SUPPLEMENTARY MATERIAL: DIFFERENTIALLY PRIVATE ONE-BIT MODEL AGGREGATION IN PERSONALIZED FEDERATED LEARNING

Muhang Lan, Qing Ling, Song Xiao, and Wenyi Zhang

APPENDIX A
PROOF OF THEOREM 1

Proof: We prove each part of Theorem 1 as follows:

1) We first need to calculate two key statistics, including

$$\begin{aligned}\mathbb{E}\left[N_i\right] &= \sum_{m=1}^{M} \mathbb{P}\left(c_i^m = 1\right) \\ &= \sum_{m=1}^{M} \mathbb{E}\left[\mathbb{P}\left(c_i^m = 1 | \delta_i^m\right)\right] \\ &= \sum_{m=1}^{M} \mathbb{E}\left[\frac{b_i + \delta_i^m}{2b_i}\right] \\ &= \frac{M}{2}\left(1 + \frac{\mathbb{E}\left[\delta_i^m\right]}{b_i}\right) \\ &= \frac{M}{2}\left(1 + \frac{\theta_i}{b_i}\right),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left[N_i^2\right] \\ &= \mathbb{E}\left[\sum_{m=1}^{M} \mathbb{I}^2\left\{c_i^m = 1\right\}\right. \\ &\qquad \left.+ \sum_{i \neq j} \mathbb{I}\left\{c_i^m = 1\right\} \mathbb{I}\left\{c_j^m = 1\right\}\right] \\ &= \sum_{m=1}^{M} \mathbb{E}\left[\mathbb{I}\left\{c_i^m = 1\right\}\right] \\ &\qquad + \sum_{i \neq j} \mathbb{E}\left[\mathbb{I}\left\{c_i^m = 1\right\} \mathbb{I}\left\{c_j^m = 1\right\}\right] \\ &= M\mathbb{P}\left(c_i^m = 1\right) + M(M-1)\mathbb{E}^2\left[\mathbb{I}\left\{c_i^m = 1\right\}\right] \\ &= M\left(\frac{b_i + \theta_i}{2b_i}\right) + M(M-1)\left(\frac{b_i + \theta_i}{2b_i}\right)^2 \\ &= M\left(\frac{b_i + \theta_i}{2b_i}\right)\left(1 + (M-1)\frac{b_i + \theta_i}{2b_i}\right) \\ &= M\left(\frac{b_i + \theta_i}{2b_i}\right)\left(\frac{(M+1)\,b_i + (M-1)\,\theta_i}{2b_i}\right).\end{aligned}$$

By calculating the variance, we get

$$\begin{aligned}&\mathbb{E}\left[\left(\theta_i - \hat{\theta}_i\right)^2\right] \\ &= \mathbb{E}\left[\frac{2N_i - M}{M} b_i\right]^2 - \theta_i^2 \\ &= \left(\frac{b_i}{M}\right)^2 \mathbb{E}\left[4N_i^2 - 4N_i M + M^2\right] - \theta_i^2 \\ &= \left(\frac{b_i}{M}\right)^2 \cdot \left(4\mathbb{E}\left[N_i^2\right] - 4M\mathbb{E}\left[N_i\right] + M^2\right) - \theta_i^2 \\ &= \frac{b_i^2 - \theta_i^2}{M}. \qquad (14)\end{aligned}$$

From this, we can obtain the transmission error as

$$\begin{aligned}\mathbb{E}\left[\left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right\|^2\right] &= \mathbb{E}\left[\sum_{i=1}^{d}\left(\theta_i - \hat{\theta}_i\right)^2\right] \\ &= \frac{\sum_{i=1}^{d}\left(b_i^2 - \theta_i^2\right)}{M},\end{aligned}$$

where d is the length of the vector. ■

APPENDIX B
PROOF OF THEOREM 2

Proof of Theorem 2: According to the definition of privacy loss [8], we have

$$\begin{aligned}PL &= \ln \frac{\mathbb{P}\left(\boldsymbol{c}^m | \boldsymbol{\delta}^m + \boldsymbol{v}^m\right)}{\mathbb{P}\left(\boldsymbol{c}^m | \boldsymbol{\delta}^m\right)} \\ &= \sum_{i=1}^{d} \ln \frac{\mathbb{P}\left(c_i^m | \delta_i^m + v_i^m\right)}{\mathbb{P}\left(c_i^m | \delta_i^m\right)},\end{aligned}$$

where $\boldsymbol{c}^m$ represents the values after stochastic quantization. We further analyze the privacy loss for the i-th dimension. Consider the case when $c_i^m = 1$, we have

$$\begin{aligned}PL_i &= \ln \frac{\mathbb{P}\left(c_i^m = 1 | \delta_i^m + v_i^m\right)}{\mathbb{P}\left(c_i^m = 1 | \delta_i^m\right)} \\ &= \ln \frac{\left(b_i + \delta_i^m + v_i^m\right)/\left(2b_i\right)}{\left(b_i + \delta_i^m\right)/\left(2b_i\right)} \\ &= \ln\left(1 + \frac{v_i^m}{b_i + \delta_i^m}\right) \\ &\leq \frac{v_i^m}{b_i + \delta_i^m}.\end{aligned}$$

From the given condition on b_i, it holds

$$b_i \geq \max_m \left|\delta_i^m\right| + \left(1 + \frac{1}{\epsilon}\right)\Delta_1 \geq -\delta_i^m + \frac{\Delta_1}{\epsilon}.$$

By simple manipulation of the equation, we can obtain

$$\frac{1}{b_i + \delta_i^m} \leq \frac{\epsilon}{\Delta_1}.$$

Summing PL_i over all dimensions, we get

$$\sum_{i=1}^{d} PL_i \leq \sum_{i=1}^{d} \frac{\left|v_i^m\right|}{b_i + \delta_i^m} \leq \frac{\epsilon}{\Delta_1} \sum_{i=1}^{d} \left|v_i^m\right| \leq \epsilon,$$

where the last inequality follows from the l_1-sensitivity.

Similarly, when $c_i^m = -1$, the privacy loss is

$$PL_i = \ln \frac{\mathbb{P}\left(c_i^m = -1 | \delta_i^m\right)}{\mathbb{P}\left(c_i^m = -1 | \delta_i^m + v_i^m\right)}$$

$$= \ln \frac{(b_i - \delta_i^m)\,/\,(2b_i)}{(b_i - \delta_i^m - v_i^m)\,/\,(2b_i)}$$
$$= \ln\left(1 + \frac{v_i^m}{b_i - \delta_i^m - v_i^m}\right)$$
$$\leq \frac{v_i^m}{b_i - \delta_i^m - v_i^m}.$$

From the given condition on b_i, we have

$$b_i \geq \max_m |\delta_i^m| + \left(1 + \frac{1}{\epsilon}\right)\Delta_1 \geq \delta_i^m + v_i^m + \frac{\Delta_1}{\epsilon},$$

which holds

$$\frac{1}{b_i - \delta_i^m - v_i^m} \leq \frac{\epsilon}{\Delta_1}.$$

Summing PL_i over all dimensions, we get

$$\sum_{i=1}^{d} PL_i \leq \sum_{i=1}^{d} \frac{|v_i^m|}{b_i - \delta_i^m - v_i^m} \leq \frac{\epsilon}{\Delta_1}\sum_{i=1}^{d}|v_i^m| \leq \epsilon.$$

Combining both cases for $c_i^m = 1$ and $c_i^m = -1$, we have $PL = \sum_{i=1}^{d} PL_i \leq \epsilon$, thus the stochastic quantization mechanism satisfies $(\epsilon, 0)$-DP. ■

APPENDIX C
PROOFS OF LEMMA 1 AND LEMMA 2

Proof of Lemma 1: We define $\boldsymbol{e}_m^{t+1} = \nabla f_m(\boldsymbol{w}_m^{t+1}) + \lambda(\boldsymbol{w}_m^{t+1} - \boldsymbol{w}^t)$, then the client m's local training satisfies

$$\|\boldsymbol{e}_m^{t+1}\| \leq \gamma \|\nabla f_m(\boldsymbol{w}^t)\|, \tag{15}$$

which comes from the γ-inexact solution in Definition 2. Define $\bar{\boldsymbol{w}}^{t+1} = \mathbb{E}_m\left[\boldsymbol{w}_m^{t+1}\right] = \frac{1}{R}\sum_{m=1}^{R}\boldsymbol{w}_m^{t+1}$, we get

$$\bar{\boldsymbol{w}}^{t+1} - \boldsymbol{w}^t = \frac{-1}{\lambda}\mathbb{E}_m\left[\nabla f_m(\boldsymbol{w}_m^{t+1})\right] + \frac{1}{\lambda}\mathbb{E}_m\left[\boldsymbol{e}_m^{t+1}\right].$$

With $\bar{\lambda} = \lambda - L_- > 0$, we obtain

$$\begin{aligned}\nabla^2 h_m - \bar{\lambda}\boldsymbol{I} &= \nabla^2 f_m + \lambda \boldsymbol{I} - (\lambda - L_-)\boldsymbol{I}\\ &= \nabla^2 f_m + L_- \boldsymbol{I}\\ &\succeq \boldsymbol{0}, \end{aligned}\tag{16}$$

where (16) comes from the Assumption 2 and shows that h_m is $\bar{\lambda}$-strongly convex. Further define $\tilde{\boldsymbol{w}}_m^{t+1} = \arg\min_{\boldsymbol{w}} h_m(\boldsymbol{w}; \boldsymbol{w}^t)$, then we get

$$\begin{aligned}\bar{\lambda}\|\tilde{\boldsymbol{w}}_m^{t+1} - \boldsymbol{w}_m^{t+1}\| &\leq \|\nabla h(\tilde{\boldsymbol{w}}_m^{t+1}) - \nabla h(\boldsymbol{w}_m^{t+1})\|\\ &= \|\nabla h(\boldsymbol{w}_m^{t+1})\|\\ &= \|\boldsymbol{e}_m^{t+1}\|\\ &\leq \gamma\|\nabla f_m(\boldsymbol{w}^t)\|. \end{aligned}\tag{17}$$

Similarly, using the $\bar{\lambda}$-strong convexity of h_m, we obtain

$$\begin{aligned}\bar{\lambda}\|\tilde{\boldsymbol{w}}_m^{t+1} - \boldsymbol{w}^t\| &\leq \|\nabla h_m(\boldsymbol{w}^t)\|\\ &= \|\nabla f_m(\boldsymbol{w}^t) + \lambda(\boldsymbol{w}^t - \boldsymbol{w}^t)\|\\ &= \|\nabla f_m(\boldsymbol{w}^t)\|. \end{aligned}\tag{18}$$

By applying the triangle inequality to (17) and (18), we get

$$\|\boldsymbol{w}_m^{t+1} - \boldsymbol{w}^t\| \leq \frac{1+\gamma}{\bar{\lambda}}\|\nabla f_m(\boldsymbol{w}^t)\|. \tag{19}$$

From this, we can derive

$$\begin{aligned}\|\bar{\boldsymbol{w}}^{t+1} - \boldsymbol{w}^t\| &\leq \mathbb{E}_m[\|\boldsymbol{w}_m^{t+1} - \boldsymbol{w}^t\|] &(20)\\ &\leq \frac{1+\gamma}{\bar{\lambda}}\mathbb{E}_m[\|\nabla f_m(\boldsymbol{w}^t)\|]\\ &\leq \frac{1+\gamma}{\bar{\lambda}}\sqrt{\mathbb{E}_m[\|\nabla f_m(\boldsymbol{w}^t)\|^2]} &(21)\\ &\leq \frac{B(1+\gamma)}{\bar{\lambda}}\|\nabla F(\boldsymbol{w}^t)\|, &(22)\end{aligned}$$

where (20) follows from Jensen's inequality; (21) follows from the inequality $\sqrt{\mathbb{E}[A^2]} \geq \mathbb{E}[A]$; (22) follows from Assumption 1.

Further define $\boldsymbol{G}_{t+1}$ to describe $\bar{\boldsymbol{w}}^{t+1} - \boldsymbol{w}^t = \frac{-1}{\lambda}(\nabla F(\boldsymbol{w}^t) + \boldsymbol{G}_{t+1})$, then we have

$$\boldsymbol{G}_{t+1} = \mathbb{E}_m\left[\nabla f_m(\boldsymbol{w}_m^{t+1}) - \nabla f_m(\boldsymbol{w}^t) - \boldsymbol{e}_m^{t+1}\right],$$

whose upper bound satisfies

$$\begin{aligned}\|\boldsymbol{G}_{t+1}\| &\leq \mathbb{E}_m\left\|\left[\nabla f_m(\boldsymbol{w}_m^{t+1}) - \nabla f_m(\boldsymbol{w}^t) - \boldsymbol{e}_m^{t+1}\right]\right\|\\ &\leq \mathbb{E}_m\left[\|\nabla f_m(\boldsymbol{w}_m^{t+1}) - \nabla f_m(\boldsymbol{w}^t)\| + \|\boldsymbol{e}_m^{t+1}\|\right]\\ &\leq \mathbb{E}_m\left[L\|\boldsymbol{w}_m^{t+1} - \boldsymbol{w}_m^t\| + \|\boldsymbol{e}_m^{t+1}\|\right]\\ &\leq \left(\frac{L(1+\gamma)}{\bar{\lambda}} + \gamma\right) \times \mathbb{E}_m\left[\|\nabla f_m(\boldsymbol{w}^t)\|\right] &(23)\\ &\leq \left(\frac{L(1+\gamma)}{\bar{\lambda}} + \gamma\right) B\|\nabla F(\boldsymbol{w}^t)\|, &(24)\end{aligned}$$

where (23) follows from (19) and (15), and (24) is derived in the same manner as (22).

Since F is a convex combination of the local loss functions f_m, it is also L-Lipschitz smooth, which gives us

$$\begin{aligned}F(\bar{\boldsymbol{w}}^{t+1}) &\leq F(\boldsymbol{w}^t) + \langle \nabla F(\boldsymbol{w}^t), \bar{\boldsymbol{w}}^{t+1} - \boldsymbol{w}^t\rangle\\ &\quad + \frac{L}{2}\|\bar{\boldsymbol{w}}^{t+1} - \boldsymbol{w}^t\|^2 &(25)\\ &\leq F(\boldsymbol{w}^t) - \frac{1}{\lambda}\|\nabla F(\boldsymbol{w}^t)\|^2 - \frac{1}{\lambda}\langle \nabla F(\boldsymbol{w}^t), \boldsymbol{G}_{t+1}\rangle\\ &\quad + \frac{L(1+\gamma)^2B^2}{2\bar{\lambda}^2}\|\nabla F(\boldsymbol{w}^t)\|^2 &(26)\\ &\leq F(\boldsymbol{w}^t) - \left(\frac{1-\gamma B}{\lambda} - \frac{LB(1+\gamma)}{\bar{\lambda}\lambda}\right.\\ &\quad \left. - \frac{L(1+\gamma)^2B^2}{2\bar{\lambda}^2}\right) \times \left\|\nabla F(\boldsymbol{w}^t)\right\|^2, &(27)\end{aligned}$$

where substituting the definition of $\boldsymbol{G}_{t+1}$ and (22) into (25) yields (26), and (27) follows from the Cauchy-Schwarz inequality, i.e.,

$$\begin{aligned}&-\langle \nabla F(\boldsymbol{w}^t), \boldsymbol{G}_{t+1}\rangle\\ \leq\ &\left\|\langle \nabla F(\boldsymbol{w}^t), \boldsymbol{G}_{t+1}\rangle\right\|\\ \leq\ &\left\|\nabla F(\boldsymbol{w}^t)\right\|\|\boldsymbol{G}_{t+1}\|\\ \leq\ &\left\|\nabla F(\boldsymbol{w}^t)\right\|\left(\frac{L(1+\gamma)}{\bar{\lambda}} + \gamma\right)B\|\nabla F(\boldsymbol{w}^t)\|\\ =\ &\left(\frac{L(1+\gamma)}{\bar{\lambda}} + \gamma\right)B\|\nabla F(\boldsymbol{w}^t)\|^2.\end{aligned}$$

■

Proof of Lemma 2: Further extrapolating the result to a vector notation, we have

$$\mathbb{E}\left[\left\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right\|\right] \leq \sqrt{\mathbb{E}\left[\sum_{i=1}^{d}\left(\theta_i-\hat{\theta}_i\right)^2\right]} \tag{28}$$

$$= \sqrt{\sum_{i=1}^{d}\frac{b_i^2-\theta_i^2}{M}}$$

$$= \sqrt{\frac{1}{M}\sum_{i=1}^{d}\left(b_i+\theta_i\right)\left(b_i-\theta_i\right)}$$

$$\leq \sqrt{\frac{2}{M}\left(1+\frac{1}{\epsilon}\right)\Delta_1\sum_{i=1}^{d}b_i} \tag{29}$$

$$= \sqrt{\frac{2\|\boldsymbol{b}\|_1}{M}\left(1+\frac{1}{\epsilon}\right)\Delta_1},$$

where (28) follows from the Jensen's inequality; (29) follows from the DP requirement in Theorem 2. ■

Appendix D
Proof of Theorem 3

Proof of Theorem 3:

We utilize Lemma 1 and Lemma 2 to prove Theorem 3. Lemma 1 derives an upper bound of FL convergence rate with lossless FedAvg aggregation on regular clients. However, under the influence of the proposed transmission mechanism, we need to further consider the impact of quantized transmission and Byzantine attacks. The quantized aggregated parameter $\boldsymbol{w}^{t+1}$ and the lossless aggregated parameter $\bar{\boldsymbol{w}}^{t+1}$ satisfy

$$\begin{aligned} F(\boldsymbol{w}^{t+1})-F(\bar{\boldsymbol{w}}^{t+1}) &\leq L_0\|\boldsymbol{w}^{t+1}-\bar{\boldsymbol{w}}^{t+1}\| \\ &= L_0\|(\bar{\boldsymbol{w}}^{t+1}-\boldsymbol{w}^t)-(\boldsymbol{w}^{t+1}-\boldsymbol{w}^t)\| \\ &= L_0\|\boldsymbol{\theta}^t-\hat{\boldsymbol{\theta}}^t\|, \end{aligned} \tag{30}$$

where $\boldsymbol{\theta}^t=\bar{\boldsymbol{w}}^{t+1}-\boldsymbol{w}^t$ is the FedAvg aggregation result and $\hat{\boldsymbol{\theta}}^t$ is the one-bit aggregation result with Byzantine attacks and privacy protection. Substituting the result of Lemma 2 back into (30), we get

$$\mathbb{E}\left[F(\boldsymbol{w}^{t+1})\right]-\mathbb{E}\left[F(\bar{\boldsymbol{w}}^{t+1})\right] \leq L_0\sqrt{\frac{2\|\boldsymbol{b}\|_1}{M}\left(1+\frac{1}{\epsilon}\right)\Delta_1}. \tag{31}$$

Integrating (13) and (31) yields

$$\begin{aligned} \mathbb{E}\left[F(\boldsymbol{w}^{t+1})\right] &\leq \mathbb{E}\left[F(\boldsymbol{w}^t)\right]-\rho\left\|\nabla F(\boldsymbol{w}^t)\right\|^2 \\ &+ L_0\sqrt{\frac{2\|\boldsymbol{b}\|_1}{M}\left(1+\frac{1}{\epsilon}\right)\Delta_1}, \end{aligned} \tag{32}$$

where

$$\begin{aligned} \rho &\triangleq \frac{1-\gamma B}{\lambda}-\frac{LB(1+\gamma)}{\lambda\bar{\lambda}}-\frac{L(1+\gamma)^2B^2}{2\bar{\lambda}^2} \\ &= \frac{1}{\lambda}-\mathcal{O}\left(B^2\right). \end{aligned}$$

By selecting appropriate γ, λ, and M to ensure $\rho > 0$, it follows from (32) that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\left\|\nabla F(\boldsymbol{w}^t)\right\|^2 &\leq \frac{1}{\rho}\left(\frac{F(\boldsymbol{w}_0)-F^\star}{T}\right. \\ &\left.+ L_0\sqrt{\frac{2\|\boldsymbol{b}\|_1}{M}\left(1+\frac{1}{\epsilon}\right)\Delta_1}\right). \end{aligned}$$

■