Supplementary Material: Topology-Aware Differential Privacy in Federated Learning

This supplementary material provides full proofs of Theorem V.2 and Theorem V.3 from the main paper. All equation, figure, and table numbering continues consecutively from the main paper. Lemmas introduced here are local to this document and are not referenced in the main text.

APPENDIX

A. Statement

Theorem (V.2 — Per-client conditional MI bound)**.** *Under (IA1) and (IA2) from the main paper, for any deterministic adversary strategy f, any prior $\mathbb{P} \in \mathcal{F}_{\mathcal{G},\omega}$, and any client i:*

$$I_{\mathbb{P}}(p_i;\, \hat{p}_i \mid \mathcal{G},\, \omega,\, \{\sigma_j\}) \;\le\; \underbrace{\frac{T_{\max}}{2\sigma_i^2\,|B|^2}}_{\textit{controllable}} \;+\; \underbrace{\ell_i^{\circ}}_{\textit{uncontrollable}}. \tag{6}$$

B. Proof Structure

The proof decomposes into four lemmas, assembled in Section D.

- Lemmas A.1 and A.4 isolate the *uncontrollable* lateral-leakage term via a chain-rule inequality and the definition of structural leverage.
- Lemmas A.2 and A.3 bound the *controllable* mechanism term using the Gaussian noise structure of DP-SGD.

Three standard tools are used throughout: the chain rule of mutual information and the data processing inequality; Mironov's RDP composition for the Gaussian mechanism [1]; and the Cuff-Yu max-KL to MI-DP conversion [2].

C. Lemmas

Lemma A.1 (Chain-rule decomposition)**.** *For any random variables X, Y, Z:*

$$I(X;\, Y) \;\le\; I(X;\, Z) \;+\; I(X;\, Y \mid Z).$$

Proof. By the MI conditioning identity:

$$I(X;Y) \;=\; I(X;Z) \;+\; I(X;Y \mid Z) \;-\; I(X;Z \mid Y).$$

The result follows since $I(X;Z \mid Y) \ge 0$. □

Lemma A.2 (Per-round Gaussian-mechanism KL bound)**.** *Under (IA1) and (IA2), for each round t, any D_{-i}, and any adjacent databases D_i, D_i':*

$$D_{\mathrm{KL}}\Big(\mathbb{P}_{\theta_i^{(t)} \mid D_i,\, \theta^{(t-1)},\, D_{-i}} \;\Big\|\; \mathbb{P}_{\theta_i^{(t)} \mid D_i',\, \theta^{(t-1)},\, D_{-i}}\Big) \;\le\; \frac{1}{2\sigma_i^2\,|B|^2}. \tag{7}$$

Consequently, by the Cuff-Yu conversion [2]:

$$I\Big(D_i;\, \theta_i^{(t)} \;\Big|\; \theta^{(t-1)},\, D_{-i}\Big) \;\le\; \frac{1}{2\sigma_i^2\,|B|^2}.$$

Proof. Conditional on $\theta^{(t-1)}$ and D_{-i}, the round-t update (Eq. (1) of the main paper) applies the Gaussian mechanism to the per-sample average of clipped gradients of D_i.

Sensitivity. Each clipped gradient has norm at most C, so the sensitivity of the averaged gradient is:

$$\Delta \;=\; \frac{C}{|B|}.$$

Noise scale. The per-round noise is $\xi_i^{(t)} \sim \mathcal{N}(0, \sigma_i^2 C^2 I)$, giving effective noise scale $s = \sigma_i C$ on the averaged gradient, where $\sigma_i > 0$ is the dimensionless multiplier of Abadi et al. [3].

KL bound. The KL divergence between adjacent mechanism outputs is:

$$\frac{\Delta^2}{2s^2} \;=\; \frac{(C/|B|)^2}{2(\sigma_i C)^2} \;=\; \frac{C^2/|B|^2}{2\sigma_i^2 C^2} \;=\; \frac{1}{2\sigma_i^2\,|B|^2},$$

establishing (7). The Cuff-Yu max-KL bound [2] then implies MI-DP with the same parameter, giving the stated MI inequality. □

Lemma A.3 (Sequential composition)**.** *Under (IA1) and (IA2), sequential composition of the per-round bound across all $T_{\max}$ rounds gives:*

$$\begin{aligned} I(D_i;\, \Theta_i \mid D_{-i}) \;&\le\; \sum_{t=1}^{T_{\max}} I\Big(D_i;\, \theta_i^{(t)} \;\Big|\; \theta^{(t-1)},\, D_{-i}\Big) \\ &\le\; \frac{T_{\max}}{2\sigma_i^2\,|B|^2}. \end{aligned}$$

Proof. Step 1: Chain rule. By the MI chain rule:

$$I(D_i;\Theta_i \mid D_{-i}) \;=\; \sum_{t=1}^{T_{\max}} I\Big(D_i;\, \theta_i^{(t)} \;\Big|\; \theta_i^{(<t)},\, D_{-i}\Big).$$

Step 2: Conditioning inequality. For each term, we claim:

$$I\Big(D_i;\, \theta_i^{(t)} \;\Big|\; \theta_i^{(<t)},\, D_{-i}\Big) \;\le\; I\Big(D_i;\, \theta_i^{(t)} \;\Big|\; \theta^{(t-1)},\, D_{-i}\Big).$$

To see this, note that conditional on D_{-i} and $\theta_i^{(<t)}$, the global state $\theta^{(t-1)}$ is a deterministic function of $\theta_i^{(<t)}$, $\Theta_{-i}^{(<t)}$, and the aggregation rule. Since $\Theta_{-i}^{(<t)}$ depends only on D_{-i} and noise $\{\xi_j^{(<t)}\}_{j\neq i}$, which is independent of D_i by (IA1) and (IA2), $\theta^{(t-1)}$ is conditionally independent of D_i given $(\theta_i^{(<t)}, D_{-i})$. The Markov chain

$$D_i \;\longrightarrow\; \theta_i^{(<t)} \;\longrightarrow\; \big(\theta_i^{(<t)},\, \theta^{(t-1)}\big)$$

and the data processing inequality therefore imply that conditioning additionally on $\theta^{(t-1)}$ cannot increase the MI.

Step 3: Apply Lemma A.2. Each term is bounded by $1/(2\sigma_i^2|B|^2)$. Summing over $t = 1, \ldots, T_{\max}$ completes the proof. □

Lemma A.4 (Lateral-leakage bound)**.** *Under (IA2):*

$$I(p_i;\ D_{-i}) \ \le\ \ell_i^{\circ}.$$

Proof. Immediate from Definition V.1 of the main paper, which defines ℓ_i° as $\sup_{\mathbb{P}\in\mathcal{F}_{\mathcal{G},\omega}} I_{\mathbb{P}}(p_i; D_{-i})$. □

D. Assembly

We now combine the four lemmas. Let $(\cdot)$ denote conditioning on $(\mathcal{G}, \omega, \{\sigma_j\})$ throughout; these are deployment constants, not random variables.

Step 1 (data processing). Since $\hat{p}_i = f(\Theta)$ is a deterministic function of the observations:

$$I(p_i;\ \hat{p}_i \mid \cdot) \ \le\ I(p_i;\ \Theta \mid \cdot).$$

Step 2 (chain rule). Applying Lemma A.1 with $X = p_i, Y = \Theta$, $Z = D_{-i}$:

$$I(p_i;\ \Theta \mid \cdot) \ \le\ I(p_i;\ D_{-i}) \ +\ I(p_i;\ \Theta \mid D_{-i}).$$

Step 3 (lateral term). By Lemma A.4:

$$I(p_i;\ D_{-i}) \ \le\ \ell_i^{\circ}.$$

Step 4 (independence of other clients). Under (IA1), Θ_{-i} is independent of D_i given D_{-i}. Therefore:

$$I(p_i;\ \Theta \mid D_{-i}) \ =\ I(p_i;\ \Theta_i \mid D_{-i}).$$

Step 5 (data processing on p_i). Since p_i is a deterministic function of D_i:

$$I(p_i;\ \Theta_i \mid D_{-i}) \ \le\ I(D_i;\ \Theta_i \mid D_{-i}).$$

Step 6 (composition bound). By Lemma A.3:

$$I(D_i;\ \Theta_i \mid D_{-i}) \ \le\ \frac{T_{\max}}{2\sigma_i^2\,|B|^2}.$$

Chaining Steps 1–6 gives:

$$I(p_i;\ \hat{p}_i \mid \cdot) \ \le\ \frac{T_{\max}}{2\sigma_i^2\,|B|^2} \ +\ \ell_i^{\circ},$$

which is the statement of Theorem V.2. □

E. Statement

Theorem (V.3 — Balanced min-max allocation)**.** *For any utility budget $U > 0$, the unique solution to:*

$$\min_{\{\sigma_i^2>0\}}\ \max_i \left[\frac{a}{\sigma_i^2} + \ell_i^{\circ}\right] \quad \textit{subject to} \quad \sum_{i=1}^{n}\sigma_i^2 \le U, \tag{8}$$

where $a := T_{\max}/(2|B|^2)$, is:

$$\sigma_i^{*\,2} \ =\ \frac{a}{K^{\star} - \ell_i^{\circ}},$$

where $K^{\star}$ is the unique solution to:

$$\sum_{i=1}^{n}\frac{a}{K^{\star} - \ell_i^{\circ}} \ =\ U, \qquad K^{\star} > \max_i \ell_i^{\circ}. \tag{9}$$

The achieved worst-case per-client MI bound is $K^{\star}$, equilibrated uniformly across all clients.

F. Proof Structure

Problem (8) has a non-smooth $\max$ objective. The proof proceeds in three steps.

1) *Reformulation.* Introduce a slack variable K to convert the $\max$ objective into a linear one, yielding a convex programme.
2) *Strong duality.* Verify Slater's condition to establish that KKT conditions are necessary and sufficient.
3) *KKT analysis.* Derive the closed-form solution and establish uniqueness of $K^{\star}$.

G. Convex Reformulation

Introduce slack variable K to rewrite (8):

$$\min_{K,\,\{\sigma_i^2>0\}} K \quad \text{subject to} \quad \frac{a}{\sigma_i^2} + \ell_i^{\circ} \le K\ \ \forall\, i, \quad \sum_{i=1}^{n}\sigma_i^2 \le U. \tag{10}$$

The objective is linear in K. The per-client constraints are convex since a/σ_i^2 is convex and strictly decreasing on $\sigma_i^2 > 0$. The budget constraint is linear.

Slater's condition. Setting $\sigma_i^2 = U/(2n)$ for all i is strictly feasible, with:

$$K \ =\ \frac{2na}{U} + \max_i \ell_i^{\circ} \ >\ \frac{a}{\sigma_i^2} + \ell_i^{\circ} \quad \forall\, i.$$

Strong duality therefore holds, and KKT conditions are necessary and sufficient for optimality.

H. KKT Analysis

Lemma A.5 (KKT optimality)**.** *The unique solution to* (10) *satisfies:*

$$\sigma_i^{*\,2} \ =\ \frac{a}{K^{\star} - \ell_i^{\circ}} \quad \forall\, i,$$

where $K^{\star}$ is the unique root of $g(K) = \sum_i a/(K - \ell_i^{\circ}) = U$ on $(\max_i \ell_i^{\circ},\ \infty)$.

Proof. The Lagrangian for (10) is:

$$\mathcal{L} \ =\ K \ +\ \sum_{i=1}^{n}\mu_i\left(\frac{a}{\sigma_i^2} + \ell_i^{\circ} - K\right) \ +\ \lambda\left(\sum_{i=1}^{n}\sigma_i^2 - U\right),$$

with multipliers $\mu_i \ge 0$ and $\lambda \ge 0$.

Stationarity in K.

$$\frac{\partial\mathcal{L}}{\partial K} \ =\ 1 - \sum_{i=1}^{n}\mu_i \ =\ 0 \ \implies\ \sum_{i=1}^{n}\mu_i \ =\ 1.$$

Stationarity in σ_i^2.

$$\frac{\partial\mathcal{L}}{\partial\sigma_i^2} \ =\ -\frac{\mu_i a}{\sigma_i^4} + \lambda \ =\ 0 \ \implies\ \sigma_i^{*\,2} \ =\ \sqrt{\frac{\mu_i a}{\lambda}}.$$

All constraints are active. Suppose the per-client constraint for some i were inactive. Then $\mu_i = 0$ by complementary slackness. Stationarity in σ_i^2 then requires $\lambda = 0$. But $\lambda = 0$ forces $\mu_j = 0$ for all j, contradicting $\sum_i \mu_i = 1$. Therefore all per-client constraints are active at the optimum:

$$\frac{a}{\sigma_i^{*\,2}} + \ell_i^{\circ} \ =\ K^{\star} \quad \forall\, i,$$

which gives the closed form $\sigma_i^{*\,2} = a/(K^\star - \ell_i^\circ)$.

Uniqueness of $K^\star$. Define $g(K) = \sum_i a/(K - \ell_i^\circ)$ on the domain $(K_{\min}, \infty)$ where $K_{\min} = \max_i \ell_i^\circ$. Then:

- g is continuous and strictly decreasing on $(K_{\min}, \infty)$,
- $g(K) \to +\infty$ as $K \downarrow K_{\min}$,
- $g(K) \to 0$ as $K \to +\infty$.

By the intermediate value theorem, for any $U > 0$ there exists a unique $K^\star \in (K_{\min}, \infty)$ with $g(K^\star) = U$. □

I. Strict Improvement over Uniform Allocation

Corollary (V.4 — Strict improvement over uniform allocation)**.** *Let $K_{\text{uniform}} = an/U + \max_i \ell_i^\circ$ denote the worst-case MI bound under uniform allocation $\sigma_i^2 = U/n$. Then:*

$$K^\star \;\le\; K_{\text{uniform}},$$

with equality if and only if all ℓ_i° are equal.

Proof. Uniform allocation $\sigma_i^2 = U/n$ for all i is feasible for (10) and achieves objective value K_{uniform}. Since $K^\star$ is the minimum of (10), we have $K^\star \le K_{\text{uniform}}$. Equality holds if and only if uniform allocation satisfies the KKT conditions of Lemma A.5, which requires:

$$\frac{a}{U/n} + \ell_i^\circ \;\equiv\; K^\star \quad \forall\, i.$$

This holds if and only if all ℓ_i° are equal. □

J. Edge Cases and Computational Note

Small utility budget. The KKT solution requires $K^\star > \max_i \ell_i^\circ$, which holds for all $U > 0$ by Lemma A.5. However, when U is very small, the solution concentrates the budget on high-leverage clients ($\sigma_i^2 \to 0$ for low-leverage clients), falling below the practically deployable range for DP-SGD. This regime is not analysed further.

Computation. Solving $g(K^\star) = U$ via one-dimensional bisection on K requires $O(n \log(1/\varepsilon))$ operations for tolerance ε, which is negligible for $n \le 500$.

REFERENCES

[1] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 263–275.

[2] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 43–54.

[3] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.