Optimal Compression of Locally Differentially Private Mechanisms

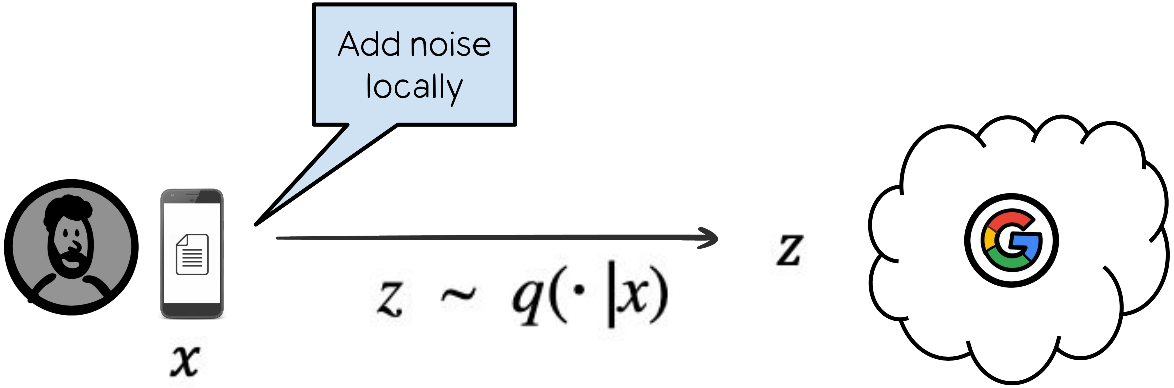
Abhin Shah

Joint work with Wei-Ning Chen, Johannes Balle, Peter Kairouz, Lucas Theis

Private and efficient distributed learning

1. Preserving the **privacy** of the user's local data
2. **Communicating** the privatized data efficiently to a central server.
3. Achieving high **accuracy** on a task (e.g., mean estimation or frequency estimation)

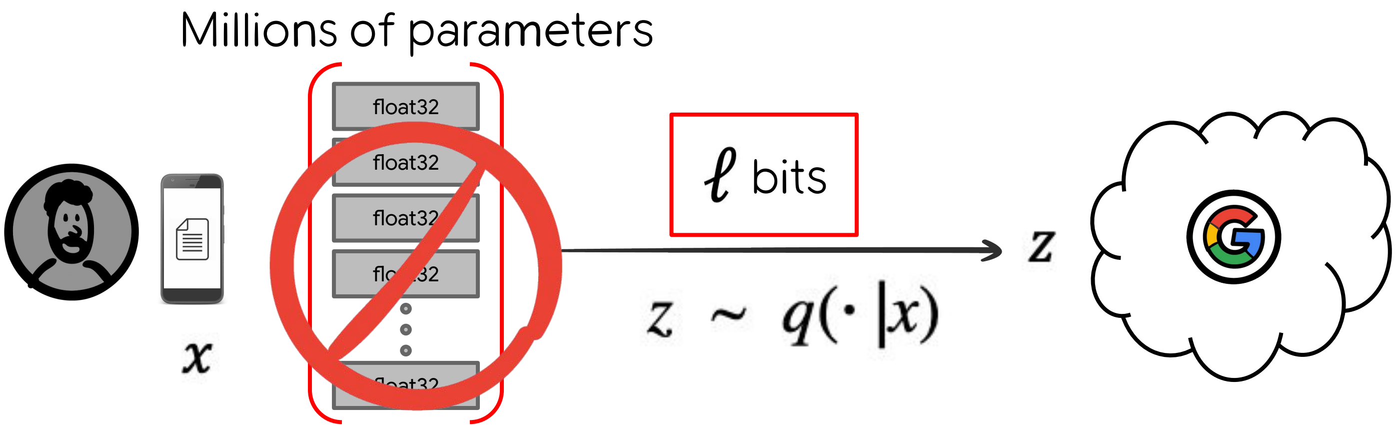
Local Differential Privacy (LDP)



$$\forall\, x,\ x',\ z,\quad \boxed{q(z|x) \le \exp(\varepsilon)\, q(z|x')}$$

Smaller ϵ ⟹ larger privacy

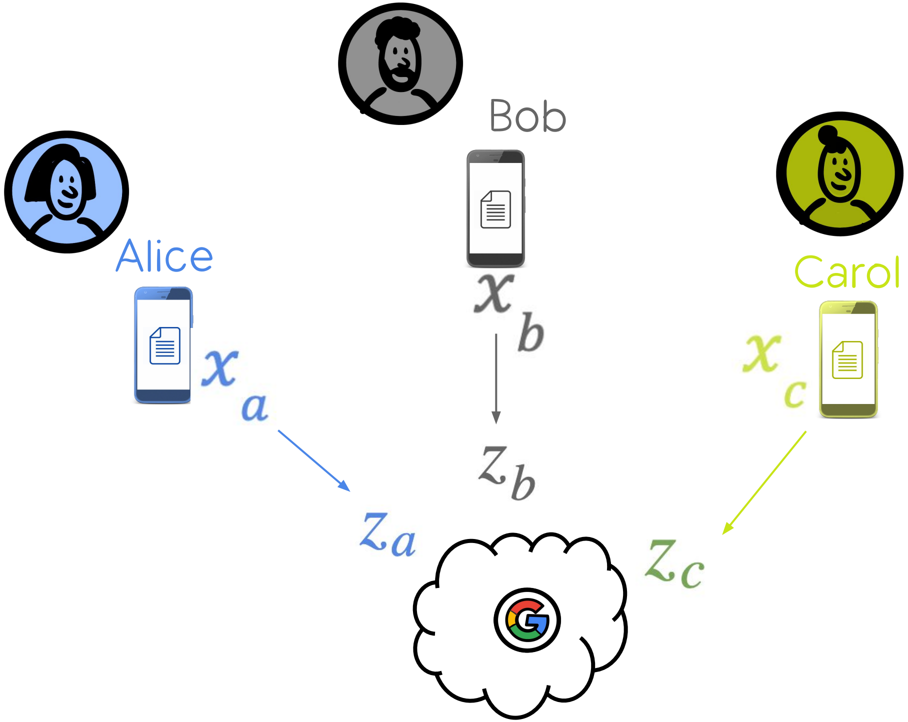
Communication cost



Mean estimation

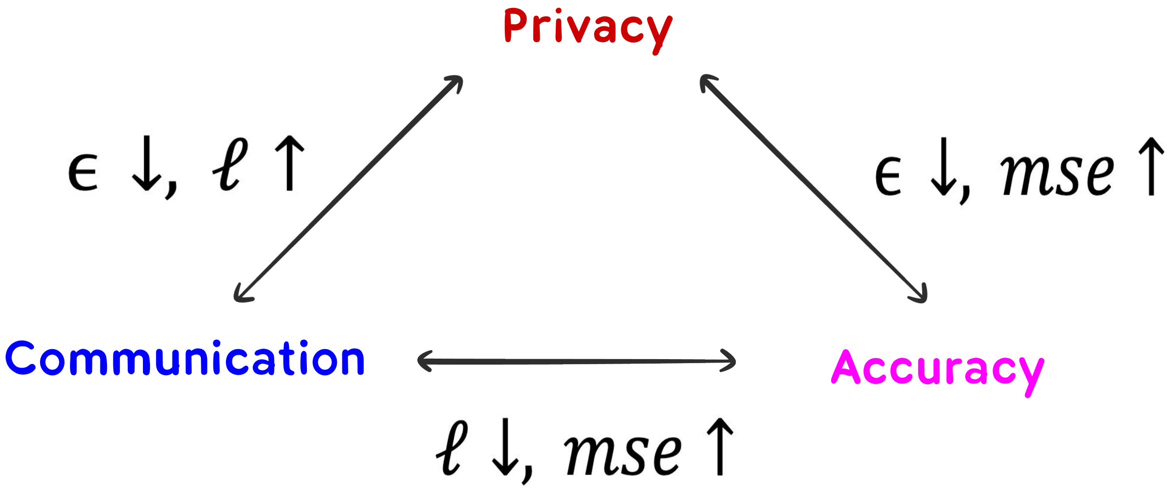
$$\mu = \frac{x_a + x_b + x_c + \cdots}{n}$$

$$\hat{\mu} = \frac{z_a + z_b + z_c + \cdots}{n}$$



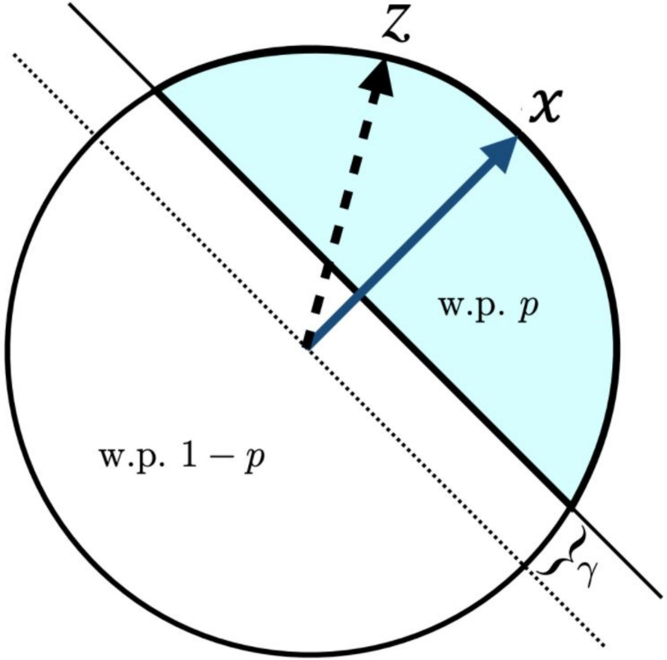
$$mse = E[\|\hat{\mu} - \mu\|^2]$$

Privacy-Accuracy-Communication tradeoffs



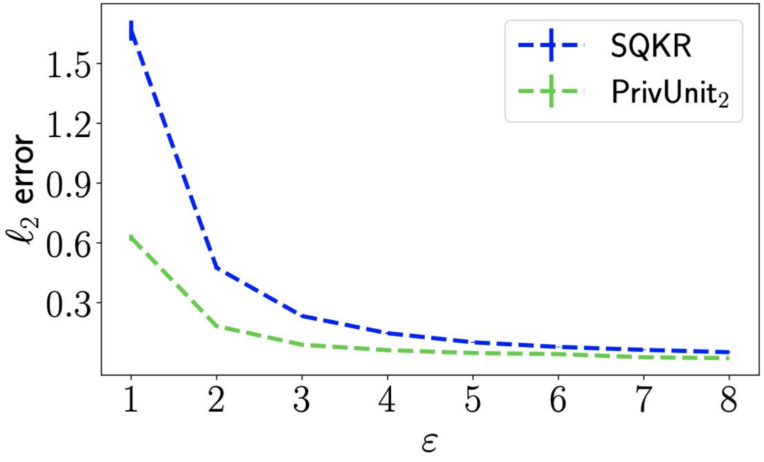
Best-known Privacy-Accuracy Tradeoff

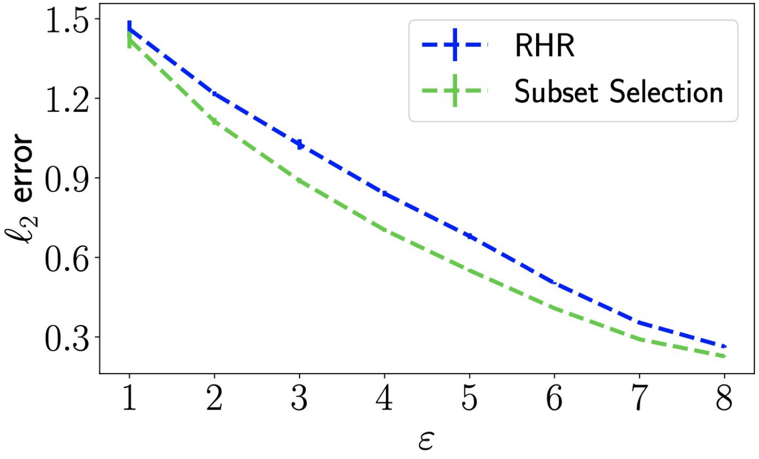
- PrivUnit and Subset Selection are the ε-LDP schemes that provide the best-known accuracy for mean estimation and frequency estimation.
- However, their communication cost scales as *O(d)*.



SQKR and RHR

- Chen et al. (2020) presented minimax order-optimal mechanisms for mean estimation (SQKR) and frequency estimation (RHR) that required only ε bits, by using shared randomness.
- However, SQKR and RHR are not competitive in terms of accuracy with PrivUnit and Subset Selection.





Main Question

Can we attain the best known accuracy under ε-LDP for mean estimation and frequency estimation while only using on the order of ε bits of communication?

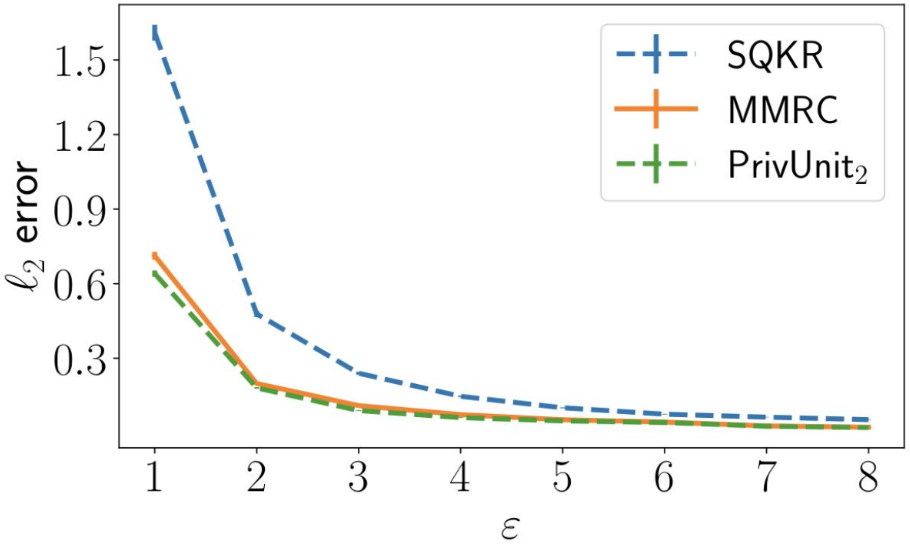
Yes! We leverage a technique based on importance sampling called Minimal Random Coding (MRC)

Pathway / Contributions

- MRC can compress any ε-LDP mechanism in a near-lossless fashion using only on the order of ε bits of communication. The resulting compressed mechanism is 2ε-LDP.
- A modified version, MMRC, can compress a large class of ε-LDP mechanisms in a near-lossless fashion using only on the order of ε bits of communication. The resulting compressed mechanism is ε-LDP.
- The class of LDP mechanisms MMRC can simulate includes the best-known schemes for mean and frequency estimation.

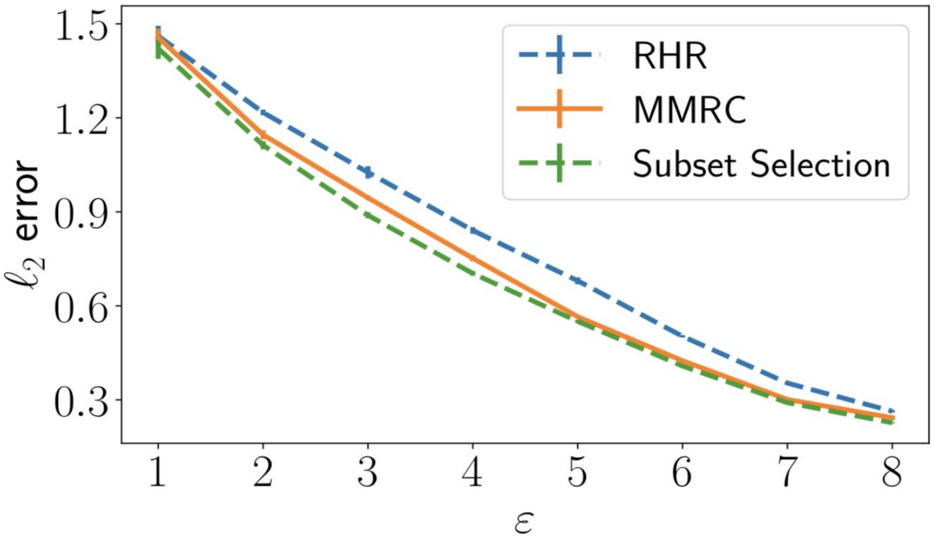
Empirical comparison

Mean estimation



d = 500, n = 5000, #bits = max{ (ε/ ln 2) + 2, 8 }

Frequency estimation



d = 500, n = 5000, #bits = max{ (ε/ ln 2) + 3, 8 }

Thank you! Please visit our poster!