Unified Privacy Guarantees for Decentralized Learning via Matrix Factorization

Aurélien Bellet[1] Edwige Cyffers[2] Davide Frey[3] Romaric Gaudel[3] Dimitri Lerévérend[3] François Taïani[3]

[1]Inria, Université de Montpellier, INSERM, Montpellier, France
[2]Institute of Science and Technology Austria, Klosterneuburg, Austria
[3]Université de Rennes, Inria, CNRS, IRISA, Rennes, France

Decentralized Learning

- Fixed undirected and connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$ nodes
- Gossip matrix $W \in [0,1]^{n\times n}$ over the graph $\mathcal{G}$ is a doubly stochastic matrix ($W\mathbb{1} = W^\top\mathbb{1} = \mathbb{1}$) with $W_{uv} > 0$ if and only if there exists an edge between u and v
- For Decentralized Gradient Descent, nodes aim to optimize an objective function of the form $f(\theta) = \sum_{v=1}^{n} f_v(\theta, D_v)$ where θ represents the parameters of the model and L is some differentiable loss and D_v the local dataset of node v. Let θ_v^t be an arbitrary initialization of the parameters at each node v. We denote the local gradient of node v at iteration t (scaled by η) by $g_t^v = \nabla f_v(\theta_t^v, x_v)$. We note G_t the stacked gradient across all nodes. Then, D-SGD can be written as:

$$\text{Gradient update: } \theta_{t+\frac{1}{2}} = \theta_t - \eta G_t, \qquad \text{Gossiping step: } \theta_{t+1} = W\theta_{t+\frac{1}{2}}.$$

Privacy in Decentralized Learning

- D-SGD is vulnerable to privacy attacks. In [2], a node can reconstruct datapoints from other nodes, even if they are far away from the attackers in the graph
- When differential privacy mechanisms are applied at nodes' level, privacy guarantees are amplified by decentralization as noise accumulate [1].
- No existing analysis able to take into account the noise correlation between nodes

Differential Privacy in Central Setting

A randomized mechanism $\mathcal{M}$ satisfies *μ-Gaussian Differential Privacy* (μ-GDP) if, for any neighboring datasets $\mathcal{D} \simeq \mathcal{D}'$, there exists a (possibly randomized) function h such that

$$h(Z) \overset{d}{=} \mathcal{M}(\mathcal{D}), \quad Z \sim \mathcal{N}(0,1), \qquad h(Z') \overset{d}{=} \mathcal{M}(\mathcal{D}'), \quad Z' \sim \mathcal{N}(\mu,1),$$

where $\overset{d}{=}$ denotes equality in distribution.

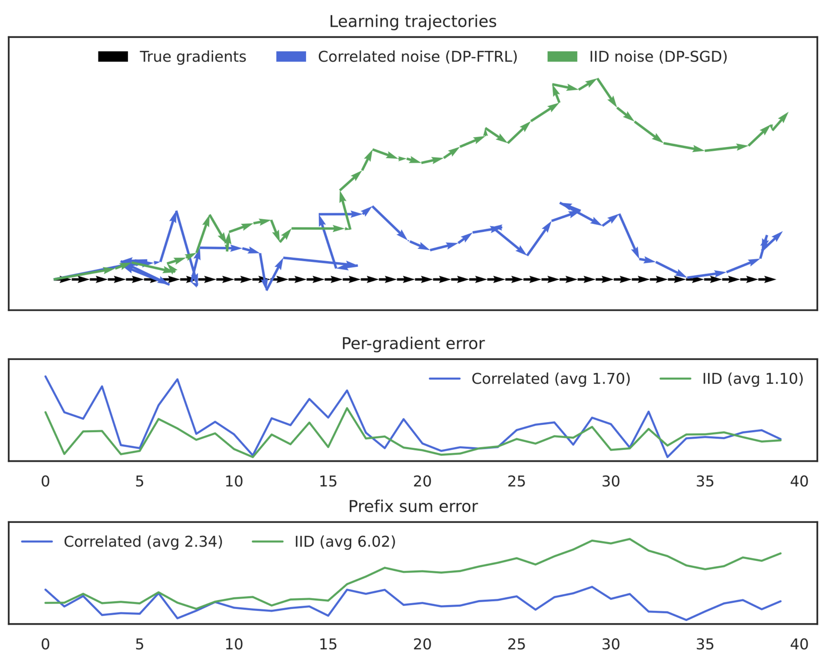
DP is achieved by clipping and adding Gaussian noise to the gradients.

- Noise **accumulate** through iterations and using correlated noise decrease the overall amount of noise injected in the system
- For $A_{ij}^{\text{pre}} = 1_{i\geq j}$, DP-SGD is:

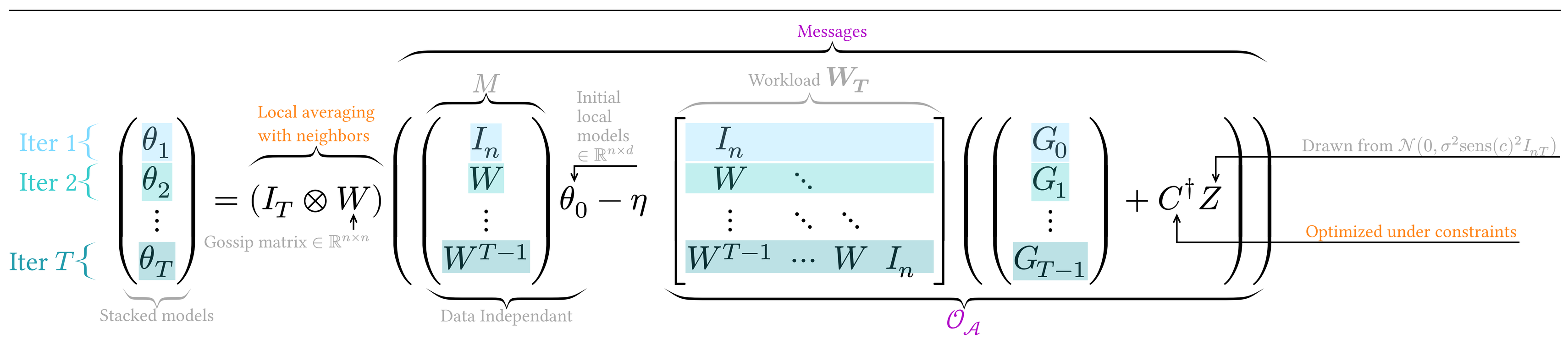
$$\theta = 1_T \otimes \theta_0 - (A^{\text{pre}}G + Z), \quad G \in \mathbb{R}^{T\times d}$$

- For any factorization $A^{\text{pre}} = BC$, one can rewrite $A^{\text{pre}}G + BZ$ as $A^{\text{pre}}(G + C^\dagger Z)$
- Goal: minimize $\text{sens}(C)^2\|B\|^2$ with

$$\text{sens}_\Pi(C) = \max_{G \simeq_\Pi G'} \|C(G - G')\|_F .$$



Framing MF-D-SGD as a MF problem



- Gradient step: $\theta_{t+1/2} = \theta_t - \eta\,(G_t + C_t^\dagger Z)$, and Gossiping step: $\theta_{t+1} = W\theta_{t+1/2}$
- Whole algorithm can then be summarized as

$$\theta \;=\; (I_T \otimes W)\left(M\theta_0 - \eta\,\mathbf{W}T(G + C^\dagger Z)\right)$$

Correlated noise in D-SGD

Algorithm 1: MF-D-SGD: Matrix Factorization Decentralized SGD

Inputs: $W \in \mathbb{R}^{n\times n}$, C, T, Δ_g, σ, $\theta_0 \in \mathbb{R}^{n\times d}$, $Z \sim \mathcal{N}(0,\, \Delta_g^2\sigma^2)^{nT\times d}$

forall *node u in parallel* **do**
  for $t = 1$ ***to*** T **do**
    $g_t^{(u)} \leftarrow \text{clip}[\nabla f_u(\theta_t^{(u)}, \xi_t^{(u)}),\, \Delta_g]$ with $\xi_t^{(u)} \sim \mathcal{D}_u$`// Clipped gradient`
    $\theta_{t+\frac{1}{2}}^{(u)} \leftarrow \theta_t^{(u)} - \eta(g_t^{(u)} + (C^\dagger Z)_{[nt+u]})$`// Local update`
    Send $\theta_{t+\frac{1}{2}}^{(u)}$ to all neighbors $v \in \Gamma_u$;
    Receive $\theta_{t+\frac{1}{2}}^{(v)}$ from all neighbors $v \in \Gamma_u$;
    $\theta_{t+1}^{(u)} \leftarrow \sum_{v\in\Gamma_u} W_{[u,v]}\theta_{t+\frac{1}{2}}^{(v)}$ `// Local average`

return $\theta_{T+1}^{(u)}$, $\forall u \in \{1,\ldots,n\}$

Privacy Guarantees

Theorem

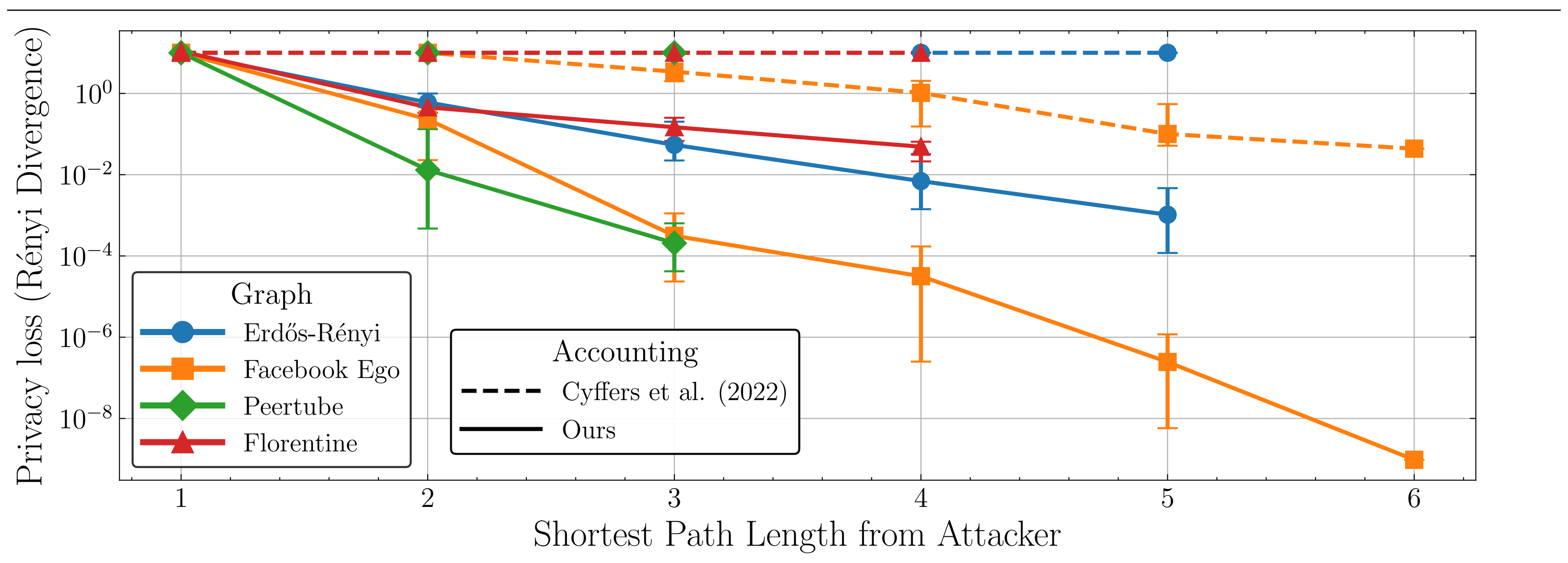
Let $\mathcal{O}_{\mathcal{A}} = AG + BZ$ be the attacker knowledge of a trust model , and denote $\mathcal{M}(G)$ the corresponding mechanism. Let Π be a participation scheme for G. For $Z \sim \mathcal{N}(0,\nu^2)^{m\times d}$, when A is a column-echelon matrix and there exists some matrix C such that $A = BC$ with

$$\nu = \sigma\,\text{sens}_\Pi(C;B) \quad \text{with} \quad \text{sens}_\Pi(C;B) \leq \max_{\pi\in\Pi} \sum_{s,t\in\pi} \left|\left(C^\top B^\dagger BC\right)_{s,t}\right|,$$

then $\mathcal{M}$ is $\frac{1}{\sigma} - GDP$, even when G is chosen adaptively.

- Now we have a dependency in B because the workload matrix is not squared
- This is a generalization of previous results.
- Hold for more general algorithms than MF-D-SGD
- Captures different trust models: LDP, PNDP [1], SecLDP.

Better accounting



MAFALDA

MAtrix FActorization for Local Differential PrivAte-SGD (MAFALDA-SGD)

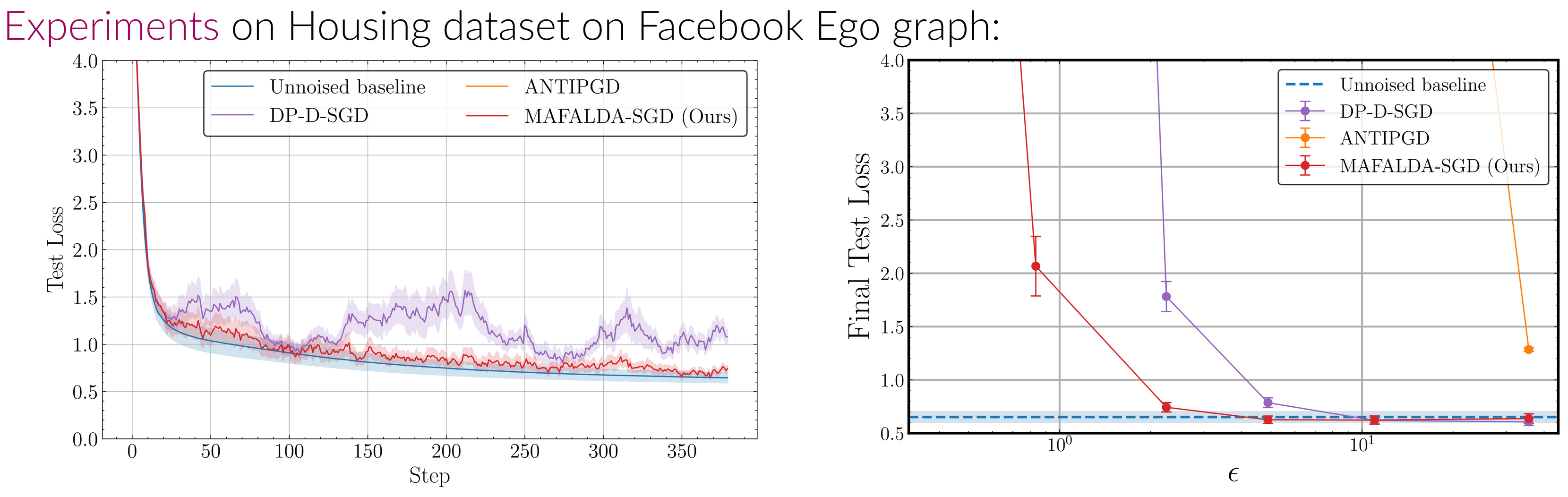
- For Local DP, nodes can only rely on their own noise for protection
- For computational reasons, we want the same noise pattern for all nodes.
- The new minimization problem is

$$\mathcal{L}_{\text{opti}}\left(\mathbf{W}_T, C_{\text{local}}\right) = \underset{\Pi_{\text{local}}}{\text{sens}}\left(C_{\text{local}}\right)^2 \left\|LC_{\text{local}}^\dagger\right\|_F^2$$

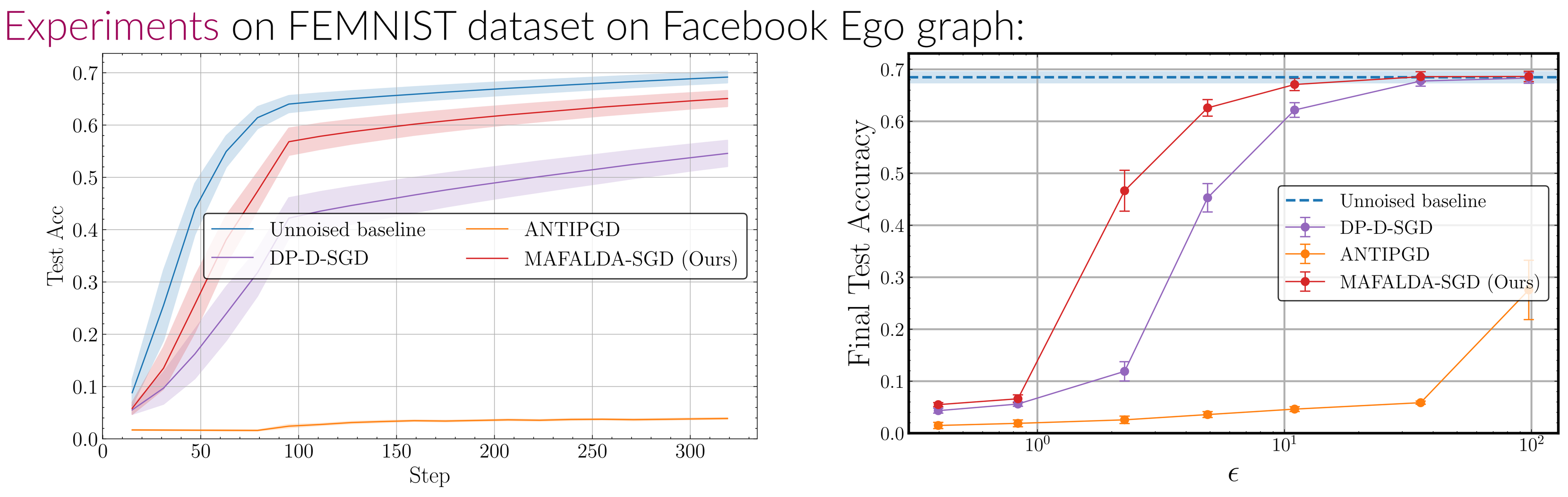
with L the Choleski decomposition of

$$\sum_{i=1}^{n} \left[(I_T \otimes W)\,\mathbf{W}_T\mathbf{K}^{(T,n)}\right]^\top_{[:,iT:(i+1)T-1]} \left[(I_T \otimes W)\,\mathbf{W}_T\mathbf{K}^{(T,n)}\right]_{[:,iT:(i+1)T-1]}$$

Experiments on Housing dataset on Facebook Ego graph:



Experiments on FEMNIST dataset on Facebook Ego graph:



References

[1] Edwige Cyffers, Mathieu Even, Aurélien Bellet, and Laurent Massoulié. Muffliato: Peer-to-Peer Privacy Amplification for Decentralized Optimization and Averaging. NeurIPS, 2022.

[2] Abdellah El Mrini, Edwige Cyffers, and Aurélien Bellet. Privacy attacks in decentralized learning. ICML'24. JMLR.org, 2024.

[3] Krishna Pillutla, Jalaj Upadhyay, Christopher A. Choquette-Choo, Krishnamurthy Dvijotham, Arun Ganesh, Monika Henzinger, Jonathan Katz, Ryan McKenna, H. Brendan McMahan, Keith Rush, Thomas Steinke, and Abhradeep Thakurta. Correlated noise mechanisms for differentially private learning, 2025.

dimitri.lereverend@inria.fr