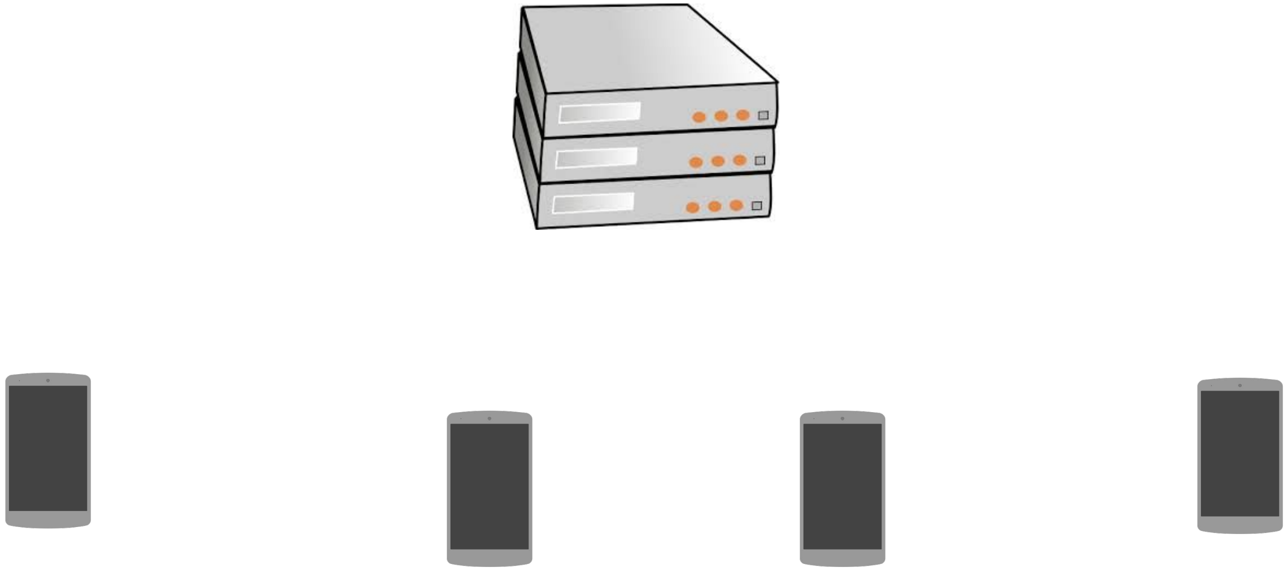
cpSGD: **c**ommunication-efficient and differentially-**p**rivate distributed SGD

Naman Agarwal, Ananda Theertha Suresh, **Felix X. Yu**
Sanjiv Kumar, H. Brendan McMahan

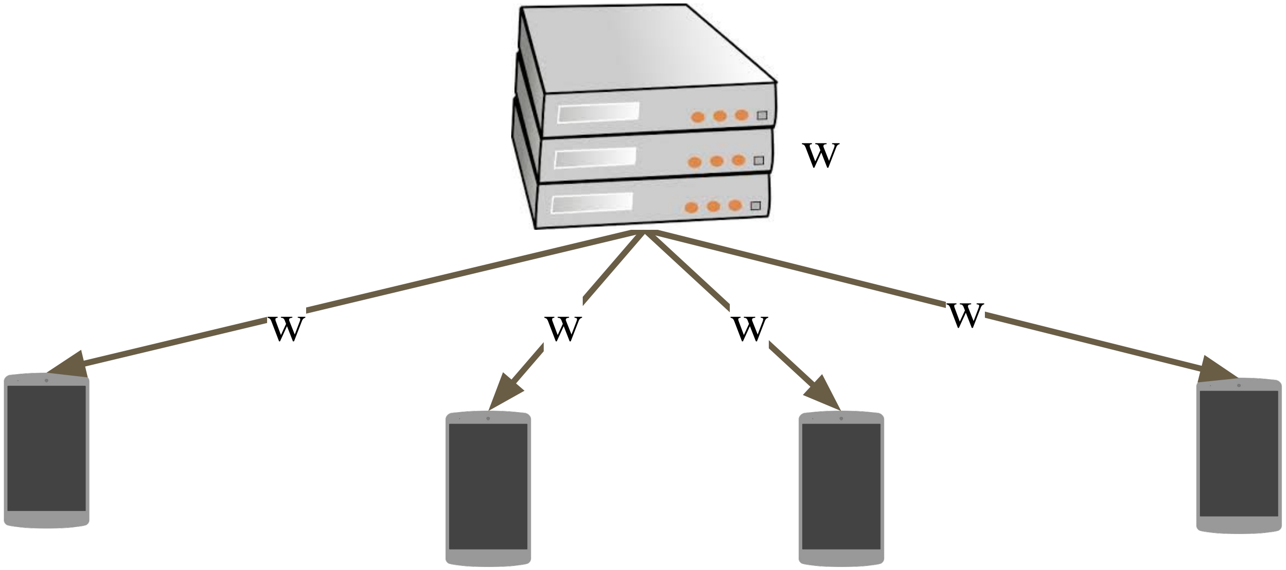
Google AI

Distributed learning with mobile devices



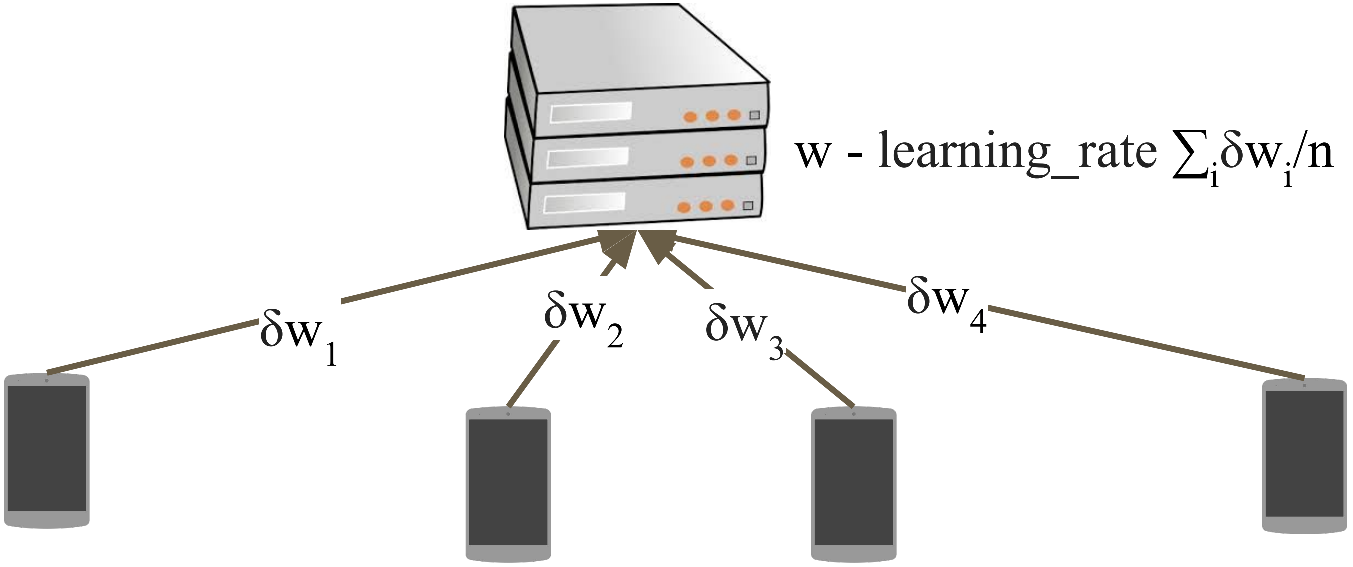
Train a centralized model; data stays on mobile phones.
In each iteration...

Server sends model to clients...



$w \in R^d$: the model vector

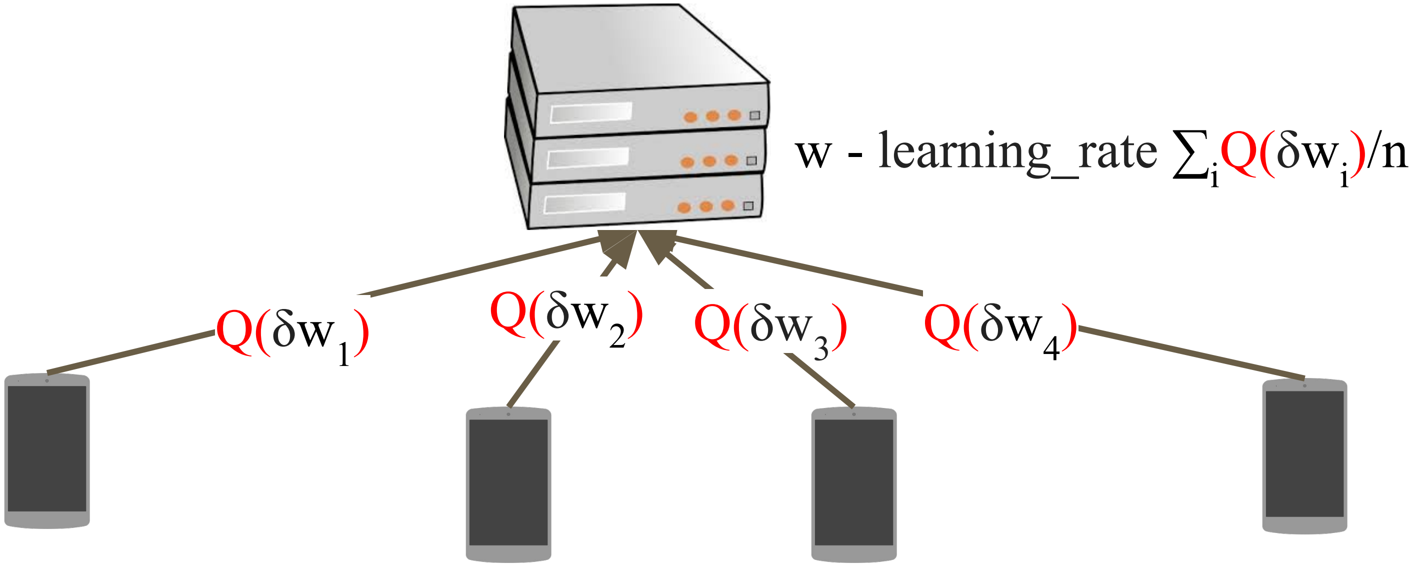
Clients send updates back...



n: number of clients

δw_i: gradient of the i-th client

Challenge I: uplink communication is expensive



- Q: quantization

How to design the quantization?

- Convergence of SGD depends on the MSE of the estimated gradient.

- Sufficient to study:

  bits vs. quantization error in distributed mean estimation.

  - No compression (float): 32 bits per coordinate; 0 MSE.
  - Binary quantization: 1 bit; O(d/n) MSE
  - Variable length coding: O(1/n) MSE
  - [Suresh et al., 17] [Alistarh et al., 17] [Wen et al., 17] [Bernstein et al., 18]

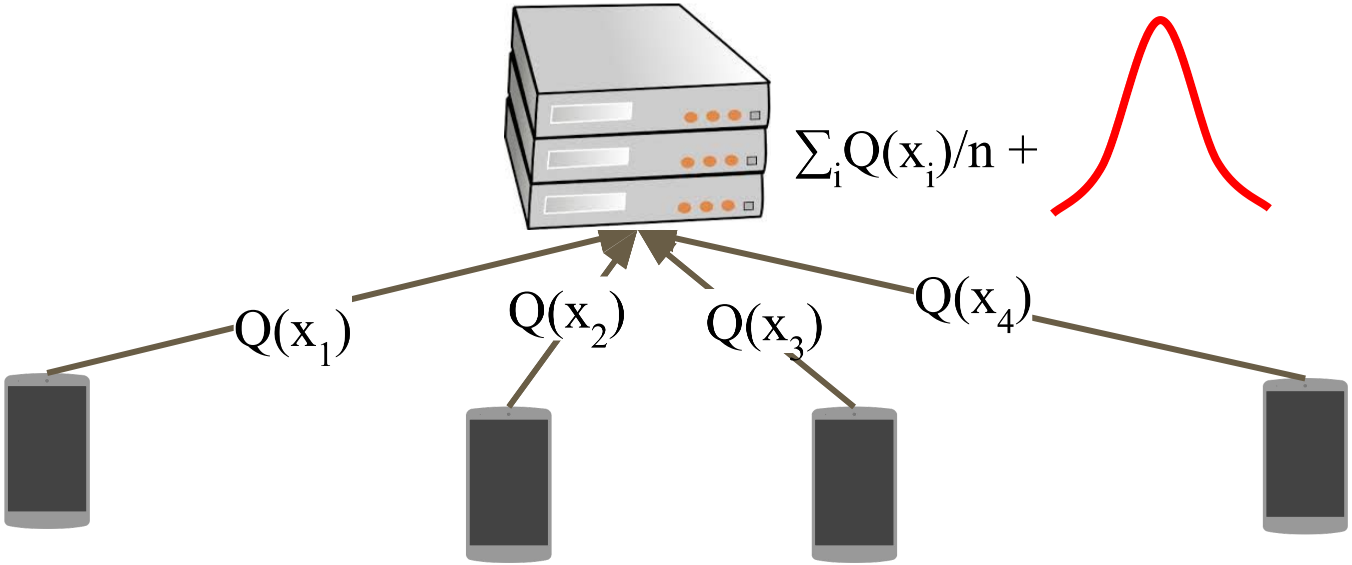
Challenge II: user privacy is important

- Differential privacy (DP)
  - Removing or changing single client's data should not result in big difference in the estimated mean
  - Adding Gaussian noise [Abadi et al., 16]

Goal of this paper

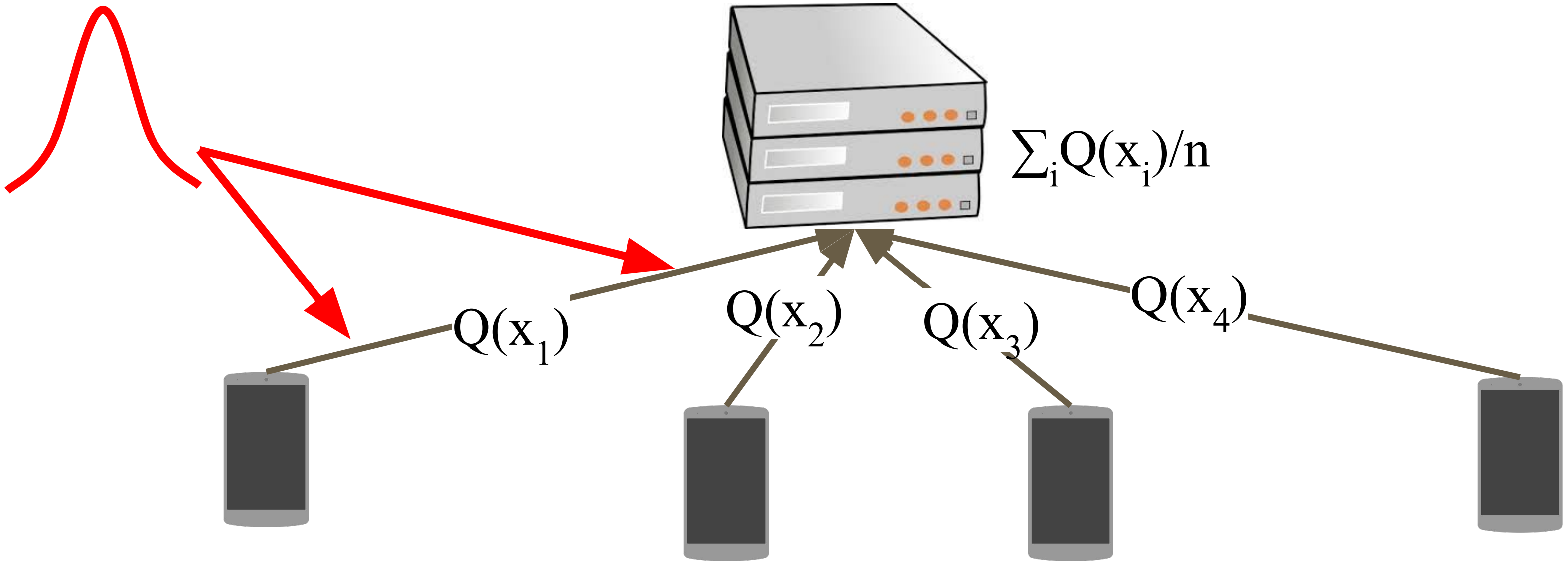
- **Both communication efficiency and differential privacy**

Attempt 1: add Gaussian noise on the server



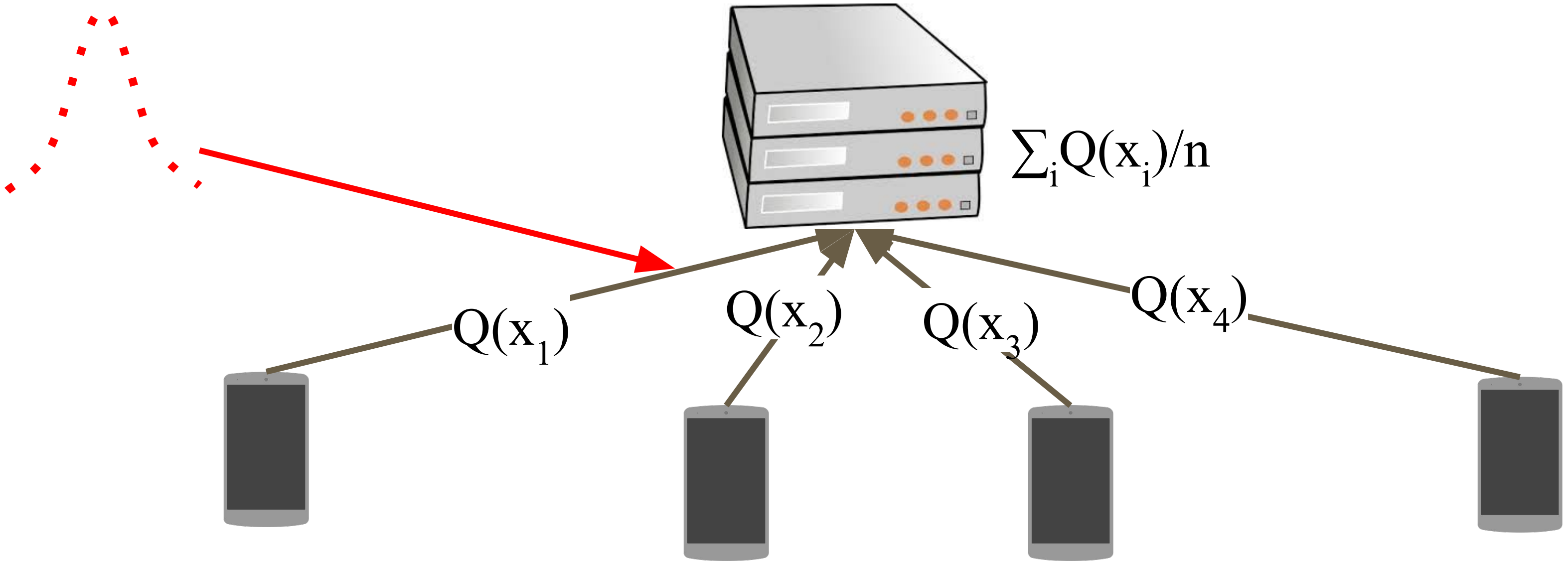
- DP results readily available
  - Assuming L2 norm of the gradient is bounded (gradient clipping).
- Server has to be trustworthy.

Attempt 2: add Gaussian noise on the client



- After quantization: no communication efficiency.
- Before quantization: hard to analyze.

cpSGD: add binomial noise after quantization

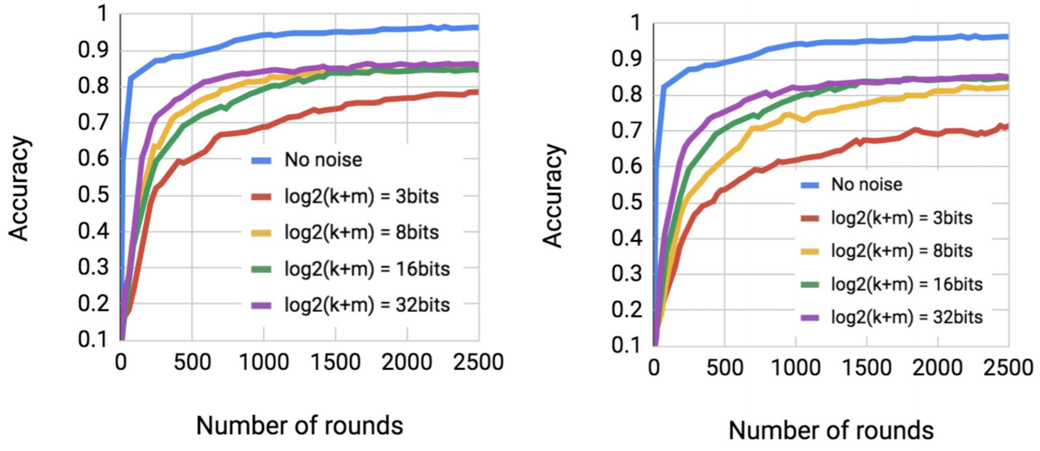


cpSGD

- Maintains communication efficiency
    - Binomial is discrete.
- Differentially private
    - Binomial similar to Gaussian.
    - Extended to d-dimension with improved bound.
- Works if server is negligent but not malicious
- Works even if clients do not trust the server
    - Secure aggregation.

For d variables and n ≈ d clients, cpSGD uses

- *O(log log(nd)) bits of communication per client per coordinate*
- *Constant privacy*



(a) $\varepsilon = 4.0$ (b) $\varepsilon = 2.0$

Tue Dec 4th 05:00 -- 07:00 PM Room 210 & 230 AB #27