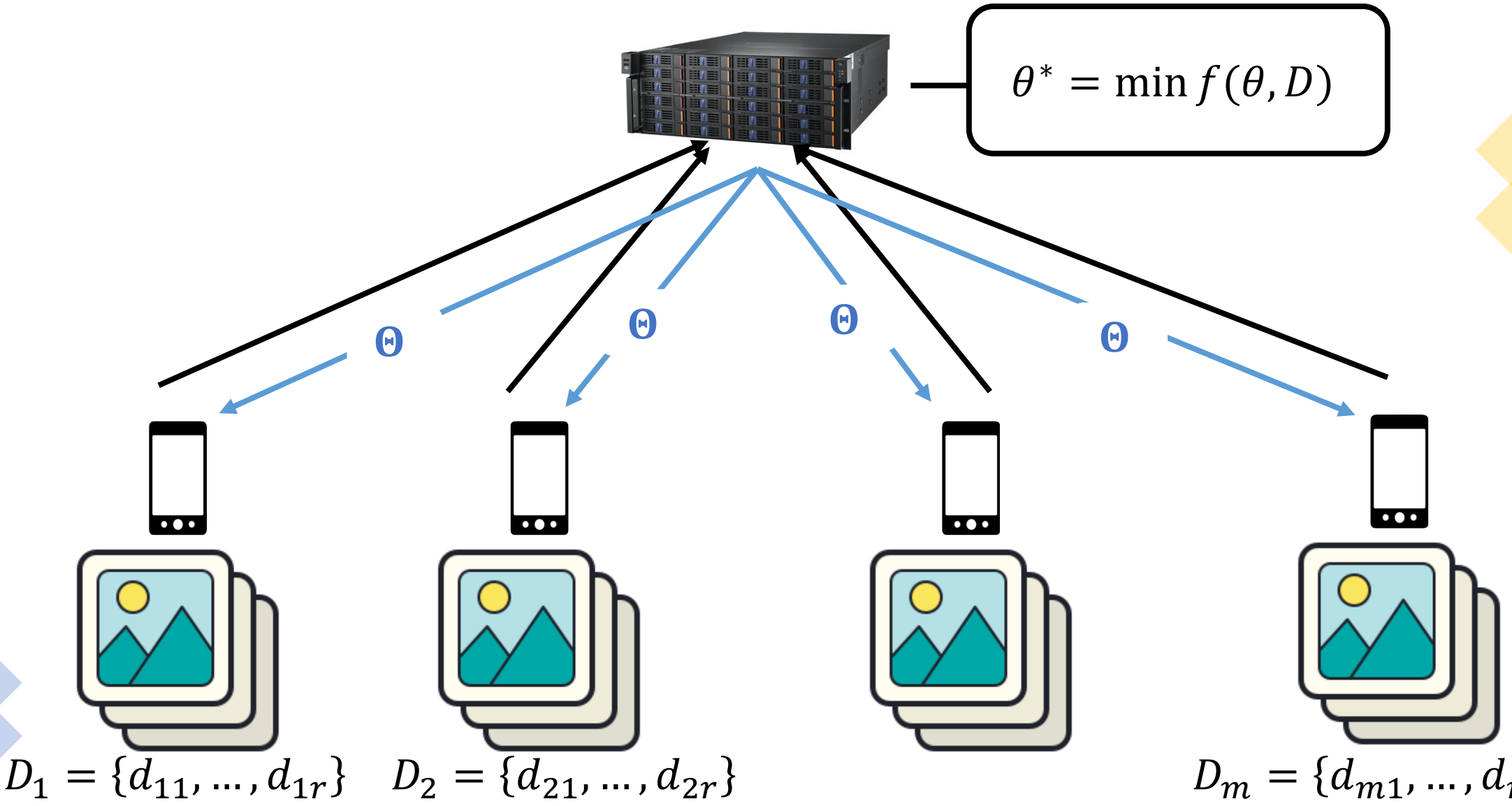
Google

Shuffled Model of Differential Privacy in Federated Learning

Antonious M. Girgis, Deepesh Data, Suhas Diggavi,
Peter Kairouz, and Ananda Theertha Suresh

Context and Motivation



Context and Motivation

1- Privacy

Context and Motivation

2- Communication

Context and Motivation

Communication-Efficient and Privacy Preserving Machine Learning with good learning performance

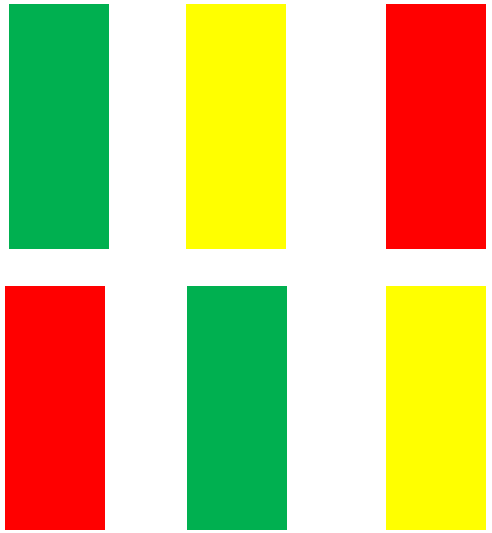
Local Differential Privacy[Kasiviswanathan et al. 2011]

Let $\mathcal{X}$ de a set of all possible inputs

A mechanism R is $\epsilon_l - LDP$ if
$\Pr[R(d) \in S] \leq e^{\epsilon_l} \Pr[R(d') \in S]$ For all $d, d' \in \mathcal{X}$

d

d'



Issue: Too much randomization/noise is needed.
We need ways to improve this.

CLDP-Stochastic Gradient Descent (CLDP-SGD)

At round $t \in T$

Client Sampling

$\theta_{t+1} = \theta_t - \eta_t G_t$

$G_t = \frac{1}{k} \sum \mathcal{R}_p(\mathrm{d}_{\mathrm{i}s_i})$

π :

$\mathcal{R}_p(\mathrm{d}_{1s_1})$

$\mathcal{R}_p(\mathrm{d}_{2s_2})$

$\mathcal{R}_p(\mathrm{d}_{3s_3})$

$\mathcal{R}_p(\mathrm{d}_{4s_4})$

Local SGD:
Sampling local dataset

Local SGD:
Sampling local dataset

Stochastic Gradient Descent (SGD)

Our contributions are mainly two ingredients:

1. Design (unbiased) compressed and private mean estimation under LDP for different ℓ_p geometries that requires $\mathcal{O}(\log(d))$ bits per gradient.
2. New privacy amplification theorems that combine sampling and shuffling.

Client/data sampling is non-uniform sampling of data points.

Privacy Results of CLDP-SGD

Theorem. Choose k out of m client at each round

Privacy: for $\mathrm{n} = \mathrm{mr}, q = \frac{k}{n}$, after T iterations and $\epsilon_l = \mathcal{O}(1)$:

$$\epsilon = \mathcal{O}\left(q\,\epsilon_l\sqrt{\frac{\mathrm{T}\log(1/\delta)\log(T/\delta)}{qn}}\right)$$

- Privacy Amplification by sub-sampling: $\mathcal{O}(q)$
  - Client/data sampling together result in non-uniform sampling from the entire dataset. Existing results cannot be applied.
- Privacy Amplification by Shuffling: $\mathcal{O}\left(\sqrt{\frac{\log(T/\delta)}{qn}}\right)$
- Strong Composition: $\mathcal{O}\left(\sqrt{T\log(1/\delta)}\right)$

CLDP- Stochastic Gradient Descent (CLDP-SGD)

Theorem.

Convergence: For convex, Lipschitz continuous objective F w.r.t. the dual of ℓ_p −norm, i.e., gradients have bounded ℓ_p −norm for $p \geq 1$, CLDP-SGD converges with rate:

$$\mathbb{E}[F(\theta_T)] - F(\theta^*) = \mathcal{O}\left(\frac{D \log T \sqrt{\max\left(d, d^{2-2/p}\right)}}{\sqrt{qTn}\,\epsilon_l}\right)$$

Communication: Each client sends $\frac{k}{m} \times \log\left(2^b\right)$ bits in expectation per iteration, where $b = \log(d) + 1$ if $p \in [1, \infty]$ and $b = d$otherwise.

Main Results:

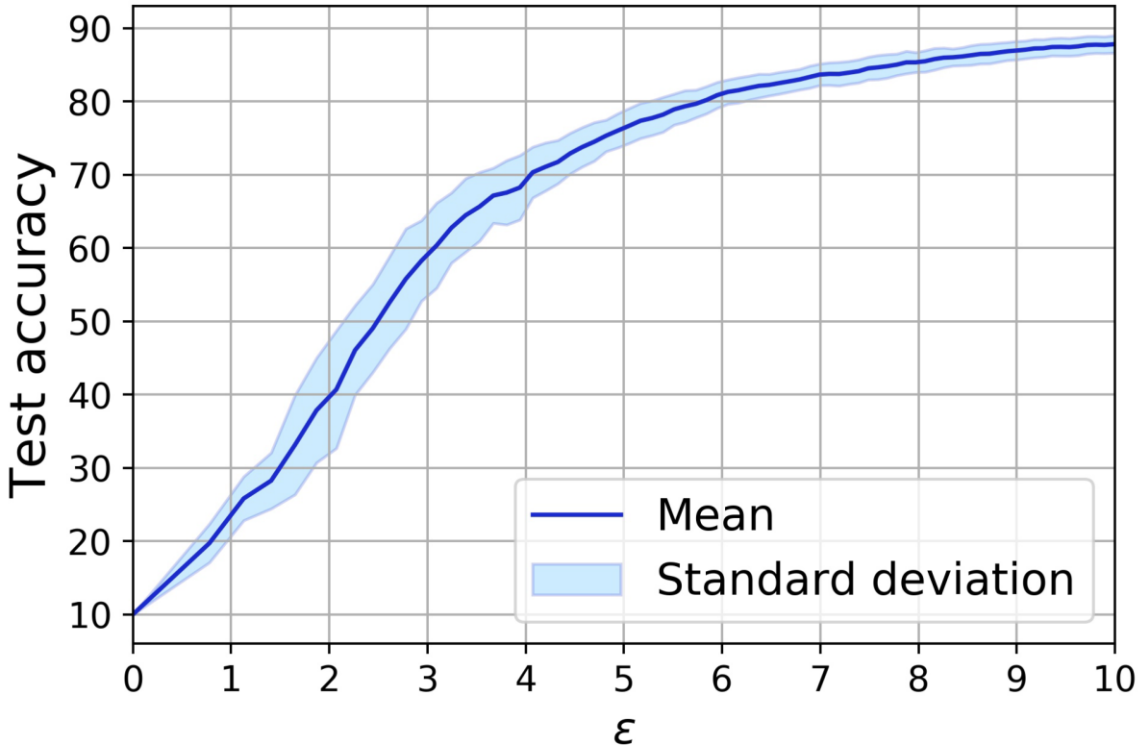
For ℓ_2 Lipschitz functions:

CLDP-SGD algorithm has optimal privacy-convergence rate (same as Bassily et. al [1]) and optimal privacy (same as ESA++[2]), while it uses only $\log(d) + 1$ bits of communication per sample.

[1] Bassily, Raef, Adam Smith, and Abhradeep Thakurta. "Private empirical risk minimization: Efficient algorithms and tight error bounds." *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 2014.
[2] U´ . Erlingsson, V. Feldman, I. Mironov, A. Raghunathan, S. Song, K. Talwar, and A. Thakurta, "Encode, shuffle, analyze privacy revisited: formalizations and empirical evaluation," arXiv preprint arXiv:2001.03618, 2020.

Numerical Results:

- MNIST dataset.
- LDP: $\epsilon_l = 2$
- Privacy loss: $\delta = 10^{-5}$
- Num. of parameters: $d = 13{,}170$
- Communication: $\log(d) + 1 = 15$ bits per gradient.



THANK YOU!