A Proof of Lemma 3.7

As discussed in Section 2, the privacy loss occurred at $\tilde{h}^{(i)}$ is lower bounded by the privacy loss occurred at $\tilde{w}^{\text{global}}$. To keep our analysis general for all algorithms that fit in `PriFedSync`, we shall assume no knowledge of F_i and analyze $\tilde{w}^{\text{global}}$.

Without sampling, we can write the update of $\tilde{w}^{\text{global}}$ as:

$$(1-\eta)\tilde{w}^{\text{global}} + \frac{\eta}{m}\sum_{i\in[m]} \tilde{w}^{(i)}(S^{(i)}). \tag{A.1}$$

This is fully invertible function of $\tilde{w}^{(j)}$, so that the privacy loss is of the updated $\tilde{w}^{\text{global}}$ would be the same $\tilde{w}^{(j)}$, i.e.,

$$T\Big(\tilde{H}_i(\boldsymbol{S}), \tilde{H}_i(\boldsymbol{S}^{'j})\Big) = T\Big(\tilde{w}^{(j)}(S^{(j)}), \tilde{w}^{(j)}(S^{'(j)})\Big).$$

This lemma thus follows by the assumption.

B Proof of Lemma 3.8.

Let $\omega \in [0,1]^m$ be the indicator vector of the Possion sampling outcome: $\omega_i = 1$ if Client i is selected in synchronization, i.e. $i \in \Omega$. We use p_ω to denote the probability that ω appears, namely, $p_\omega = p^s(1-p)^{m-s}$ if ω has s nonzero entries.

Let $E = \{\omega : \omega_i = \omega_j = 1\}$ denote that event that both Client i and j are selected, and let E^c denote the complementary event that not both of them are selected. The output distribution of the subsampled algorithm $\tilde{H} \circ \mathtt{Sample}_p$ on dataset S can be written as a mixture model

$$\tilde{H}_i \circ \mathtt{Sample}_p(\boldsymbol{S}) = \sum_{\omega\in E} p_w P_\omega + \sum_{\omega\in E^c} p_w Q_\omega, \tag{B.1}$$

where we use P_ω to denote the output distribution associated with ω if $\omega \in E$, and use Q_ω for the other case. It is easy to see that P_ω depends on dataset $S^{(j)}$ but Q_ω does not. With the neighboring dataset $\boldsymbol{S}^{'j}$, the distribution $\tilde{H}_i \circ \mathtt{Sample}_p(S')$ can also be written as a mixture, yet only the components corresponding to cases where both i and j are selected will change. Specifically,

$$\tilde{H}_i \circ \mathtt{Sample}_p(\boldsymbol{S}^{'j}) = \sum_{\omega\in E} p_w P'_\omega + \sum_{\omega\in E^c} p_w Q_\omega. \tag{B.2}$$

The following technical lemma helps us bound the trade-off function between $\tilde{H}_i \circ \mathtt{Sample}_p(\boldsymbol{S})$ and $\tilde{H}_i \circ \mathtt{Sample}_p(\boldsymbol{S}^{'j})$.

Lemma B.1. *Let $\mathcal{F}$ be an event space and $\mathcal{F} = E \cup E^c$ is a valid partition of $\mathcal{F}$. Let ω denote an arbitrary event in $\mathcal{F}$, whose probability is p_w. We have $\sum_{w\in\mathcal{F}} p_w = 1$. For each event $\omega \in \mathcal{F}$, P_w, P'_w and Q_w are distributions reside on a common sample space. Consider two mixture distributions $A = \sum_{\omega\in E} p_\omega P_\omega + \sum_{\omega\in E^c} p_\omega Q_\omega$ and $B = \sum_{\omega\in E} p_\omega P'_\omega + \sum_{\omega\in E^c} p_\omega Q_\omega$. If there exists a trade-off function f such that $T(P_\omega, P'_\omega) \geq f$ for all ω, it holds that*

$$T(A,B)(\alpha) \geq \max\left\{f(\alpha), 1-\alpha-p_E\right\}.$$

Under the context of our problem, it holds that $\mathbb{P}(E) = p^2$ and $\mathbb{P}(E^c) = 1-p^2$ due to the independence of sampling Client i and j. Besides, for any fixed $\omega \in E$, using the same argument for Lemma 3.7, we have $T(P_\omega, P'_\omega) = T\big(\tilde{H}_i(\boldsymbol{S}_\Omega), \tilde{H}_i(\boldsymbol{S}^{'j}_\Omega)\big) \geq f_j$. This proofs our results.

B.1 Proof of Lemma B.1

Proof. Let $p_E = \mathbb{P}(w \in E)$. We can write

$$A = p_E \sum_{\omega\in E} p_{\omega|E} P_\omega + (1-p_E) \sum_{\omega\in E^c} p_{\omega|E^c} Q_\omega$$

and

$$B = p_E \sum_{\omega \in E} p_{\omega|E} P'_\omega + (1 - p_E) \sum_{\omega \in E^c} p_{\omega|E^2} Q_\omega.$$

Suppose a rejection rule ϕ achieves type I error α:

$$\alpha = \mathbb{E}_A[\phi] = p_E \sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P_\omega}[\phi] + (1 - p_E) \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi]. \tag{B.3}$$

The type II error of ϕ is

$$\begin{aligned}
1 - \mathbb{E}_B[\phi] &= 1 - p_E \sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P'_\omega}[\phi] - (1 - p_E) \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \\
&= 1 - p_E + p_E \left(1 - \sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P'_\omega}[\phi] \right) - (1 - p_E) \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \\
&= p_E \left(1 - \sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P'_\omega}[\phi] \right) + (1 - p_E) \left(1 - \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&= p_E \left(\sum_{\omega \in E} p_{\omega|E} \left(1 - \mathbb{E}_{P'_\omega}[\phi] \right) \right) + (1 - p_E) \left(1 - \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&\overset{(i)}{\geq} p_E \left(\sum_{\omega \in E} p_{\omega|E} f(\mathbb{E}_{P_\omega}[\phi]) \right) + (1 - p_E) \left(1 - \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&\overset{(ii)}{\geq} p_E \left(\sum_{\omega \in E} p_{\omega|E} f(\mathbb{E}_{P_\omega}[\phi]) \right) + (1 - p_E) f \left(\sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&\overset{(iii)}{\geq} p_E f \left(\sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P_\omega}[\phi] \right) + (1 - p_E) f \left(\sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&\overset{(iv)}{\geq} f \left(p_E \sum_{\omega \in E} p_{\omega|E} \, \mathbb{E}_{P_\omega}[\phi] + (1 - p_E) \sum_{\omega \in E^c} p_{\omega|E^c} \, \mathbb{E}_{Q_\omega}[\phi] \right) \\
&= f(\alpha),
\end{aligned} \tag{B.4}$$

where

(i) follows from the definition of the trade-off function: $T(P_\omega, P'_\omega) \geq f$ implies $1 - \mathbb{E}_{P'_\omega}[\phi] \geq f(\mathbb{E}_{P_\omega}[\phi])$,

(ii) follows from the property of trade-off functions: $f(\alpha) \leq 1 - \alpha, \forall \alpha \in [0, 1]$,

(iii) and (iv) follows from the Jensen's inequality for convex functions (f is convex).

It also holds that

$$
\begin{aligned}
1-\mathbb{E}_B[\phi] &= 1-p_E\sum_{\omega\in E}p_{\omega|E}\,\mathbb{E}_{P'_\omega}[\phi]-(1-p_E)\sum_{\omega\in E^c}p_{\omega|E^c}\,\mathbb{E}_{Q_\omega}[\phi]\\
&\overset{(v)}{=} 1-p_E\sum_{\omega\in E}p_{\omega|E}\,\mathbb{E}_{P'_\omega}[\phi]-\left\{\alpha-p_E\sum_{\omega\in E}p_{\omega|E}\,\mathbb{E}_{P_\omega}[\phi]\right\}\\
&= 1-\alpha-p_E+p_E\left\{1-\sum_{\omega\in E}p_{\omega|E}\,\mathbb{E}_{P'_\omega}[\phi]+\sum_{\omega\in E}p_{\omega|E}\,\mathbb{E}_{P_\omega}[\phi]\right\}\\
&= 1-\alpha-p_E+p_E\left\{\sum_{\omega\in E}p_{\omega|E}\left(1-\mathbb{E}_{P'_\omega}[\phi]+\mathbb{E}_{P_\omega}[\phi]\right)\right\}\\
&\overset{(vi)}{\geq} 1-\alpha-p_E+p_E\left\{\sum_{\omega\in E}p_{\omega|E}\big(1-\mathrm{TV}(P'_\omega,P_\omega)\big)\right\}\\
&\overset{(vii)}{\geq} 1-\alpha-p_E.
\end{aligned}
\tag{B.5}
$$

The equality (v) follows from Equation (B.3). For (vi) and (vii), consider the rejection rule ϕ for testing P_ω versus P'_ω. The type I error is $\alpha_\omega=\mathbb{E}_{P_\omega}[\phi]$ and type II error is $\beta_\omega=1-\mathbb{E}_{P'_\omega}[\phi]$. It is well known that

$$\alpha_\omega+\beta_\omega\geq 1-\mathrm{TV}(P_\omega,P'_\omega),$$

where $\mathrm{TV}(P_\omega,P'_\omega)$ is the total variation distance between P_ω and P'_ω, which takes value between 0 and 1. $\square$

C Proof of Theorem 1

Lemma 3.8 shows that for any $i\in[m]$,

$$T\big(\tilde{H}_i\circ\mathtt{Sample}_p(\boldsymbol{S}),\tilde{H}_i\circ\mathtt{Sample}_p(\boldsymbol{S}^{'j})\big)\geq g_{p,j}. \tag{C.1}$$

Recall that Equation 3.2 established the equivalence between $T\big((\tilde{H}_i\circ\mathtt{Sample}_p)^{\otimes R}(\boldsymbol{S}),(\tilde{H}_i\circ\mathtt{Sample}_p)^{\otimes R}(\boldsymbol{S}^{'j})\big)$ and $T\big(M_i(\boldsymbol{S}),M_i(\boldsymbol{S}^{'j})\big)$. By the composition theorem of f-differential privacy (Lemma 3.3), we have that for any $i\in[m]$,

$$T\big(M_i(\boldsymbol{S}),M_i(\boldsymbol{S}^{'j})\big) \;=\; T\big((\tilde{H}_i\circ\mathtt{Sample}_p)^{\otimes R}(\boldsymbol{S}),(\tilde{H}_i\circ\mathtt{Sample}_p)^{\otimes R}(\boldsymbol{S}^{'j})\big)\geq\; g_{p,j}^{\otimes R}. \tag{C.2}$$

The above result holds for a fixed Client j. Since the weak federated f-differential privacy notion (Definition 3.5) is defined for any pairs of i,j such that $i\neq j$, we need to take the "least private" trade-off function as our lower bound. That is $g_{p,j_{\min}}^{\otimes R}$, where $g_{p,j_{\min}}=\min\{g_{p,1},\ldots,g_{p,m}\}$.

Last, the strong federated privacy lower bound can be obtained by applying the composition theorem again:

$$T\left(\prod_{i\neq j}M_i(\boldsymbol{S}),\prod_{i\neq j}M_i(\boldsymbol{S}^{'j})\right)=\bigotimes_{i\neq j}T\big(M_i(\boldsymbol{S}),M_i(\boldsymbol{S}^{'j})\big)\geq g_{p,j_{\min}}^{\otimes(m-1)R}.$$

D Proof of Theorem 2

Let $g_{p,j}=\max(f_j,1-\alpha-p^2)$. By Theorem 1, it holds that

$$T\big(M_i(\boldsymbol{S}),M_i(\boldsymbol{S}^{'j})\big)\geq g_{p,j}^{\otimes R},\ i\in[m]. \tag{D.1}$$

We can apply the CLT type of result in Dong et al. (2019, Theorem 3.5) to obtain the asymptotic convergence of (D.1). Yet we found that taking the $1-\alpha-p^2$ component into account will give rise to a trade-off function that does not have an explicit form. Nonetheless, we can still lower bound

$$T\big(M_i(\boldsymbol{S}),M_i(\boldsymbol{S}^{'j})\big)\geq f_j^{\otimes R},\ i\in[m]. \tag{D.2}$$

We then utilize the following result from Dong et al. (2019) to obtain f_j.

Lemma D.1 (Dong et al. (2019))**.** *Algorithm 2 is $C_{B_j/n_j}(G_{1/\sigma_j})^{\otimes K}$-differentially private.*

Plugging $f_j = C_{B_j/n_j}(G_{1/\sigma_j})^{\otimes K}$ into Equation (D.2), we obtain

$$T\big(M_i(\boldsymbol{S}), M_i(\boldsymbol{S}'^{j})\big) \geq C_{B_j/n_j}(G_{1/\sigma_j})^{\otimes KR},\ i \in [m]. \tag{D.3}$$

The asymptotic convergence then follows from Corollary 5.4 of Dong et al. (2019): $C_{B_j/n_j}(G_{1/\sigma_j})^{\otimes KR} \to G_{\mu_j}$ if $\frac{B_j}{n_j}\sqrt{KR} \to c_j$ as $\sqrt{KR} \to \infty$ where

$$\mu_j = \sqrt{2}c_j\sqrt{e^{\sigma_j^{-2}}\Phi(1.5\sigma_j^{-1}) + 3\Phi(-0.5\sigma_j^{-1}) - 2}.$$

Similar to the argument for Theorem 1, we take the "least private" G_{μ_j} as the lower bound for the weak federated f-differential privacy notion, which is $G_{\mu_{\max}}$ with $\mu_{\max} = \max\{\mu_1, \ldots, \mu_m\}$. Likewise, the trade-off function for the strong federated privacy is $G_{\sqrt{m-1}\mu_{\max}}$.

E Additional Plots

E.1 Trade-off function $C_p(f)$

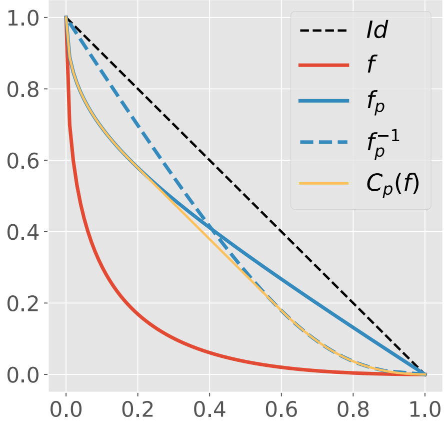


Figure E.1: The trade-off function $C_p(f)$ where $f = G_{1.8}$, $p = 0.35$.

Figure E.1 plots an example trade-off function $C_p(f)$ where f is a GDP trade-off function $G_{1.8}$, and the sampling rate $p = 0.35$.

E.2 Non-IID MNIST

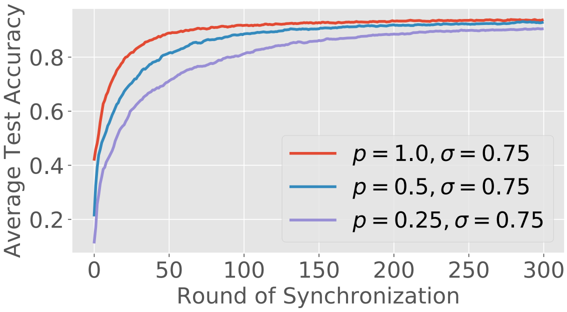
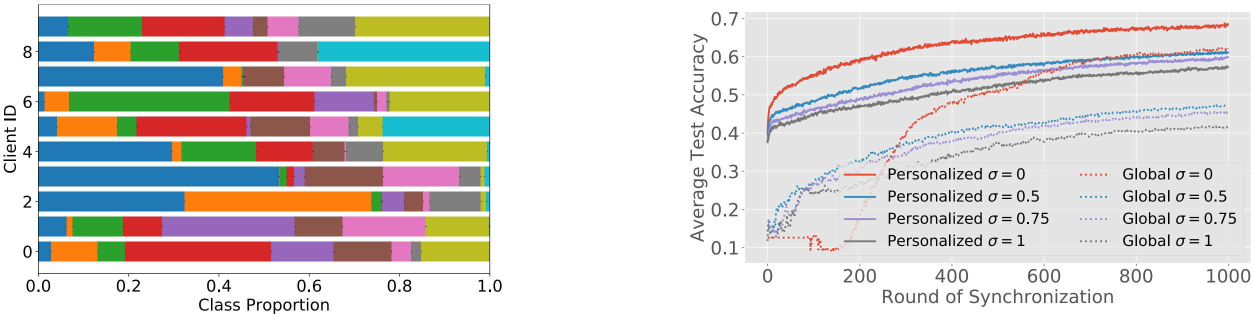


Figure E.2: MNIST experiment: A larger sampling rate leads to faster convergence.

Figure E.2 plots the average test accuracy versus the number of synchronization rounds for 3 runs with different client sampling rates in the MNIST epxeriment. It shows that the convergence is faster if we use a larger sampling rate. The noise level is set to $\sigma = 0.75$.

E.3 Non-IID CIFAR



(a)

(b)

Figure E.3: (a) The label class proportion for 10 randomly selected clients in the CIFAR-10 experiments. We use the Dirichlet prior with $\beta = 0.5$. (b) Average top-1 test accuracy vs synchronization rounds for the CIFAR-10 experiments. The client sampling rate is $p = 1$.

To illustrate the the heterogeneity of client data distributions, Figure E.3a plots the class proportion of the local data sets for 10 randomly selected clients. Figure E.3b plots the test accuracy curve for CIFAR-10 experiment when the client sampling rate is $p = 1$.