Noiseless Privacy-Preserving Decentralized Learning*

* To appear in the Proceedings on Privacy Enhancing Technologies (PoPETS) 2025

Full paper!

Sayan Biswas†, Mathieu Even‡, Anne-Marie Kermarrec†, Laurent Massoulié‡, Rafael Pires†, Rishi Sharma†, Martijn de Vos†

† SaCS Lab, EPFL, Switzerland. ‡ Inria, DI ENS, PSL University, France. Correspondence: <first name>.<last name>@epfl.ch

Motivation

- DL is becoming popular as it addresses several issues (*e.g.*, single point of failure, scalability) that centralized ML or FL are prone to
- Widely used DL algorithms like *decentralized parallel SGD* [1], *gossip learning* [2], and *epidemic learning* [3] are vulnerable to privacy violations through the sharing of model updates
- Noise-based privacy-preserving methods significantly affect model utility
- We propose SHATTER that addresses DL's privacy concerns without compromising utility or efficiency

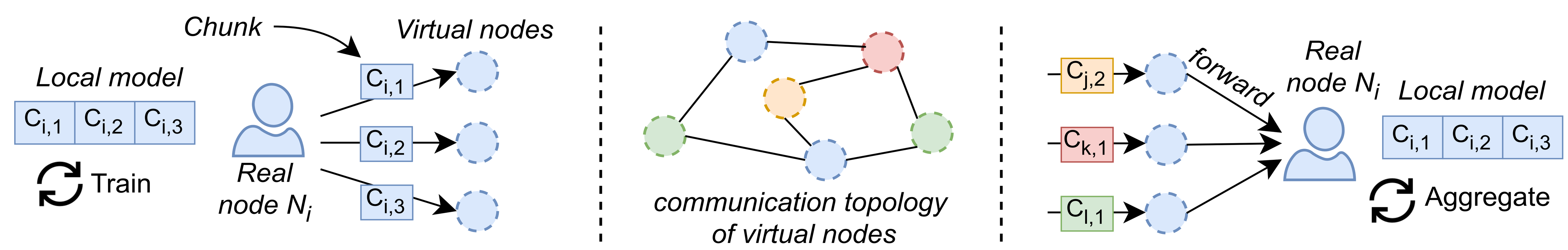
System Design

Threat Model: • Permissioned network • HbC local adversaries • No collusion

Building blocks:

- *Chunking*: restricts receiving nodes' access to a subset of model parameters (↑ privacy)
- *Full sharing*: ensures no information loss occurs (↑ utility)
- *Virtualization*: decouples nodes' identities from model chunks by means of *virtual nodes* (↑ privacy)
- *Randomized communication*: prevents structural attacks on fixed nodes (↑ privacy) and improves mixing (↑ utility)

SHATTER



SHATTER from the perspective of an arbitrary real node N_i:

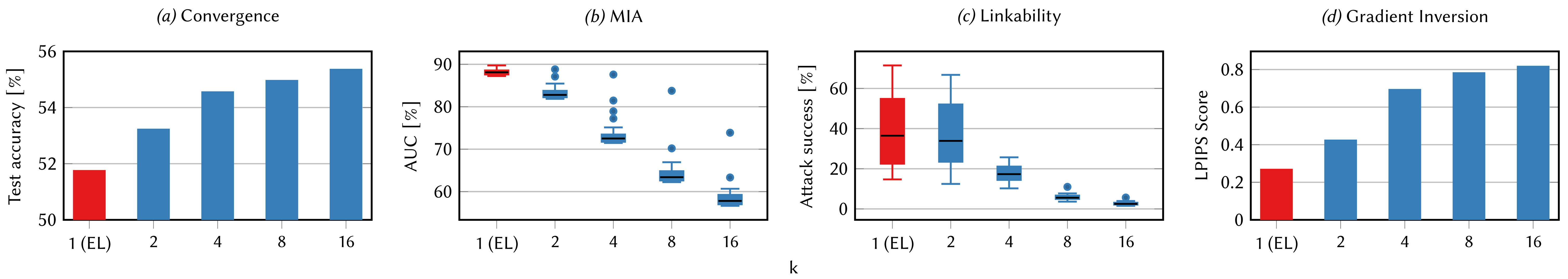
- **Initialize** $\theta_i^{(0)}$ and **spawn** k virtual nodes (VNs): $v_i(1), \ldots v_i(k)$
- For $t = 0, \ldots, T-1$:
  - $\tilde{\theta}_i^{(t,0)} \leftarrow \theta_i^{(t)}$
  - **Local training:** for $h = 1, \ldots, H$: $\tilde{\theta}_i^{(t,h+1)} \leftarrow \tilde{\theta}_i^{(t,h)} - \eta \nabla f_i(\tilde{\theta}_i^{(t,h)}, \xi_i)$
  - **Chunk** $\tilde{\theta}_i^{(t,H)}$ into k parts
  - **Forward** chunk s to $v_i(s)$ for every $s = 1, \ldots k$
  - **Randomize** r**-regular communication topology**
  - **Receive** r chunks from each of the k VNs
  - **Aggregate** the received chunks to produce $\theta_i^{(t+1)}$
- Return $\theta_i^{(T)}$

Properties

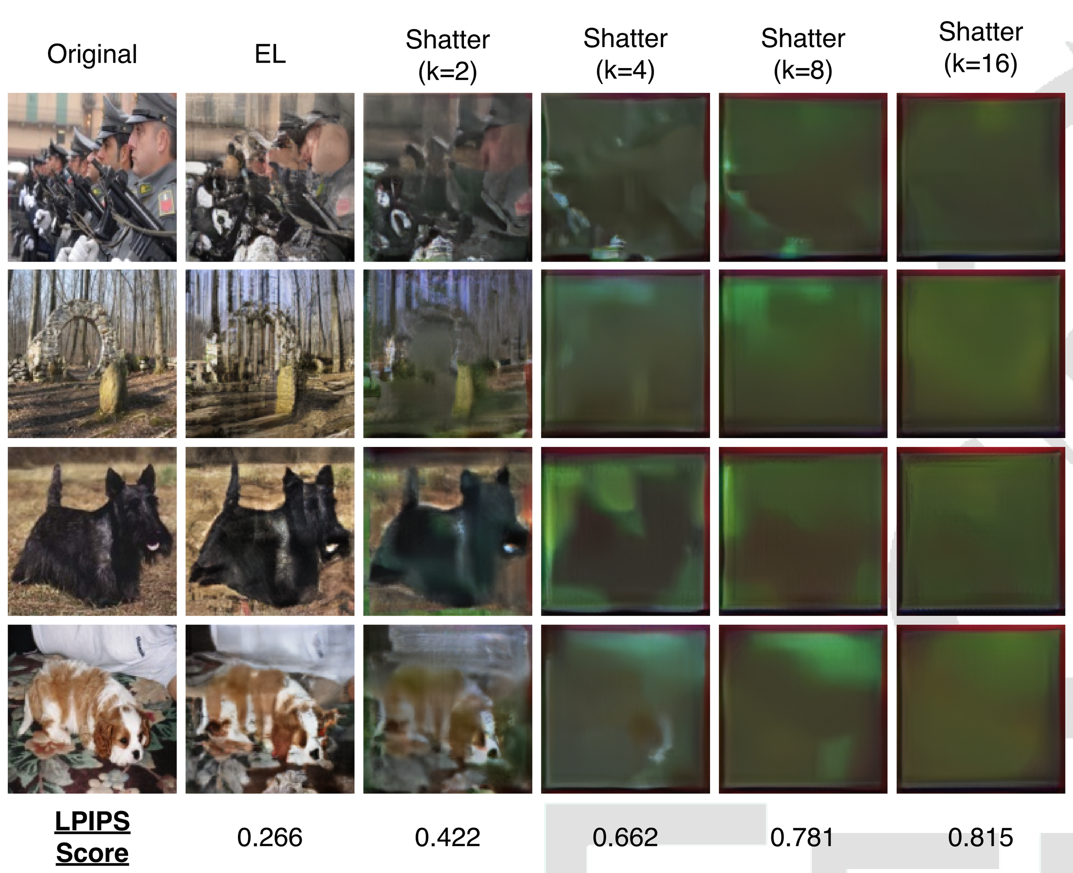
1. **Privacy guarantees**:
   - Defends better against the cutting edge *likability*, *membership inference*, and *gradient inversion* attacks than the SOTA baselines such as EL [3] and MUFFLIATO [4]
   - Improves the privacy of RNs from an information-theoretical perspective as the number of VNs operated by them increases, offering an analytical insight into the diminishing efficacy of the attacks
   - The formal privacy guarantees can be extended even when there are colluding HbC adversaries
2. **Convergence**:
   - Provably converges where the convergence rate involves regularity of local loss functions, number of local steps, number of VNs per RN, and the degree of communication graph
3. **Supports dropouts**
   - Continues to have better privacy and accuracy than its competitors even when nodes drop out at different rates during each round

Evaluation

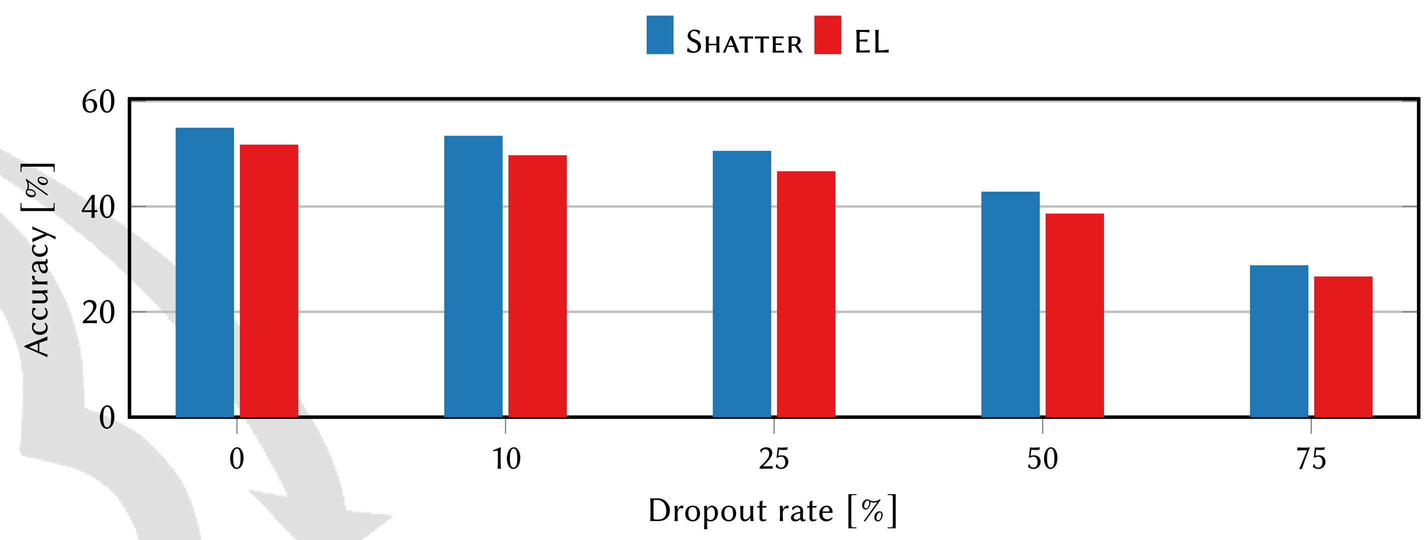
Experimental setting: • Task: Image Classification • Dataset: CIFAR-10 • Model: ResNet-18 • Training samples: 50k • Testing samples: 10k • 100 RNs • non-IID samples using Dirichlet distribution with $\alpha = 0$



Test accuracy (a, ↑ is better), MIA AUC (b, ↓ is better), attack success rate for LA (c, ↓ is better) on CIFAR-10, and GIA LPIPS score (d, ↑ is better) on ImageNet for an increasing number of VNs (k)



Reconstructed images using GIA for different numbers of VNs and the avg. LPIPS scores (↑ is better) for all the 1600 reconstructed images for each setting



Accuracy after 300 rounds at different dropout rates

- These are some of the selected illustrations from the extensive set of experiments that were performed
- Experiments with Twitter Sent-140 (task: sentiment analysis) and MovieLens (task: recommendation) datasets show similar trends w.r.t. accuracy and privacy
- In summary, SHATTER takes a pioneering step towards privacy-preserving DL without compromising utility under a marginal communication and operational overhead

[1] Xiangru Lian et al. "Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent". In: *NIPS* (2017). arXiv: 1705.09056.

[2] Róbert Ormándi, István Hegedűs, and Márk Jelasity. "Gossip learning with linear models on fully distributed data". In: *Concurrency and Computation: Practice and Experience* 25.4 (2013), pp. 556–571. DOI: 10.1002/cpe.2858.

[3] Martijn de Vos et al. "Epidemic Learning: Boosting Decentralized Learning with Randomized Communication". In: *NeurIPS*. 2023. arXiv: 2310.01972.

[4] Edwige Cyffers et al. "Muffliato: Peer-to-Peer Privacy Amplification for Decentralized Optimization and Averaging". In: *NeurIPS*. 2022. arXiv: 2206.05091.