CLIENT-LEVEL UNLEARNING IN DECENTRALIZED LEARNING

Author: Razvan Dinu (R.D.Dinu-1@student.tudelft.nl)
Supervisors: Jérémie Decouchant, Bart Cox

1. Introduction

- **Decentralized Learning (DL)** trains models via peer-to-peer communication, without a central server — offering better privacy and scalability (Fig. 1).
- When clients **leave** or **crash**, their influence remains in the model, posing privacy risks, violating the **right to be forgotten** (e.g., GDPR), or **affecting model performance**.
- Many existing Machine Unlearning (MU) methods work in centralized or federated setups, but **don't apply to DL** due to the lack of global coordination.
- This work proposes the **first client-level unlearning method tailored for DL**, supporting both **announced and unannounced** client departures.

3. Methodology

- Builds on FL SOTA by moving it to a decentralized setting, with a stronger focus on client-level unlearning. Allows for **unannounced client crashes**, a common challenge of DL.
- Each client *i* tunes synthetic dataset S_i by **minimizing the distance between gradients** on real and the generated synthetic datapoints (*gradient matching* [2]). Neighbours **cache** these datasets.
- When a client *j* disconnects, it may send out an unlearning request, or may time out after **10 minutes**. All other clients propagate the request and perform **Stochastic Gradient Ascent (SGA)** on the synthetic data S_j of the dropped node.
- Fast **recovery** of past performance is achieved by relearning on one own's synthetic dataset, **augmented** with real samples.

5. Limitations

- DL is inherently **synchronous**, which makes it **challenging** for clients to unlearn simultaneously, potentially retaining some unwanted influence of the dropped client.
- Algorithm may need further **hyperparameter tuning** for complex datasets.

References

- [1]: A. Dhasade, Y. Ding, S. Guo, A.- M. Kermarrec, M. de Vos, and L. Wu. "QuickDrop: Efficient Federated Unlearning via Synthetic Data Generation". In: Proceedings of the 25th International Middleware Conference. Middleware '24. New York, NY, USA: Association for Computing Machinery, Dec. 2024, pp. 266–278.
- [2]: B. Zhao, K. R. Mopuri, and H. Bilen. Dataset Condensation with Gradient Matching. In: International Conference on Learning Representations, 2021.

TU Delft

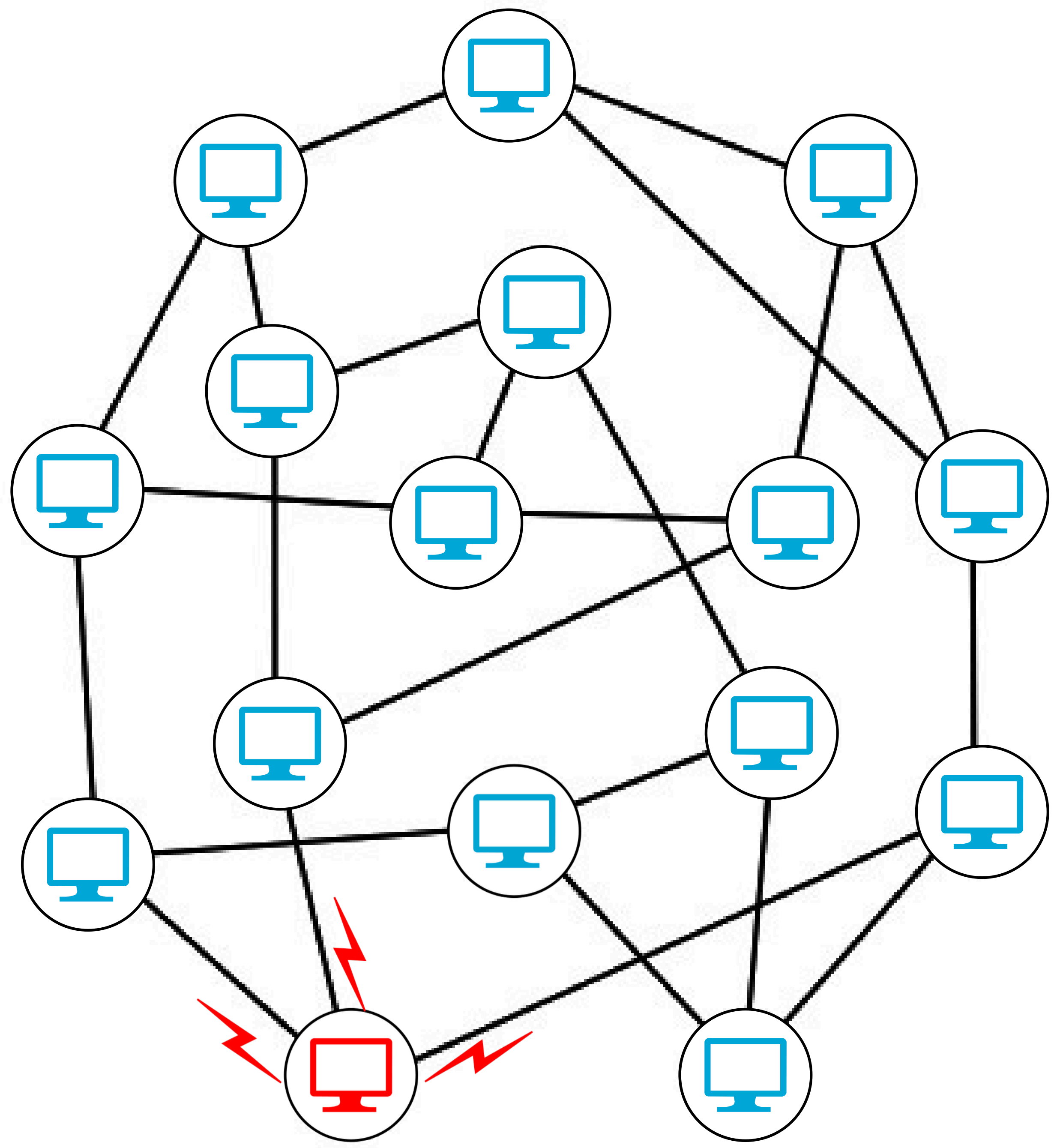

Figure 1: A **3-regular 16-client** DL system where one client disconnects. Made with *draw.io*

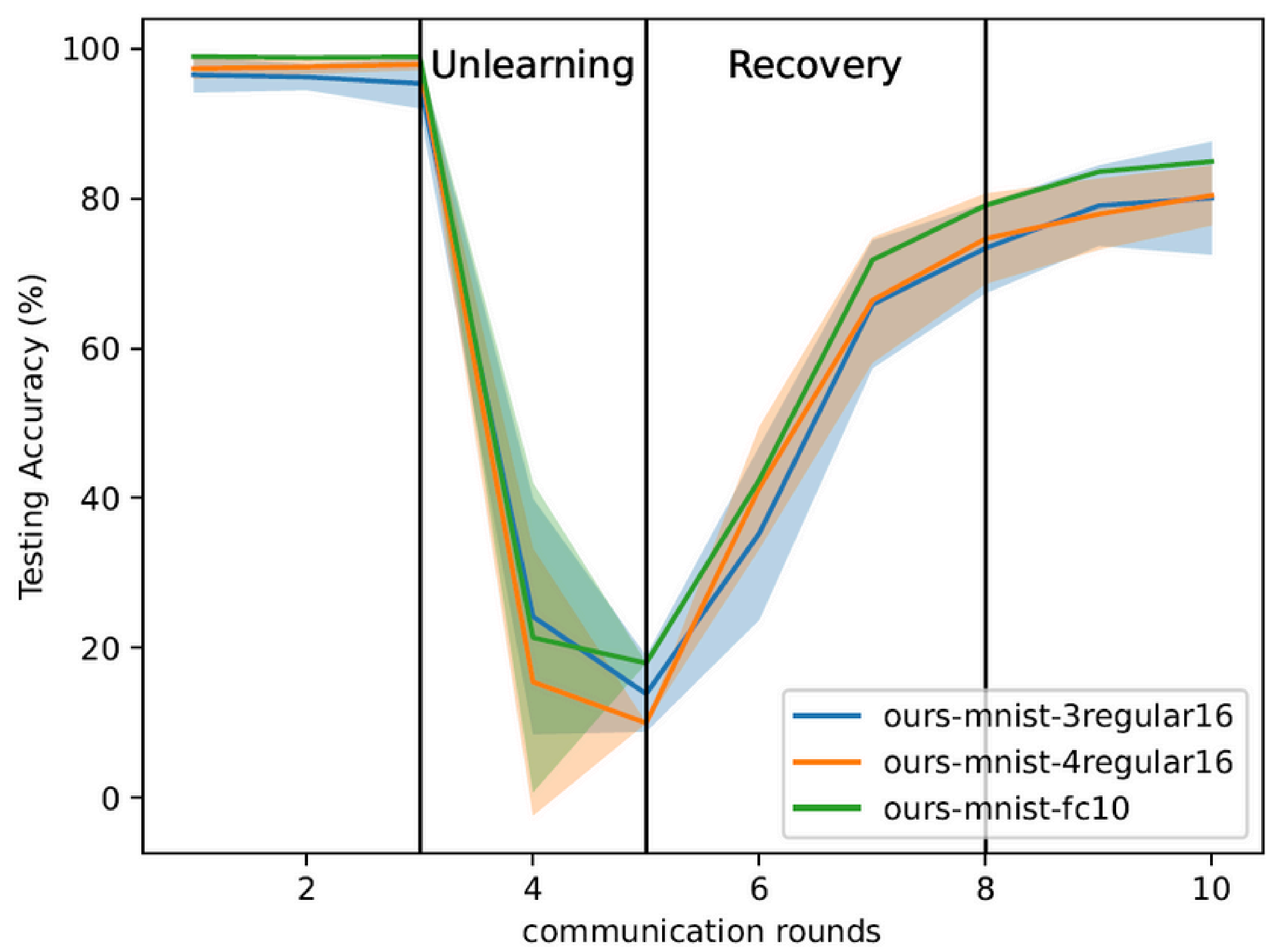

Figure 2: Average testing accuracy of the clients in the network before, during, and after unlearning and recovery the MNIST dataset. Client 8 disconnects at iteration 3. Comparison of 3 topologies: **3-regular 16-node** (*3regular16*), **4-regular 16-node** (*4regular16*), and **fully connected 10-node** (*fc10*). Standard deviation is shown by the shaded area.

2. Contributions

- We **translate** a current state-of-the-art (SOTA) machine unlearning algorithm, QuickDrop [1], from Federated to Decentralized Learning, overcoming the architectural differences between the two systems by fine-tuning its parameters. We focus on **unlearning the influence of a particular client** and report the **generalization capabilities** of the remaining model.
- We **improve** on the existing SOTA by considering crashed clients and further tuning synthetic data generation.
- We **analyse** the impact of the **network topology**, **datasets** and **disconnection frequency** w.r.t. the unlearning efficiency in DL.
- We provide a **complete implementation** of our decentralized unlearning algorithm with support for different network topologies and crash recovery, **evaluating** its practical effectiveness against **established theoretical benchmarks**.

4. Performance Evaluation

- Tested against **two** theoretically optimal baselines:
  - **Retrain-Or**: Retrains the model from scratch without the dropped client → accurate but extremely slow.
  - **SGA-Or**: Performs SGA using the dropped client's real data → fast but violates privacy.
- Tested on **three** topologies with varying degrees of connectivity (Fig. 2)
- Remains competitive with baselines, achieving a **21-30x** speedup over **Retrain-Or**, while recovering **more** of its past performance than **SGA-Or**.
- Overhead: **2%** storage and ~**80%** train time
- **Fast and effective** even after **50%** of the network crashing one by one.
- **FID score** of synthetic data: 1.57 (MNIST), 487 (CIFAR-10).

6. Conclusions

- Works even with **unannounced client crashes**.
- **Sparse & non-IID** setups make unlearning more difficult.
- Communication cost **increases** with network size.
- Extendable to **class-** and **sample-**level unlearning.
- Enables real-world, **privacy-compliant** DL systems.