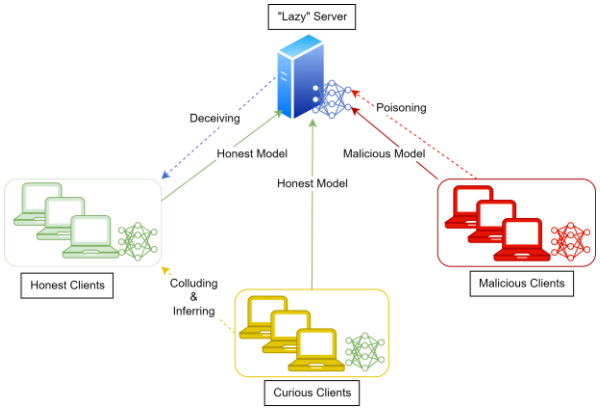
BVDFed: Byzantine-Resilient and Verifiable Aggregation for Differentially Private Federated Learning

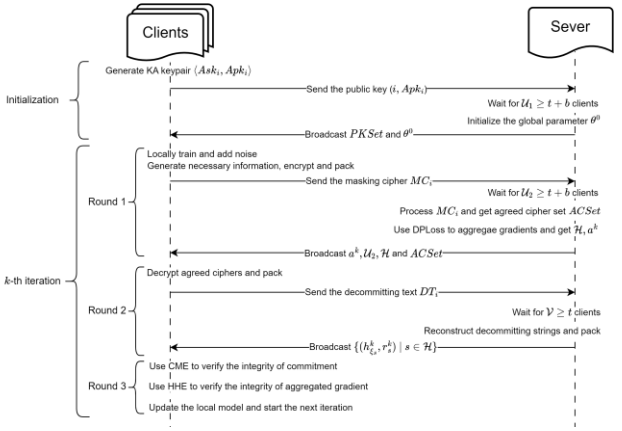
Xinwen GAO, Shaojing FU, Lin LIU, Yuchuan LUO

Frontiers of Computer Science, DOI: 10.1007/s11704-023-3142-5

Problems & Ideas

- Problems of conventional Differentially Private Federated Learning aggregation protocols:
  - Exsiting methods focus on either Byzantine-resilience or verifiability.
  - The adversarial model considers both the security issues of some malicious clients and “lazy” server, which is more threatening and challenging.
- Ideas: BVDFed, a Byzantine-resilient and verifiable aggregation for DPFL, consists of an efficient algorithm and two compatible methods to address the challenging problems.

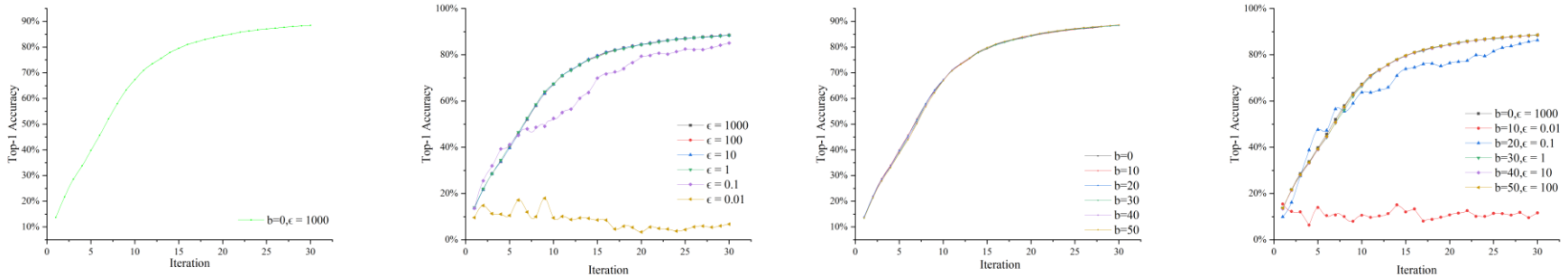




Left: The architecture of the adversarial model in Federated Learning. First, either the server or clients could be honest-but-curious. Second, some curious clients might collude with each other. Third, the malicious adversaries are capable of manipulating some participants for wrecking joint model. Fourth, The aggregation server may be ”lazy” and fabricate the results without aggregation.
Right: The high-level view of BVDFed protocol.

Main Contributions

- Contributions:
  - We propose BVDFed, a Byzantine-resilience and verifiable aggregation for DPFL. BVDFed alleviates the privacy and security issues in the adversarial model where exist the curious participants and both the corrupted clients and server;
  - Specifically, we propose the efficient privacy-preserving Differentially Private Federated Averaging algorithm (DPFA). And we propose a Byzantine-resilient aggregation rule (DPLoss) in the DP conditions. In DPLoss, we propose *Loss Score* as the indicator of noisy gradients' trustworthiness. In addition, we propose a secure verification scheme (DPVeri) that is compatible with DPFA and DPLoss;



(a) (b) (c) (d)

The Experiment on Accuracy of Classification. Note that b denotes the number of Byzantine clients and ϵ denotes the privacy budgets. (a) The baseline of accuracy of classification with small noises and no Byzantines. (b) The accuracy of classification with different sizes of privacy budgets and no Byzantines. (c) The accuracy of classification with different number of Byzantines and small noises. (d))The accuracy of classification with both noises and Byzantines in different settings.