FL-Torrent: decentralised AI for the masses

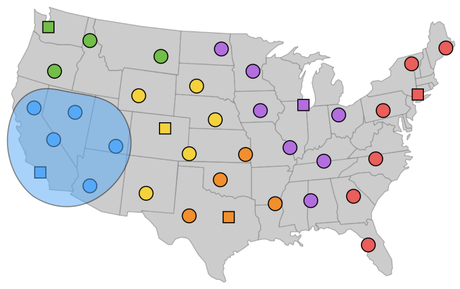
Alvaro Garcia-Recuero[1] and Nikolaos Laoutaris[1]

[1]IMDEA Networks Institute, Spain

Introduction

- **Google** [1] has democratised Federated Learning with TensorFlow but vendor lock-in is still a problem for production deployments (proprietary APIs).
- Federated Learning is employed at large Internet corporations with the promise of improved Privacy to their users but Inference and Poisoning attacks are not difficult [2].
- Decentralised learning works [3] but it is unsure how to enforce security.
- FL-Torrent aims to bring the following advancements to Federated Learning:
- Problem:
  - Previous work has not explored the impact of Poisoning or Inference attacks on decentralised AI/learning.
  - All to All Network Topology favours the malicious.
  - Probabilistic countermeasures may be just as (in)effective as in traditional Federated Learning with a centralised server component.
- Goals of this project:
  - Validate our expectations about the resilience of Centralised vs Decentralised AI in terms of security.
  - Democratisation of decentralised AI for the masses ensuring robustness against poisoning/inference attacks.
  - Application to a real use-case. RIPE ATLAS catchments form community-like structures where we want to apply decentralised vs federated AI.
  - Open source software.

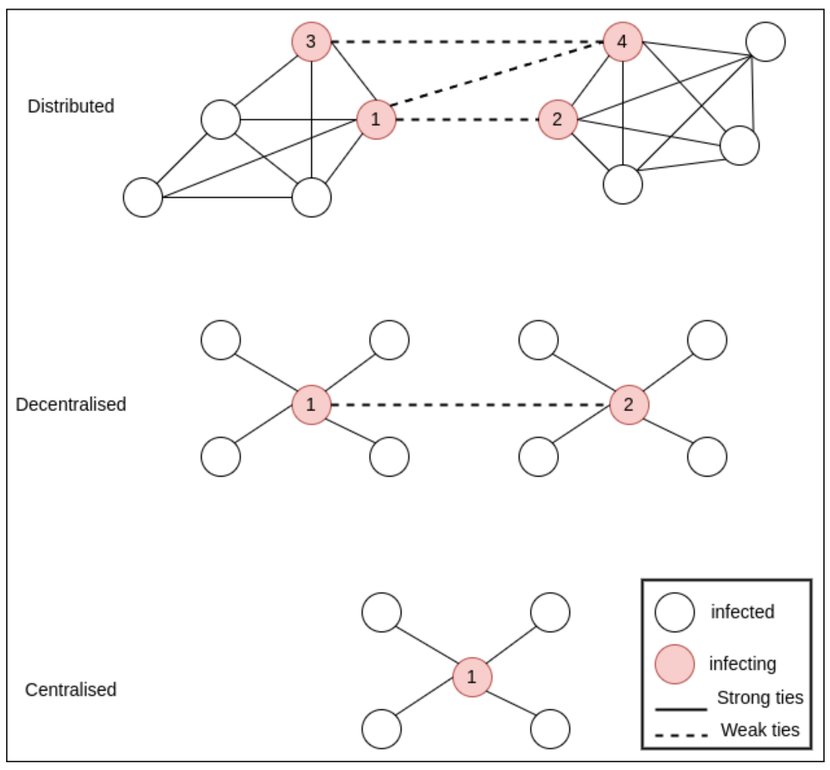
[a]Squares represent trusted nodes, circles clients



[a]Image taken from https://eng.edgecast.com/2017/11/20/seeing-the-world/

Architecture

- Chosen topology affects the resilience and robustness of attack countermeasures.
- Starting experiments with a *star* type of architecture where the central node can be malicious in order to show the security issue of the aggregator.



Collaborative Learning in trust less networks

- Trust is decentralised and thus available to malicious participants.
- Only a recent work [4] has tested attacks on decentralised AI using static, fixed network topologies, concluding they affect to the robustness of defences for decentralised AI.
- Inference attacks should be considered and defeated using security techniques for fully distributed or decentralised network topologies.

CHALLENGES

- Select *k* trusted fixed or random neighbours to exchange gradients.
- Select number of necessary peers *N* to meet probabilistic security guarantees defined by our model.
- Avoid contagion to nearby clusters or communities while allowing fast propagation of chains of good updates inside the same cluster.

Methodology

- Access control
  - Avoids having a malicious client trying to connect to all peers in order to get all the updates just like a centralised aggregator and thus launch inference attacks: mixing the peers or content of peers selected at each round.
  - Make sure that malicious peers do not keep joining a torrent with different ideas of the model gradient in order to launch poisoning attacks and defeat reputation schemes.
- Privacy: using a mixed network by mixing the layers of the model updates among participants before sending them to the aggregate server (centralised) or neighbourhood set (decentralised).
- Security: use the choke/unchoke protocol in BitTorrent over Reputation schemes [5] instead of over file hashes as a torrent does, in order to work against propagation of poisoning updates.
- Economics: reward scheme on top of FL-Torrent, e.g., if you disseminate high quality updates you get paid in Reputation or in a Reinforcing learning rate.
- Locality: Markov Decision Process with greedy approach.
  - For performance reasons (delay bandwidth etc)
  - To capture rewards so that I may prefer only updates from local clients that increase my exploration and exploitation of the network.
  - Solution can be at the tracker level (squares) by having different *fl-torrents* for different countries/cities (e.g., RIPE ATLAS).

Use Case

- Topology used to exchange gradients opportunistically.
- Locally or globally using reputation based aggregation among nodes or clusters of nodes.
- Malicious nodes in one cluster do not propagate updates to another.
- Malicious nodes inside a cluster monitored by a trusted peer or set of trusted peers defined by *k* fixed neighbours metric.

References

[1] `https://www.tensorflow.org/federated` *TensorFlow Federated: Machine Learning on Decentralised Data* . In ***Google APIs*** *- May 6th of 2022*

[2] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, Vitaly Shmatikov *How To Backdoor Federated Learning.*. In ***Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*** *- 2020*

[3] Hegedűs, István, Gábor Danner, and Márk Jelasity. *Decentralized learning works: An empirical comparison of gossip learning and federated learning.*. In ***Journal of Parallel and Distributed Computing 148: 109-124.*** *- 2021*

[4] Pasquini, Dario, Mathilde Raynal, and Carmela Troncoso. (2022). *On the Privacy of Decentralized Machine Learning.* In ***arXiv preprint arXiv:2205.08443*** *- 2022*

[5] Chu, T., Garcia-Recuero, A., Iordanou, C., Smaragdakis, G., & Laoutaris, N. (2022). *Securing Federated Sensitive Topic Classification against Poisoning Attacks.* In ***arXiv preprint arXiv:2201.13086.*** *- 2022*

Join us