A APPENDIX

A.1 Proof of Theorem 7.1

For the purpose of validation, we introduce an additional variable $\mathbf{v}_t^i$ to represent the immediate result of a single-step DPSGD update from $\mathbf{w}_t^i$. We interpret $\mathbf{w}_{t+1}^i$ as the parameter obtained after a single communication step. Consequently, the fair-clipping DPSGD in client i at iteration t transitions from Equation (14) to:

$$\mathbf{v}_{t+1}^i = \mathbf{w}_t^i - \frac{\eta}{|\mathcal{B}_i|}[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C \cdot \mathcal{N}(0, \mathbf{I})], \quad (28)$$

where:

$$C_t^{i,j} = \min\left(1 + \lambda \cdot \Delta_{loss}^j, \frac{C}{\|\nabla F_i(\mathbf{w}_t^i, \xi_j)\|}\right).$$

In our analysis, we define two virtual sequences $\mathbf{v}_t = \sum_{i=1}^N p_i \mathbf{v}_t^i$ and $\mathbf{w}_t = \sum_{i=1}^N p_i \mathbf{w}_t^i$, which is motivated by [40]. Therefore,

$$\mathbf{v}_{t+1} = \mathbf{w}_t - \sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C \cdot \mathcal{N}(0, \mathbf{I})] \quad (29)$$

A.1.1 *Key Lemma.*

LEMMA 2. *(Results of one iteration.)Assume Assumption 1-3 hold, we have:*

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

where:

- $A \triangleq G^2 C_t^3 + 3G^2 C_t^2 + 2L\Gamma C_t + \frac{2\sigma^2 C^2 d}{\hat{B}^2}$,
- $\hat{B} \triangleq \min_i |\mathcal{B}_i|$,
- $\Gamma \triangleq F^* - \sum_{i=1}^N p_i F_i^*$,
- $C_t \triangleq \sum_{i=1}^N p_i C_t^i, C_t^i \triangleq \frac{1}{|\mathcal{B}_i|} \sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}$.

Let $\Delta_t \triangleq \mathbb{E}\|\mathbf{w}_t - \mathbf{w}^*\|^2$. It is evident that we always have $\mathbf{w}_{t+1} = \mathbf{v}_{t+1}$. According to Lemma 2, this implies:

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

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

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.

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

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 $F(\cdot)$, let $\beta = \frac{2}{\mu}$, we get

$$\mathbb{E}[F(\mathbf{w}_t)] - F^* \leq \frac{L}{2}\Delta_t \leq \frac{L}{2t}v \leq \frac{L}{2t}\left(\frac{A}{\mu^2(2C_t - 1)} + \Delta_1\right)$$

A.1.2 *Proof of Lemma 2.*

By the Equation (29), we get

$$\begin{aligned}&\left\|\mathbf{v}_{t+1} - \mathbf{w}^*\right\|^2\\ &= \|\mathbf{w}_t - \mathbf{w}^*\\ &- \sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C \cdot \mathcal{N}(0, \mathbf{I})]\|^2\\ &= \left\|\mathbf{w}_t - \mathbf{w}^*\right\|^2\\ &\underbrace{-2\langle \mathbf{w}_t - \mathbf{w}^*, \sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C\mathcal{N}(0, \mathbf{I})]\rangle}_{\mathcal{A}_1}\\ &+ \underbrace{\left\|\sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C \cdot \mathcal{N}(0, \mathbf{I})]\right\|^2}_{\mathcal{A}_2}\end{aligned}$$

Firstly, we process $\mathcal{A}_2$:

$$\begin{aligned}\mathcal{A}_2 \leq &\underbrace{\left\|\sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|} \sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j)\right\|^2}_{\mathcal{B}_1}\\ &+ \underbrace{\sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}\left\langle \sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j), \sigma C \cdot \mathcal{N}(0, \mathbf{I})\right\rangle}_{\mathcal{B}_0}\\ &+ \underbrace{\|\sum_{i=1}^N p_i \frac{\eta}{|\mathcal{B}_i|}\sigma C \cdot \mathcal{N}(0, \mathbf{I})\|^2}_{\mathcal{B}_2}\end{aligned}$$

Since $\mathbb{E}[\mathcal{B}_0] = 0$, we focus on $\mathcal{B}_1$ and $\mathcal{B}_2$:

$$\mathbb{E}[\mathcal{B}_2] \leq \frac{\eta^2}{\hat{B}^2}\sum_{i=1}^N \mathbb{E}\|\sigma C\mathcal{N}(0, \mathbf{I})\|^2 \leq \frac{\eta^2\sigma^2 C^2 d}{\hat{B}^2},$$

where $\frac{1}{\hat{B}^2} \triangleq \max_i \frac{1}{|\mathcal{B}_i|}$.

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

$$\mathcal{B}_1 \leq \eta^2 \sum_{i=1}^N p_i \left\|\frac{1}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j)\right\|^2,$$

taking exception and according to assumption 3:

$$\mathbb{E}[\mathcal{B}_1] \leq \eta^2\mathbb{E}\left[\sum_{i=1}^N p_i \left\|C_t^i \nabla F_i(\mathbf{w}_t^i)\right\|^2\right] \leq \eta^2 C_t^2 G^2,$$

where $C_t \triangleq \sum_{i=1}^{N} p_i C_t^i$, $C_t^i \triangleq \frac{1}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}$.

Now, we obtain the bound for the expectation of $\mathcal{A}_2$:

$$\mathbb{E}[\mathcal{A}_2] \leq \eta^2 \left(\frac{\sigma^2 C^2 d}{\hat{B}^2} + G^2 C_t^2\right)$$

The process of $\mathcal{A}_1$ show as below:

$$\begin{aligned}
\mathcal{A}_1 = &-2\langle \mathbf{w}_t - \mathbf{w}^*, \sum_{i=1}^{N} p_i \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j)\rangle \\
&\underbrace{-2\langle \mathbf{w}_t - \mathbf{w}^*, \sum_{i=1}^{N} p_i \frac{\eta}{|\mathcal{B}_i|}\sigma C\mathcal{N}(0, \mathbf{I})\rangle}_{C_0} \\
= C_0 & \\
&\underbrace{-2\sum_{i=1}^{N} p_i \langle \mathbf{w}_t - \mathbf{w}_t^i, \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j)\rangle}_{C_1} \\
&\underbrace{-2\sum_{i=1}^{N} p_i \langle \mathbf{w}_t^i - \mathbf{w}^*, \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j)\rangle}_{C_2}
\end{aligned}$$

It's obvious that $\mathbb{E}[C_0] = 0$.

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

$$\begin{aligned}
C_1 &= -2\sum_{i=1}^{N} p_i \langle \mathbf{w}_t - \mathbf{w}_t^i, \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j)\rangle \\
&\leq \sum_{i=1}^{N} p_i \|\mathbf{w}_t - \mathbf{w}_t^i\|^2 + \|\sum_{i=1}^{N} p_i \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}\nabla F_i(\mathbf{w}_t^i, \xi_j)\|^2 \\
&= \sum_{i=1}^{N} p_i \|\mathbf{w}_t - \mathbf{w}_t^i\|^2 + \mathcal{B}_1
\end{aligned}$$

So we get

$$\mathbb{E}[C_1] \leq \sum_{i=1}^{N} p_i \mathbb{E}\|\mathbf{w}_t - \mathbf{w}_t^i\|^2 + \eta^2 G^2 C_t^2$$

According to Assumption 2, we know that

$$-\langle \mathbf{w}_t^i - \mathbf{w}^*, \nabla F_i(\mathbf{w}_t^i)\rangle \leq -\left(F_i(\mathbf{w}_t^i) - F_i(\mathbf{w}^*)\right) - \frac{\mu}{2}\|\mathbf{w}_t^i - \mathbf{w}^*\|^2$$

So we get

$$\begin{aligned}
C_2 \leq\ & 2\sum_{i=1}^{N} p_i \frac{\eta}{|\mathcal{B}_i|}\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j}. \\
&\left[-\left(F_i(\mathbf{w}_t^i, \xi_j) - F_i(\mathbf{w}^*)\right) - \frac{\mu}{2}\|\mathbf{w}_t^i - \mathbf{w}^*\|^2\right] \\
\leq\ & -2\eta C_t \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F_i(\mathbf{w}^*)\right) - \mu\eta C_t \sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 \\
=\ & -2\eta C_t \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F^* + F^* - F_i(\mathbf{w}^*)\right) \\
&- \mu\eta C_t \sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}^*\|^2 \\
=\ & \underbrace{-2\eta C_t \sum_{i=1}^{N} p_i \left(F_i(\mathbf{w}_t^i) - F^*\right)}_{\mathcal{D}_1} - 2\eta C_t \Gamma - \mu\eta C_t \|\mathbf{w}_t - \mathbf{w}^*\|^2,
\end{aligned}$$

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

Next, we proceed to handle $\mathcal{D}_1$.

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

where L-smooth inference as show:

$$\|\nabla F_i(\mathbf{w}_t^i)\|^2 \leq 2L\left(F_i(\mathbf{w}_t^i) - F_i^*\right). \tag{30}$$

Thus, we get

$$C_2 \leq 2\eta^2 C_t L\Gamma + C_t \sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}_t\|^2 - \mu\eta C_t \|\mathbf{w}_t - \mathbf{w}^*\|^2$$

To sum up,

$$
\begin{aligned}
\mathbb{E}[\mathcal{A}_1] &= \mathbb{E}\sum_{i=1}^{N} p_i \|\mathbf{w}_t - \mathbf{w}_t^i\|^2 + \eta^2 G^2 C_t^2 + 2\eta^2 C_t L\Gamma \\
&\quad + C_t \mathbb{E}\left[\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{w}_t\|^2\right] - \mu\eta C_t \mathbb{E}\|\mathbf{w}_t - \mathbf{w}^*\|^2 \\
&= (1 + C_t)\mathbb{E}\left[\sum_{i=1}^{N} p_i \|\mathbf{w}_t - \mathbf{w}_t^i\|^2\right] - \mu\eta C_t \mathbb{E}\|\mathbf{w}_t - \mathbf{w}^*\|^2 \\
&\quad + \eta^2 \left(G^2 C_t^2 + 2L\Gamma C_t\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{N} p_i \|\mathbf{w}_t - \mathbf{w}_t^i\|^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N} p_i \|(\mathbf{w}_t - \mathbf{v}_{t+1}^i) - (\mathbf{w}_t^i - \mathbf{v}_{t+1}^i)\|^2\right] \\
&\leq \mathbb{E}\left[\sum_{i=1}^{N} p_i \|\mathbf{w}_t^i - \mathbf{v}_{t+1}^i\|^2\right] \\
&\leq \mathbb{E}\left\|\sum_{i=1}^{N} p_i \frac{\eta}{|\mathcal{B}_i|}\left[\sum_{j=1}^{|\mathcal{B}_i|} C_t^{i,j} \cdot \nabla F_i(\mathbf{w}_t^i, \xi_j) + \sigma C \cdot \mathcal{N}(0, \mathbf{I})\right]\right\|^2 \\
&= \mathbb{E}\left[\mathcal{A}_2\right] = \eta^2 C_t^2 G^2 + \frac{\eta^2 \sigma^2 C^2 d}{\hat{B}^2}.
\end{aligned}
$$

So we get,

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{A}_1\right] &= (1 - \mu\eta C_t)\mathbb{E}\|\mathbf{w}_t - \mathbf{w}^*\|^2 + (1 + C_t)\eta^2 C_t^2 G^2 \\
&\quad + \eta^2 \left(2G^2 C_t^2 + 2L\Gamma C_t + 2\frac{\sigma^2 C^2 d}{\hat{B}^2}\right)
\end{aligned}
$$

All in all, we get

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

where

$$
A = G^2 C_t^3 + 3G^2 C_t^2 + 2L\Gamma C_t + \frac{2\sigma^2 C^2 d}{\hat{B}^2}.
$$

A.2 Proof of Theorem 6

THEOREM 6. *After T rounds local updates, the RDP of the $\mathbf{w}_T^i$ in i-th client satisfies:*

$$
R_{model}^i(\alpha) = \frac{T}{\alpha - 1}\sum_{k=0}^{\alpha}\binom{\alpha}{k}(1 - q)^{\alpha - i} q^k \exp\left(\frac{k^2 - k}{2\sigma^2}\right), \quad (31)
$$

where σ is noise multiplier of the w_{t+1}^i, and $\alpha > 1$ is the order.

proof. We will prove this in the following two steps: (i) use the RDP of the sampling Gaussian mechanism to calculate the privacy cost of each model update, and (ii) use the composition of RDP mechanisms to compute the privacy cost of multiple model updates.

DEFINITION 4. *(RDP privacy budget of SGM[32]). Let $SG_{q,\sigma}$, be the Sampled Gaussian Mechanism for some function f. If f has sensitivity 1, $SG_{q,\sigma}$ satisfies (α, R)-RDP whenever*

$$
R \leq \frac{1}{\alpha - 1}\log\max(A_\alpha(q, \sigma), B_\alpha(q, \sigma)) \quad (32)
$$

where

$$
\begin{cases}
A_\alpha(q, \sigma) \triangleq \mathbb{E}_{z\sim\vartheta_0}\left[(\vartheta(z)/\vartheta_0(z))^\alpha\right] \\
B_\alpha(q, \sigma) \triangleq \mathbb{E}_{z\sim\vartheta}\left[(\vartheta_0(z)/\vartheta(z))^\alpha\right]
\end{cases} \quad (33)
$$

with $\vartheta_0 \triangleq \mathcal{N}\left(0, \sigma^2\right), \vartheta_1 \triangleq \mathcal{N}\left(1, \sigma^2\right)$ and $\vartheta \triangleq (1 - q)\vartheta_0 + q\vartheta_1$

Further, it holds $\forall(q, \sigma) \in (0, 1], \mathbb{R}^{+*}, A_\alpha(q, \sigma) \geq B_\alpha(q, \sigma)$. Thus, $SG_{q,\sigma}$ satisfies $\left(\alpha, \frac{1}{\alpha-1}\log\left(A_\alpha(q, \sigma)\right)\right)$-RDP .

Finally, the existing work [32] describes a procedure to compute $A_\alpha(q, \sigma)$ depending on integer α.

$$
A_\alpha = \sum_{k=0}^{\alpha}\binom{\alpha}{k}(1 - q)^{\alpha - k} q^k \exp\left(\frac{k^2 - k}{2\sigma^2}\right) \quad (34)
$$

DEFINITION 5. *(Composition of RDP[31]). For two randomized mechanisms f, g such that f is (α, R_1)-RDP and g is (α, R_2)-RDP the composition of f and g which is defined as (X, Y)(a sequence of results), where $X \sim f$ and $Y \sim g$, satisfies $(\alpha, R_1 + R_2) - RDP$*

From Definition 4 and Definition 5, the Theorem 6 is obtained.