Low-Cost Privacy-Aware decentralized learning.

Sayan Biswas [1] Davide Frey [2] Romaric Gaudel [2]
Anne-Marie Kermarrec [1] Dimitri Lerévérend [2] Rafael Pires [1]
Rishi Sharma [1] François Taïani [2]

[1]EPFL [2]Univ Rennes, Inria, CNRS, IRISA

Motivations

- **Decentralized learning** - No central server, different from Federated Learning.
- **Local data** - non-uniform repartition across nodes.
- Collaborative learning via **model sharing** among peers.
- Models convey information about data
  - **Privacy guarantees**
  - Utility-privacy tradeoff.
- Objective: **Private** model learning through model exchanges.

1 - State of the art

- **Noisy** approaches with **Local Differential Privacy**. (Duchi et al. 2013)
  - **Global** notion of privacy.
- Pairwise network differential privacy (Cyffers et al. 2022)
  - Relaxation of **Local Differential Privacy**.
  - **Pairwise** notion of privacy: node to node instead of node to world.
  - Account for **node distance** for privacy loss: more information leaked to close neighbors.
  - **No composition theorem** when considering 1 averaging round/gradient.
  - Multiple **communication rounds**

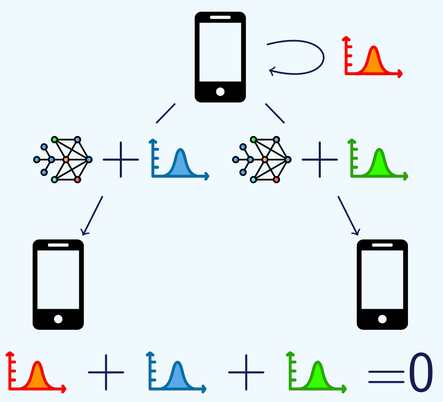
2 - Zip-DL

- **Correlated noises** to reduce individual impact on the global system.
- 1 averaging round/gradient descent.

Zip-DL-averaging for a node a.

Input: local model x_a, stepsize γ, privacy parameter ς_a.
Output: Localized model average with correlated noise.

1: Get the gossip weights W_a, $d_a \leftarrow |\Gamma_a|$
2: Draw $Y_{a\to v} \sim \mathcal{N}(0, \gamma^2\varsigma_a^2)$ for $v \in \Gamma_a$
3: $Z_{a\to v} = Y_{a\to v} - \frac{1}{d_a W_{a,v}} \sum_{j\in\Gamma_a} W_{a,j} Y_{a\to j}$
4: **for all** $v \in \Gamma_a$ **do**
5: Send $x_a + Z_{a\to v}$ to v
6: Receive $x_v + Z_{v\to a}$ from v
7: **end for**
8: **return** $\sum_{v\in\Gamma_a} W_{a,v}(x_v + Z_{v\to a})$



3 - Theoretical guarantees.

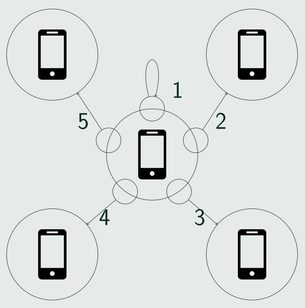
- Formal **privacy guarantees**.
- State of the art compatible **convergence rate**.

Equivalent system formulation

- Necessary to overcome **composition theorems**.
- n^2 noises vs n nodes.
- Communication across different nodes $\hat{W}$.
- Internal communication between virtual nodes $\hat{L}$.

Translates to a simplified update rule:

$$\hat{X}^{(t+1)} = \hat{L}\hat{W}\left(\hat{X}^{(t+1/2)} + \hat{Z}^{(t)}\right)$$



Theorem 1: Convergence rate

For any number of iterations T, there exists a constant stepsize γ s.t. it holds that $\frac{1}{2W_T}\sum_{t=0}^{T} w_t(\mathbb{E}\left[f(\bar{x}^{(t)})\right] - f^*) + \frac{\mu}{2}r_{T+1}$ is bounded by:

$$\mathcal{O}\left(\frac{\bar{\omega}^2}{n\mu T} + \frac{LA'}{\mu^2T^2} + \frac{r_0 L}{p}\exp\left[-\frac{\mu p(T+1)}{192\sqrt{3}L}\right]\right)$$

where $f^* = f(x^*)$, $r_t = \mathbb{E}\left[\left\|\bar{x}^{(t)} - x^*\right\|^2\right]$, $w_t = (1-\frac{\mu}{2}\gamma)^{-(t+1)}$, $W_T = \frac{1}{T}\sum_{t=1}^{T} w_t$ and $A' = \frac{16-4p}{2(16-7p)}(\bar{\omega}^2 + \frac{18}{p}\bar{\vartheta}^2) + \frac{d}{n}\frac{16-4p}{16-7p}\sum_{a,v=1}^{n} d_a\frac{(d_v-1)^2}{d_v}\varsigma_v^2$.

Theorem 2: Privacy guarantees

T iterations of Zip-DL *satisfies $\left(\alpha, \epsilon^{(T)}(a,v)\right)$-PNDP, where $\epsilon^{(T)}(a,v)$ is bounded for any two nodes $a, v \in \mathcal{V}$ by:*

$$\frac{2\alpha\gamma^2\Delta^2}{L+4\gamma^2L^2}\sum_{t=0}^{T-1}\sum_{\substack{\hat{v}\in\hat{V}\\ \hat{w}\in\hat{\Gamma}_{\hat{v}}^{(t)}}}\frac{(2+4\gamma^2L)^t - 1}{\left(\left(\tilde{W}\tilde{C}\right)^{(t)}\tilde{\Sigma}_{\tilde{Y}^{(t)}}\left(\tilde{W}\tilde{C}\right)^{(t)}\right)_{\tilde{w},\tilde{w}}},$$

where $\tilde{\Sigma}_{\tilde{Y}^{(t)}}$ is a diagonal matrix representing the noise variances of all noises generated by the algorithm up to time T, $\tilde{C}^{(t)}$ is a block-diagonal matrix representing the correlation factor at each iteration t, and $\tilde{W}^{(t)}$ is the accumulation of all the powers of the gossip matrix.

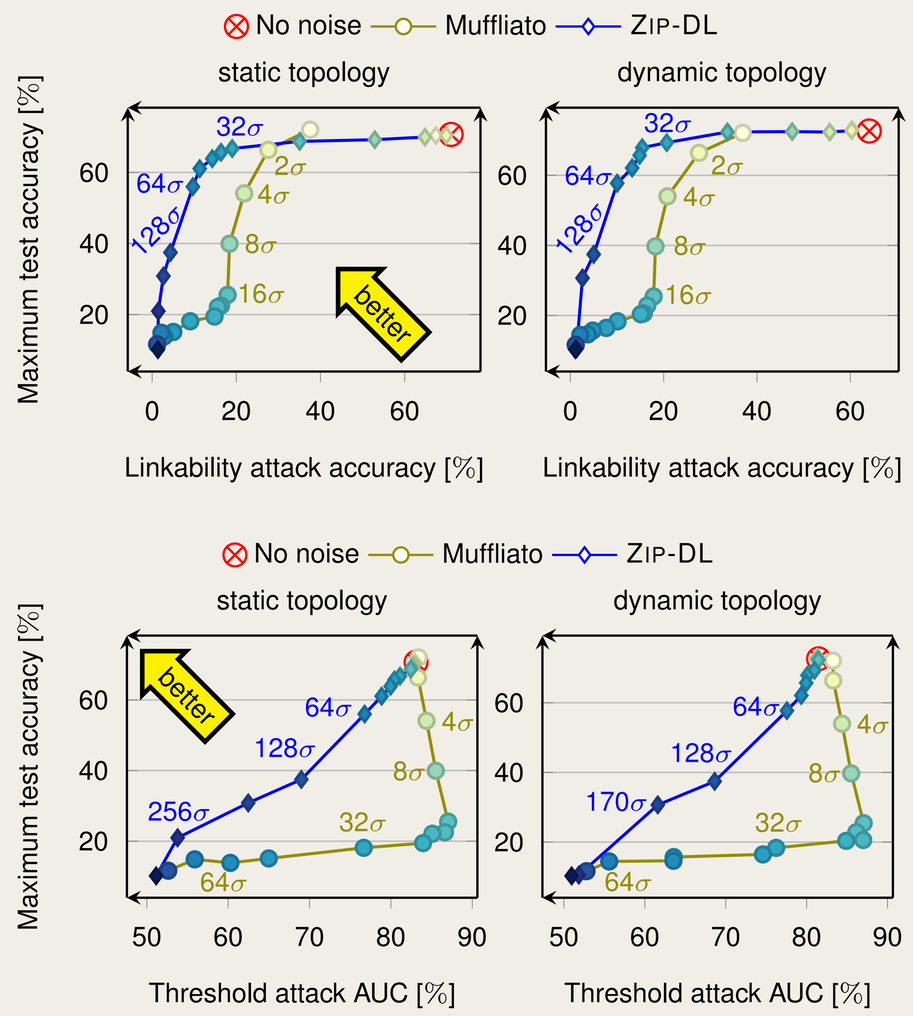
4 - Experimental results

- Decentralized **simulations** on 6-regular graphs with 128 nodes.
- $\mathrm{CIFAR10}$ partitionned in a **Non-Identically Distributed** manner.
- Compared to:
  - Decentralized Learning (DL).
  - Muffliato (Cyffers et al. 2022).
- Multiple noise levels σ.
- Several metrics gathered:
  - **Test accuracy** (utility) and **packets** sent over the network.
  - **Attacks accuracy** on models shared on the network.

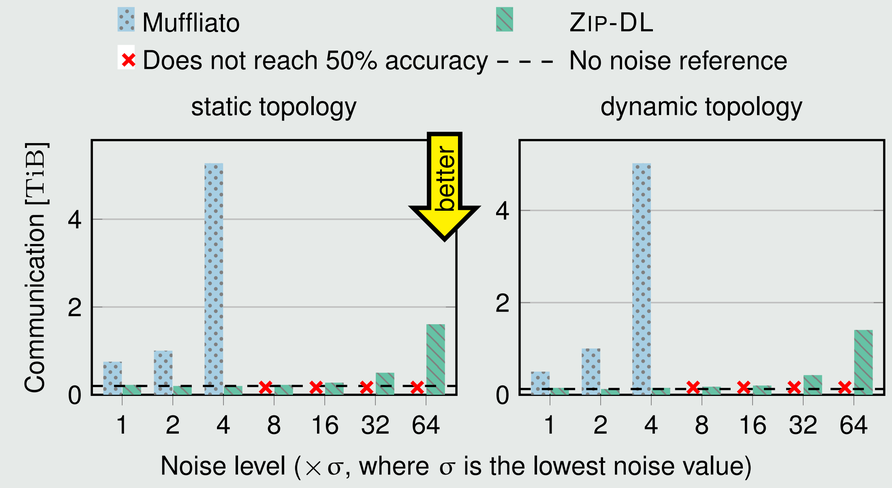
Privacy attacks

- Observation of the **loss function** over specific data samples.
- Two attacks evaluated:
  - **Linkability attack** - links the loss over a data to find the owner.
  - **Threshold attack** - Simple Membership Inference Attack with a threshold parameter.

Experimental utility-privacy tradeoff.



Communication costs



References

Cyffers, E., M. Even, A. Bellet, and L. Massoulié (2022). "Muffliato: Peer-to-Peer Privacy Amplification for Decentralized Optimization and Averaging". In: *NEURIPS*.

Koloskova, A., N. Loizou, S. Boreiri, M. Jaggi, and S. Stich (2020). "A Unified Theory of Decentralized SGD with Changing Topology and Local Updates". In: PMLR.

Duchi, J. C., M. I. Jordan, and M. J. Wainwright (2013). "Local Privacy and Statistical Minimax Rates". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*.

Contact informations

dimitri.lereverend@inria.fr. Website: Paper: