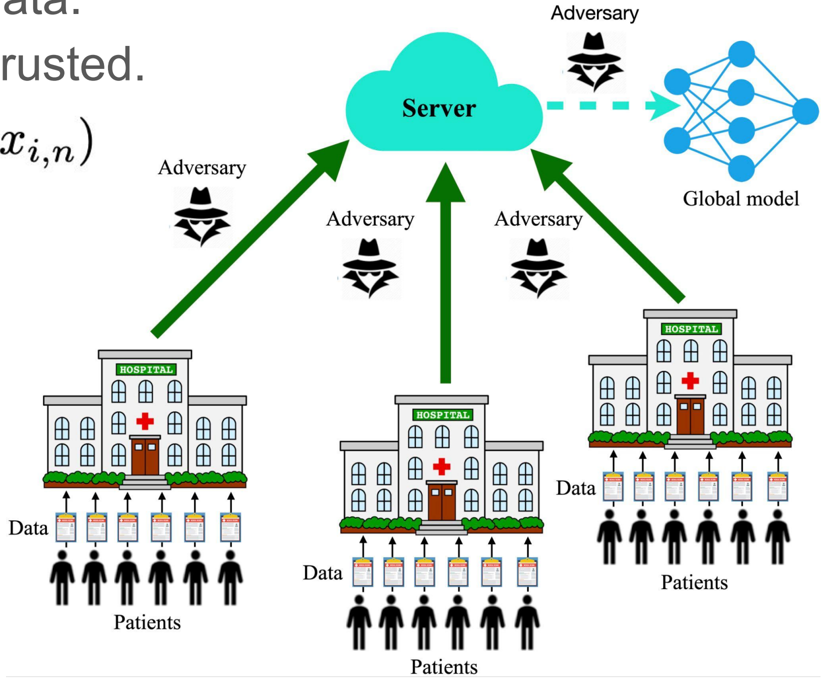
Problem Setup

- Federated Learning (FL) with sensitive data.
- The server or other silos/clients are not trusted.
- N silos, n samples each $X_i = (x_{i,1}, \cdots, x_{i,n})$
- Silo **i**'s data distribution is $\mathcal{D}_i$
- And seeks to minimize

$$F_i(w) := \mathbb{E}_{x_i \sim \mathcal{D}_i}[f(w, x_i)].$$

- Global objective (FL problem):

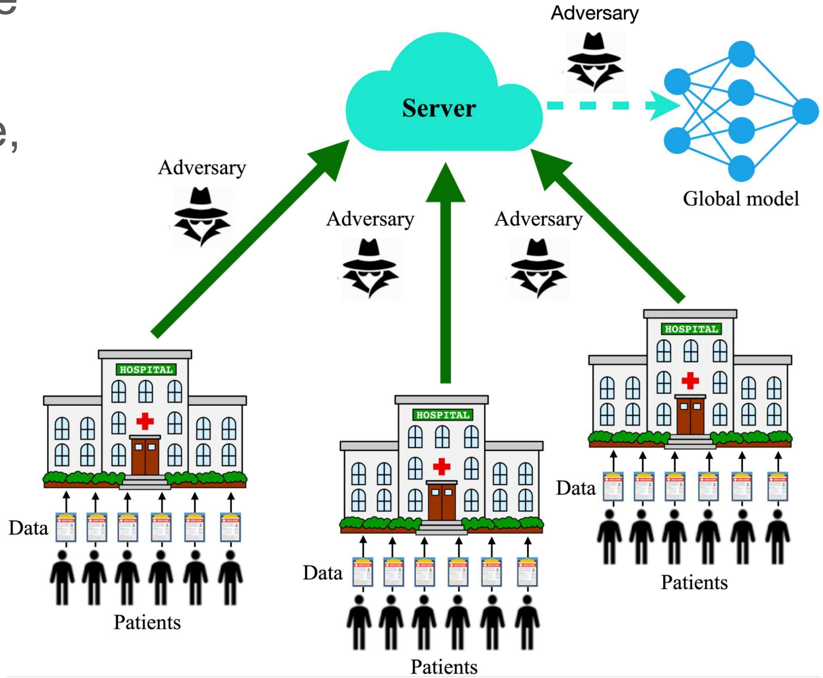
$$\min_{w \in \mathcal{W}} \left\{ F(w) := \frac{1}{N} \sum_{i=1}^{N} F_i(w) \right\}$$



Inter-Silo Record-Level Differential Privacy (ISRL-DP)

- Each silo wants to keep their data private
- They only send privatized data.
- Even if other silos and the server collude, privacy is still guaranteed.

This contrasts with *Central DP* in which the **trusted** server will run DP algorithms and ensure that the output is private.

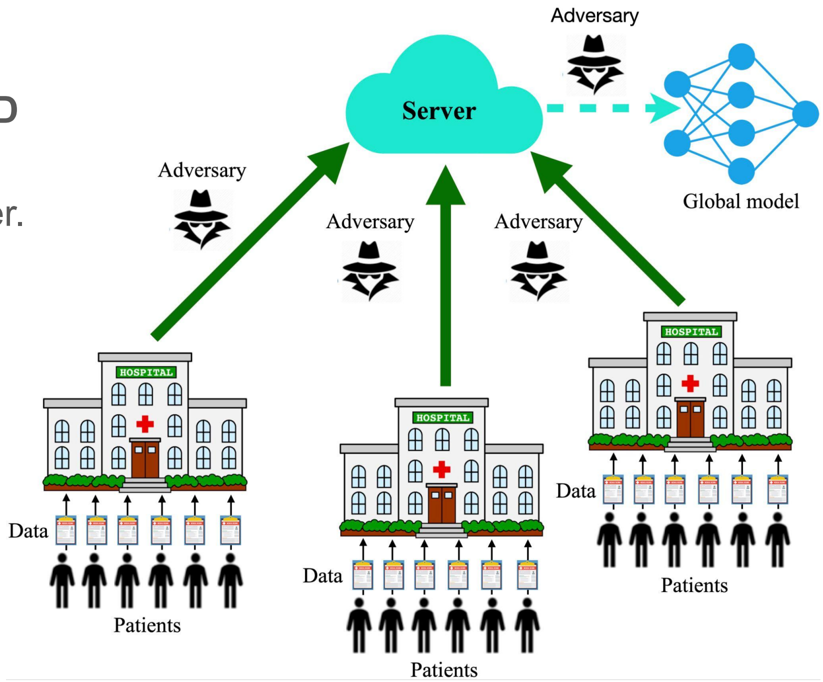


Problem Setup (cont.)

- Privacy: ISRL-DP
- Assumptions:
    - Domain ***W*** is closed and convex, with diameter **D**
    - For all **x**, f(·, x) is ***L***-Lipschitz and convex
    - Each round, **N** silos communicate with the server.
- Heterogeneous (non-iid.) setting:
- Each data distribution $\mathcal{D}_i$ may be arbitrary.
- Quality: measured by *excess risk*

$$\mathbb{E}[F(\mathcal{A}(\mathbf{X}))] - F^*$$

- Complexity
    - Communication cost R = #rounds
    - #grad evaluations



Contributions

- Optimal excess risk for *heterogeneous* (non iid.) data.
- Previous work [LR23] only has it for *homogeneous* (iid.) data.

$$\tilde{\Theta}\left(\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{n}}+\frac{\sqrt{d\log(1/\delta)}}{\varepsilon n}\right)\right)$$

- Lower communication cost and better gradient complexity.
- Similar improvements for nonsmooth losses.

Algorithm & Setting	Excess Risk	Communication Complexity	Gradient Complexity
[LR'23] Alg. 2 (i.i.d.)	optimal	Nn	N^2n^2
[LR'23] Alg. 1 (non-i.i.d.)	suboptimal	$N^{1/5}n^{1/5}$	Nn
Alg. 4 (non-i.i.d.)	optimal	$N^{1/4}n^{1/4}$	$N^{5/4}n^{1/4}+(Nn)^{9/8}$

Algorithm Overview

- Combines Iterative localization technique [FKT20] and Multi-stage Inter Silo Record Level-DP Accelerated Minibatch-SGD
- Multiple phases. In each phase:
  - Use disjoint samples
  - Solve a regularized ERM problem using Accelerated MB-SGD

$$\hat{F}_i(w) = \frac{1}{n_i N} \sum_{l=1}^{N} \sum_{x_{l,j} \in B_{i,l}} f(w; x_{l,j}) + \frac{\lambda_i}{2} \|w - w_{i-1}\|^2$$

  - Localization: increasing regularization, fewer # samples
- Extend to nonsmooth losses via smoothing

Proof Sketch

Privacy guarantees follow from our choice of parameters, advanced composition (or moment accountants) and parallel composition.

We will give a sketch proof of the excess risk bound as follows:

$$\widetilde{\Theta}\left(\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{n}}+\frac{\sqrt{d\log(1/\delta)}}{\varepsilon n}\right)\right)$$

Proof Ideas - Excess Risk

1. The DP solution for each ERM is close to the true solution of the ERM

2. By stability, we can bound the population risk via ERM

3. Observe that we can write as follows and bound each term

$$\mathbb{E}F(w_\tau) - F(w^*) = \mathbb{E}[F(w_\tau) - F(\hat{w}_\tau)] + \sum_{i=1}^{\tau} \mathbb{E}[F(\hat{w}_i) - F(\hat{w}_{i-1})].$$

$\hat{w}_i$: true solution of the ERM

Numerical Experiments

- MNIST data preprocess to simulate heterogeneous FL settings

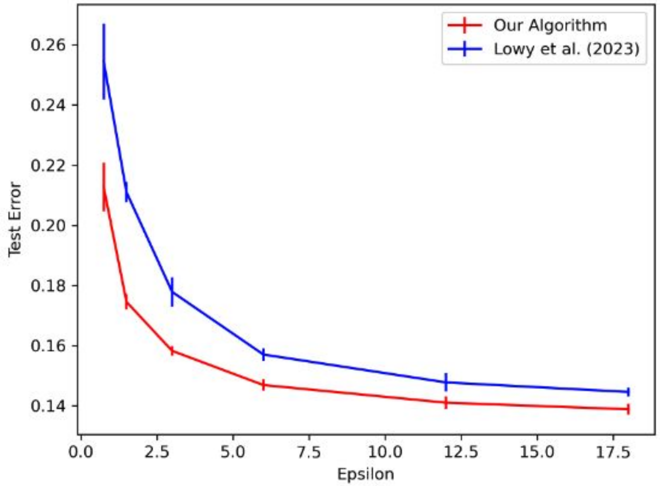


Figure 1: Reliable Communication

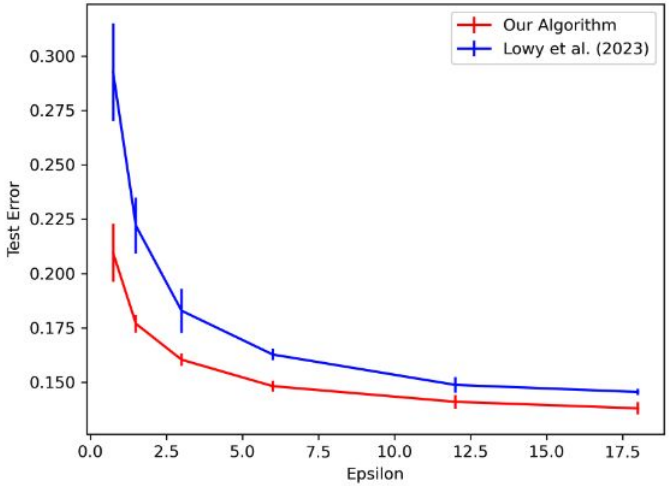


Figure 2: Unreliable Communication

Summary for DP FL without a Trusted Server

- Problem: FL without a trusted server, inter-silo record-level DP
- Method: Combine Iterative localization technique [FKT20] with DP FL version of Accelerated MB-SGD
- Results:
  - Optimal excess risk for *heterogeneous* (non iid.) data.
  - Lower communication cost and better gradient complexity.
  - Extend to nonsmooth losses via smoothing