University of British Columbia

Dancing in the Dark: Private Multi-Party Machine Learning in an Untrusted Setting

Clement Fung, Jamie Koerner, Stewart Grant, Ivan Beschastnikh

We trust cloud providers with our data for Machine Learning (**ML)**.

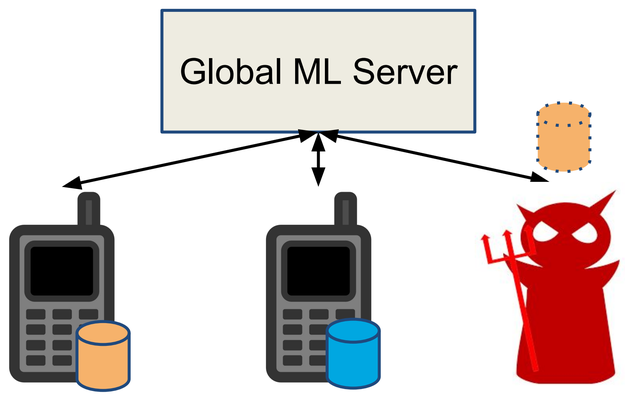
Can we support ML in **private**, **anonymous**, **untrusted** settings?

Challenges

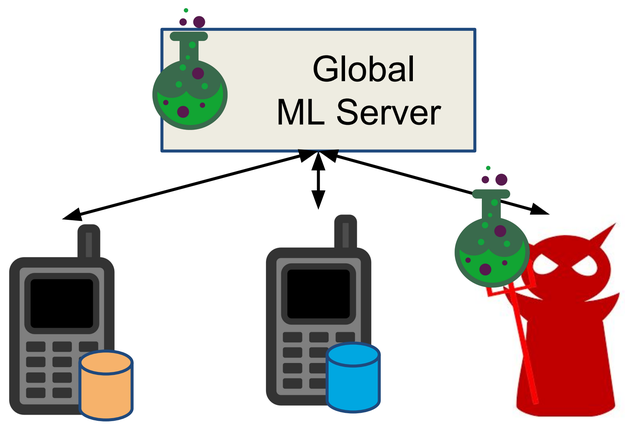
- Machine learning (ML) at scale is popular:
  - **But not privacy focused, vulnerable**
- Differentially private ML exists:
  - **But theoretical and not well applied**
- Tor allows anonymity in P2P networks
  - **But how do we use it for ML?**

ML attacks we consider

Information Leakage: Copy private training data through observing ML model while training

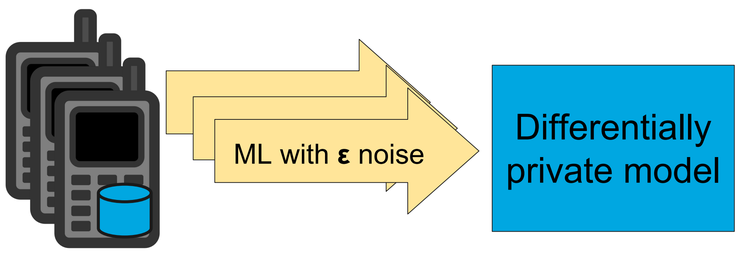


Random Model Poisoning: Manipulate performance of global model with bad updates

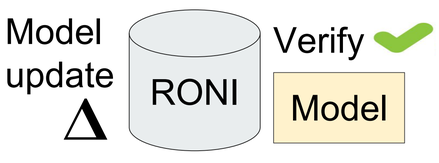


Defense ingredients

1. **Differential Privacy:** Privacy in ML from **ε** noise



2. RONI: Verify correct ML; use validation score



3. Tor: Anonymize network traffic source

Our contributions

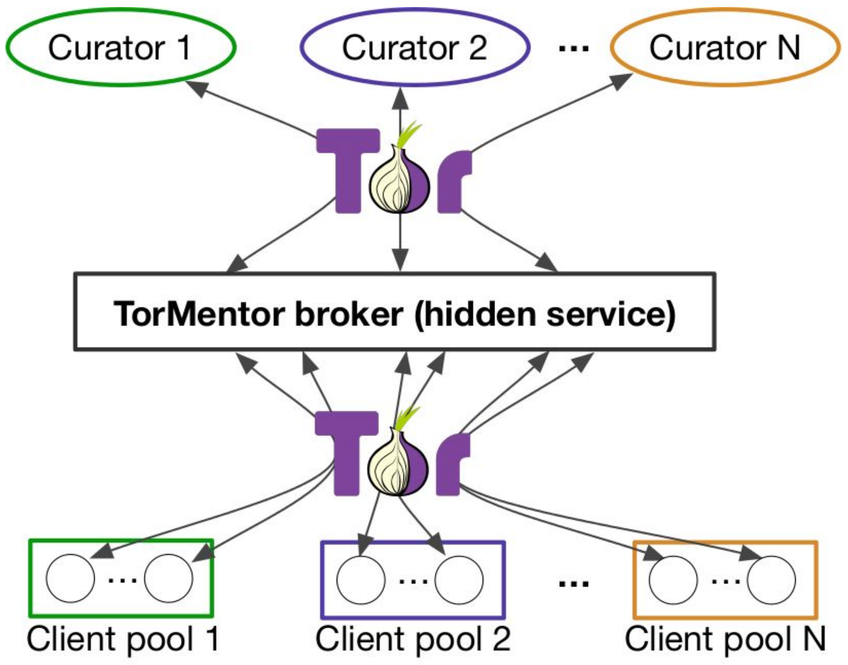
1. Brokered Learning: A design paradigm and API for ML in **untrusted** setting
2. **TorMentor**: a system for anonymous, differentially-private multi-party ML
3. We translate known **ML attacks** (Information Leakage, Poisoning) and **defenses** (RONI, Differential Privacy) to a private, untrusted setting

TorMentor design

Wants to train ML model, need training data

Provide trusted guarantees to facilitate anonymous ML

Privately use training data, (Privacy **ε**, batch size bystanders)

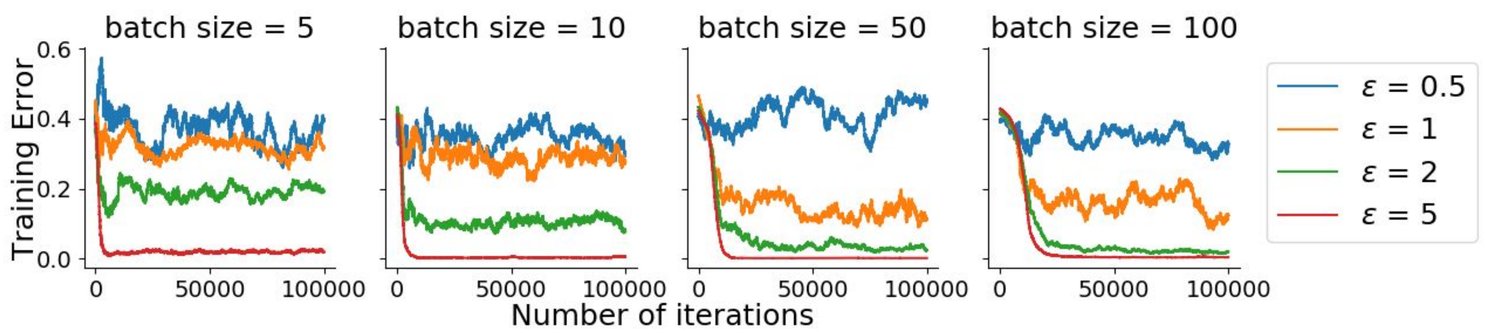


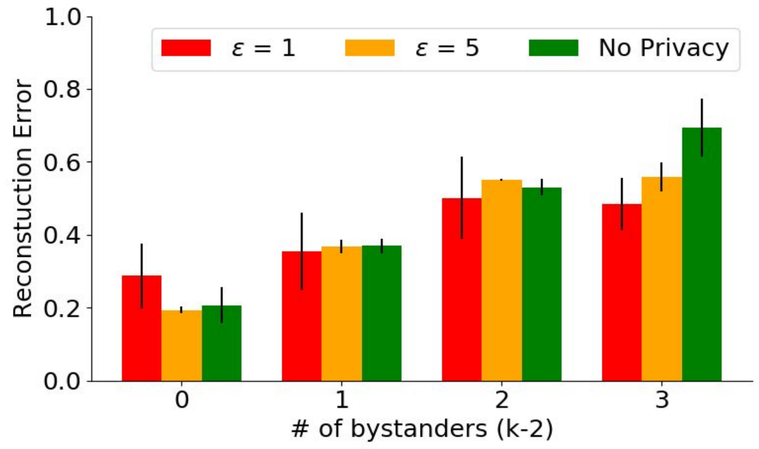
Brokered Learning API

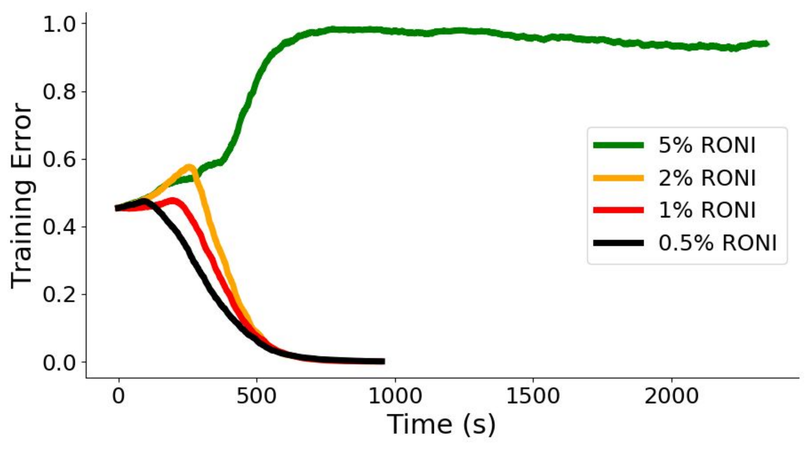
1. Curator: define model, validation set
2. Client: Join model, privacy constraints
3. Train model via differentially private SGD
4. Anonymously return model to curator

Evaluation results

1. Effects of client batch size and privacy parameters in multi-party ML
2. Defending against inversion: Including users, DP mitigates attacks
3. Defending against poisoning at scale: Tuning RONI threshold







Future Research Directions

1. Investigate client incentives for anonymous ML
2. Improve understandability of privacy parameters
3. Mitigation of attacks with fewer assumptions
4. Remove trust model in broker

References

1. McMahan et al. "*Communication-Efficient Learning of Deep Networks from Decentralized Data* ", AISTATS '17
2. Song et al. *"Stochastic Gradient Descent With Differentially Private Updates,* GlobalSIP' 13
3. Huang et al. "*Adversarial Machine Learning*", AISec '11
4. Hitaj et al. "*Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning*" CCS'17

THE UNIVERSITY OF BRITISH COLUMBIA

Partners

NSERC CRSNG

HUAWEI