n^2}\sum_{i=0}^{I-1}\mathbb{E}[\parallel g_n(y_{t,i}^n) \parallel^2] \tag{28}$$

$$= \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N} P_n^2 (K(K-1) + \frac{K}{q_n})\sum_{i=0}^{I-1}\mathbb{E}[\parallel g_n(y_{t,i}^n) \parallel^2] \tag{29}$$

For $\parallel E_2 \parallel^2$, *we have*

$$\parallel E_2 \parallel^2 \leqslant \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N}\mathbb{E}[\parallel \frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} z \parallel^2] \tag{30}$$

$$\leqslant \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N}\frac{\mathbb{E}[\mathbb{C}_n^2] P_n^2}{q_n^2}\mathbb{E}[\parallel \sum_{i=0}^{I-1} z \parallel^2] \tag{31}$$

$$= \frac{\gamma^2 N}{K^2}\sum_{n=1}^{N}(K(K-1) + \frac{K}{Q_n}) P_n^2 d I \sigma^2 \tag{32}$$

Since $\frac{1}{K} \geq q_n \geq \frac{p_n}{\sqrt{K}}$, *we have*

$$\mathbb{E}[\| E_2 \|^2] \leqslant \frac{\eta^2 N}{K} \sum_{n=1}^{N} \mathbb{E}[\| \mathbb{C}_n^t z_i \|^2] \tag{33}$$

$$\leqslant \frac{\eta^2 N}{K} \sum_{n=1}^{N} \mathbb{E}[\mathbb{C}_n^2] \mathbb{E}[\| \sum_{i=0}^{E-1} z \|^2] \tag{34}$$

$$\leqslant \eta^2 EN \sum_{n=1}^{N} ((K-1)q_n^2 + q_n) S^2 \sigma_n^2 \tag{35}$$

$$\leqslant \eta^2 EN \sum_{n=1}^{N} K q_n^2 S^2 \sigma_n^2 + q_n S^2 \sigma_n^2 \tag{36}$$

$$= \Omega(\eta^2 ENS^2 \sum_{n=1}^{N} q_n \sigma_n^2) \tag{37}$$

For $\langle E_1, E_2 \rangle$, *they are orthogonal, which means that* $\langle E_1, E_2 \rangle = 0$. *Add them up then rearrange it, sum t from 0 to T-1, and finally plug* σ_n *in corollary 1 into it, this theorem is verified.*

Theorem 4 (Non-convex Convergence Upper Bound With Parameter Perturbation)**.** *This Theorem does not appear in the paper, but we think it is important to show its detailed proof process. Let Assumptions 1, 2 and 3 hold with* γ, *T, I, N, d,* q_n^t, $\sigma_{t,n}$, *defined as above. Then, Algorithm 1 satisfies*

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\left\| \nabla f(\mathbf{x}_t) \right\|^2 \right] \leq \Phi + \frac{LNdS^2}{\gamma ITK} \sum_{t=0}^{T-1} \sum_{n=1}^{N} \theta_i p_n^2 \sigma_{t,n}^2 \tag{38}$$

Proof Sketch Follow the main steps of [**?**] Recall that

$$x_{t+1} - x_t = \frac{1}{N} \sum_{n=1}^{N} \frac{\mathbb{C}_n^t P_n}{q_n} (y_{t,I}^n + z - y_{t,0}^n) \tag{39}$$

$$= -\frac{\gamma}{k} \sum_{n=1}^{N} \frac{\mathbb{C}_n^t P_n}{q_n} \sum_{i=0}^{I-1} g_n(y_{t,i}^n) + \frac{1}{K} \sum_{n=1}^{N} \frac{\mathbb{C}_n^t P_n}{q_n} z \tag{40}$$

Then from L-smooth, we have

$$\mathbb{E}[f(x_{t+1}|x_t)] \leqslant f(x_t) + \langle \nabla f(x_t), \mathbb{E}(x_{t+1} - x_t|x_t) \rangle + \frac{L}{2}\mathbb{E}[\| x_{t+1} - x_t \|^2 |x_t] \tag{41}$$

$$= f(x) \tag{42}$$

$$+ \underbrace{\left\langle \nabla f(x_t), \mathbb{E}[-\frac{\gamma}{k}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n) + \frac{1}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n} z] \right\rangle}_{A} \tag{43}$$

$$+ \underbrace{\frac{L}{2}\mathbb{E}[\| -\frac{\gamma}{k}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n) + \frac{1}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n} z \|^2]}_{B} \tag{44}$$

For A, $\left\langle \nabla f(x_t), \mathbb{E}[\frac{1}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n} z|x_t] \right\rangle = 0$, and because of $\mathbb{E}[\mathbb{C}_n^t|x_t] = Kq_n$, the rest part of A can be bounded.

$$\left\langle \nabla f(x_t), \mathbb{E}[-\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n)] \right\rangle \tag{45}$$

$$= -\gamma\sum_{i=0}^{I-1}\mathbb{E}[\left\langle \nabla f(x_t), \sum_{n=1}^{N} P_n \nabla f_n(y_{t,i}^n) - \nabla f(x_t) + \nabla f(x_t) \right\rangle] \tag{46}$$

$$\leqslant \frac{\gamma I}{2}\mathbb{E} \| \nabla f(x_t) \|^2 + \frac{\gamma}{2}\sum_{i=0}^{I-1}\mathbb{E}[\sum_{n=1}^{N} P_n(\nabla f_n(x_t) - \nabla f_n(y_{t,i}^n))] \tag{47}$$

$$- \gamma I \mathbb{E}[\| \nabla f(x_t) \|^2] \tag{48}$$

$$\leqslant \frac{\gamma L^2 IN}{2}\sum_{n=1}^{N}\mathbb{E}[P_n^2 \| x_t - y_{t,i}^n \|^2] - \frac{\gamma I}{2}\mathbb{E}[\| \nabla f(x_t) \|^2] \tag{49}$$

$$\leqslant \frac{\gamma L^2 NI(I-1)}{2}\sum_{n=1}^{N} P_n^2 \sum_{j=0}^{i-1}\mathbb{E}[\| g_n(y_{t,j}^n) \|^2] - \frac{\gamma I}{2}\mathbb{E}[\| \nabla f(x_t) \|^2] \tag{50}$$

For B, we denote that $E_1 = -\frac{\gamma}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n}\sum_{i=0}^{I-1} g_n(y_{t,i}^n)$, $E_2 = \frac{1}{K}\sum_{n=1}^{N}\frac{\mathbb{C}_n^t P_n}{q_n} z$, Then $B = \frac{L}{2}(\parallel E_1 \parallel^2 + \parallel E_2 \parallel^2 + 2\langle E_1, E_2\rangle)$. For $\parallel E_1 \parallel^2$, we have

$$\mathbb{E}[\parallel E_1 \parallel^2] \leqslant \frac{\gamma^2 N}{K^2} \sum_{n=1}^{N} \mathbb{E}[\parallel \frac{\mathbb{C}_n^t P_n}{q_n} \sum_{i=0}^{I-1} g_n(y_{t,i}^n) \parallel^2] \tag{51}$$

$$\leqslant \frac{\gamma^2 N}{K^2} \sum_{n=1}^{N} \frac{\mathbb{E}[\mathbb{C}_n^2] P_n^2}{q_n^2} \sum_{i=0}^{I-1} \mathbb{E}[\parallel g_n(y_{t,i}^n) \parallel^2] \tag{52}$$

$$= \frac{\gamma^2 N}{K^2} \sum_{n=1}^{N} P_n^2 (K(K-1) + \frac{K}{q_n}) \sum_{i=0}^{I-1} \mathbb{E}[\parallel g_n(y_{t,i}^n) \parallel^2] \tag{53}$$

For $\parallel E_2 \parallel^2$, we have

$$\mathbb{E}[\parallel E_2 \parallel^2] \leqslant N \sum_{n=1}^{N} \mathbb{E}[\parallel \mathbb{C}_n^t z_i \parallel^2] \tag{54}$$

$$\leqslant N \sum_{n=1}^{N} (K(K-1)q_n^2 + Kq_n) S_n^2 \sigma_i^2 \tag{55}$$

$$\tag{56}$$

For $\langle E_1, E_2\rangle$, they are orthogonal, which means that $\langle E_1, E_2\rangle = 0$. Add them up then rearrange it, sum t from 0 to T-1, finally plug $\sigma = \frac{2Cs_i}{m\epsilon_i'} \sqrt{2\ln\left(\frac{1.25}{\delta_i'}\right)\left(KTq_i + \frac{\sqrt{KTq_i(1-q_i)}}{k}\ln\left(\frac{1}{\gamma_i'}\right)\right)}$ into it, we have the form of this theorem.

Corollary 2. *If Assumption 3 holds, $R < (log(1/\delta'''))^2/(q_iKk^2)$, plugging σ_i in Theorem 3, we have*

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\|\nabla f(\mathbf{x}_t)\|^2\right] = \mathcal{O}\left(\phi + \frac{LN\sqrt{R}}{\gamma E} \sum_{n=1}^{N} B_n \left(\sqrt{q_n} + \frac{1}{\sqrt{Kq_n}}\right)\right) \tag{57}$$

where $\phi = \frac{2(f(\mathbf{x}_0)-f^*)}{\gamma EKR} + \gamma^2 L^2 N(E-1)^2 \sum_{n=1}^{N} G_n^2 + \frac{\gamma LN}{K} \sum_{n=1}^{N} p_n^2 G_n^2 \theta_n$ and $B_n = dS^2 \ln(\frac{1}{\delta'}) ln(\frac{1}{\delta'''}) s_n^2 p_n^2 / \epsilon_n^2$.