EDEN: Communication-Efficient and Robust Distributed Mean Estimation for Federated Learning

Shay Vargaftik (VMware Research)

Ran Ben-Basat (University College London)

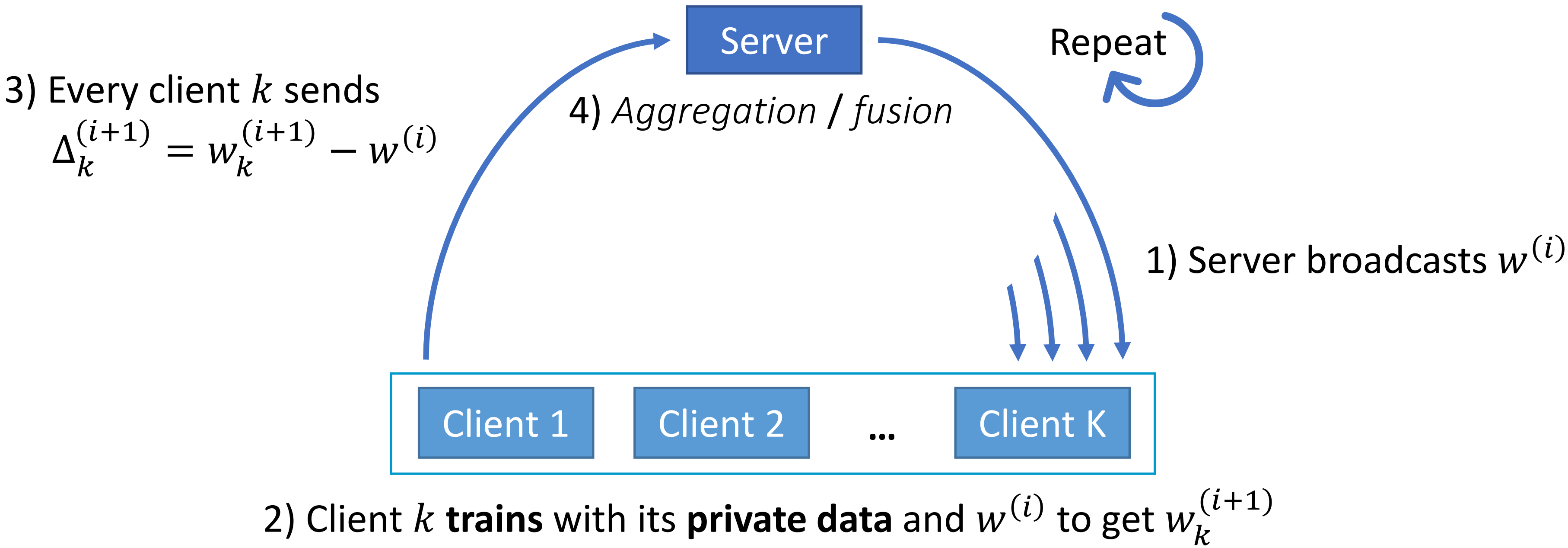
Amit Portnoy (Ben-Gurion University of the Negev)

Gal Mendelson (Stanford University)

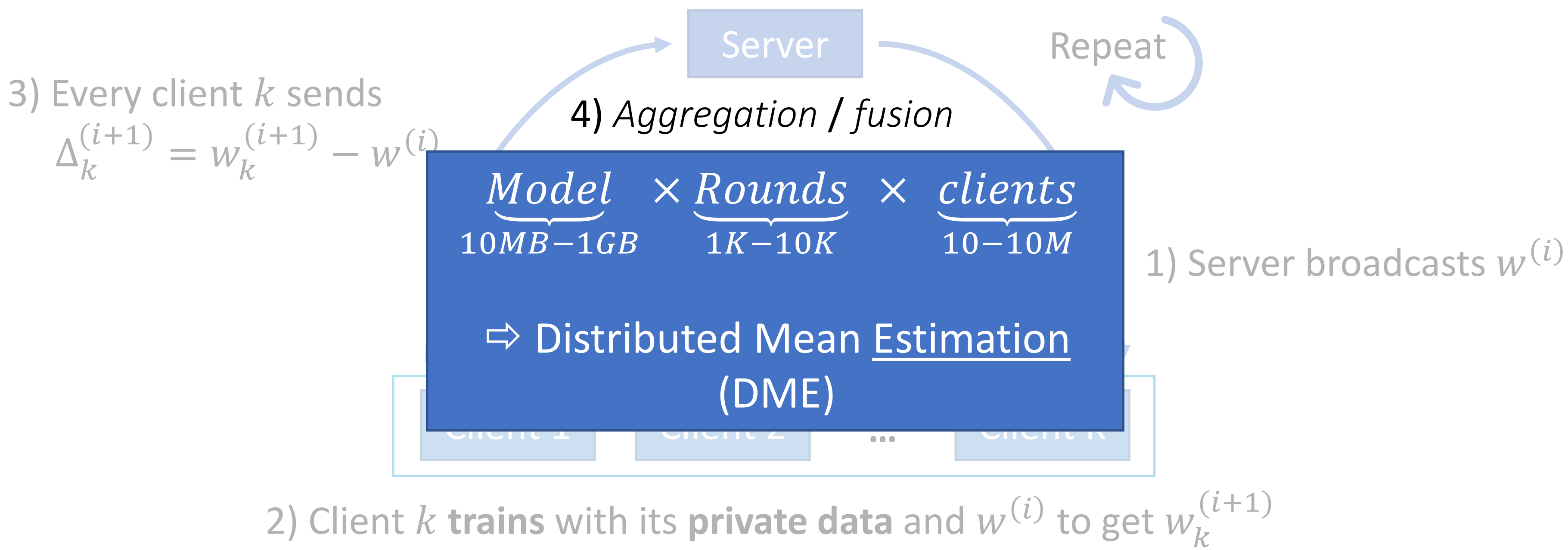
Yaniv Ben-Itzhak (VMware Research)

Michael Mitzenmacher (Harvard University)

Federated Optimization



Federated Optimization



Efficient DME for diverse Networks (EDEN)

1) First, we **randomly rotate** the input vector, $x \in \mathbb{R}^d$, after which each coordinate distribution is similar to a normal distribution:

$$\mathcal{N}\left(0, \frac{\|x\|_2^2}{d}\right)$$

2) We use this fact, multiply the post-rotation vector by $\frac{\sqrt{d}}{\|x\|_2}$, and apply a **deterministic quantization** over $\mathcal{N}(0,1)$ instead of a stochastic one.

3) With proper **scaling**, after applying the inverse rotation, the reconstructed vector, $\hat{x}$, is unbiased, and the eventual distributed mean estimation is more accurate.

Efficient DME for diverse Networks (EDEN)

We define a post-rotation quantization, $\mathcal{Q}$, as follows:

Given **a set of intervals**, $\mathcal{I}$, that constitutes a symmetric partition of $\mathbb{R}$.

Each coordinate, $a \in I \in \mathcal{I}$, is (deterministically) quantized into I's center of mass:

$$\mathcal{Q}(a) = \mathbb{E}[\ z \ | \ z \in I \ \ and \ \ z \sim \mathcal{N}(0,1)\ ]$$

<u>Key theoretical result</u>

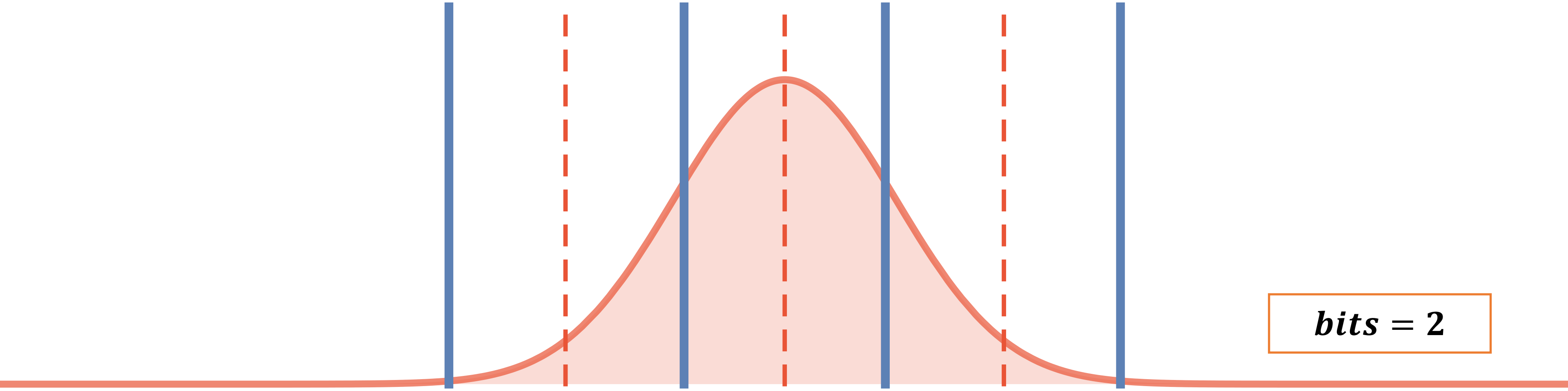
$$\frac{\mathbb{E}[\|x - \hat{x}\|_2^2]}{\|x\|_2^2} \leq \frac{1}{1 - \mathbb{E}\left[\left(\boldsymbol{z} - \boldsymbol{\mathcal{Q}}(\boldsymbol{z})\right)^2\right]} - 1$$

where $d \to \infty$ and $z \sim \mathcal{N}(0,1)$

Efficient DME for diverse Networks (EDEN)

Two principal quantization strategies:

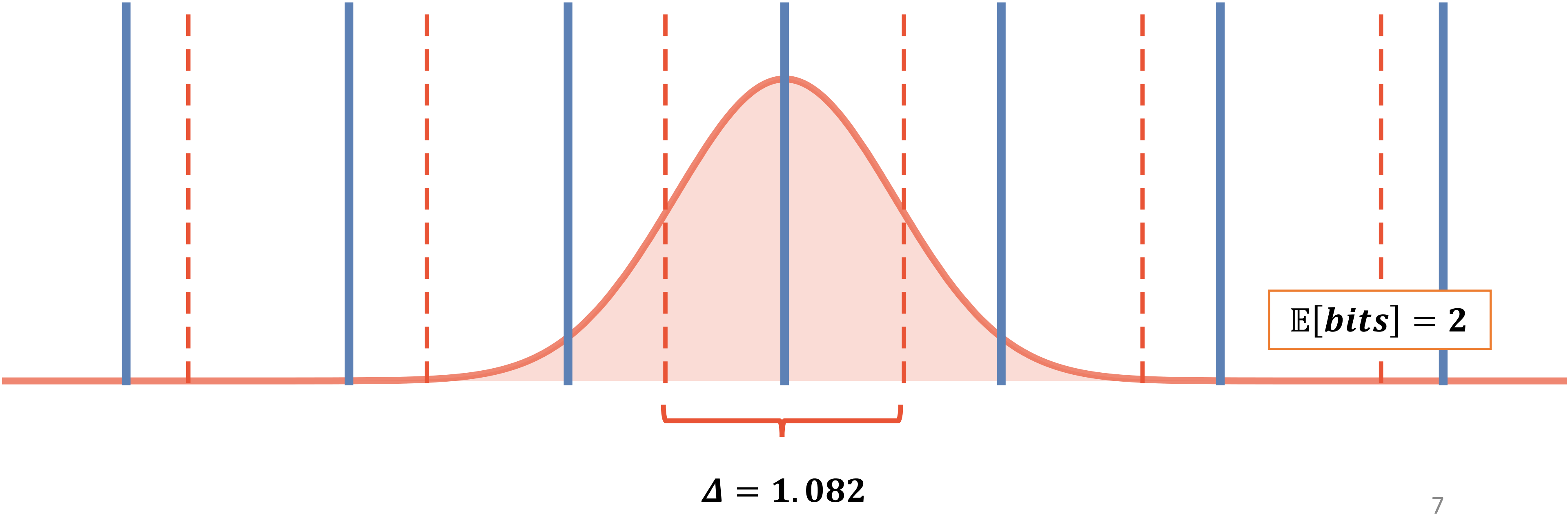
1) Lloyd-Max provides the **optimal** quantization for a **fixed** number of bits by minimizing $\mathbb{E}\left[\left(z - Q(z)\right)^2\right]$.



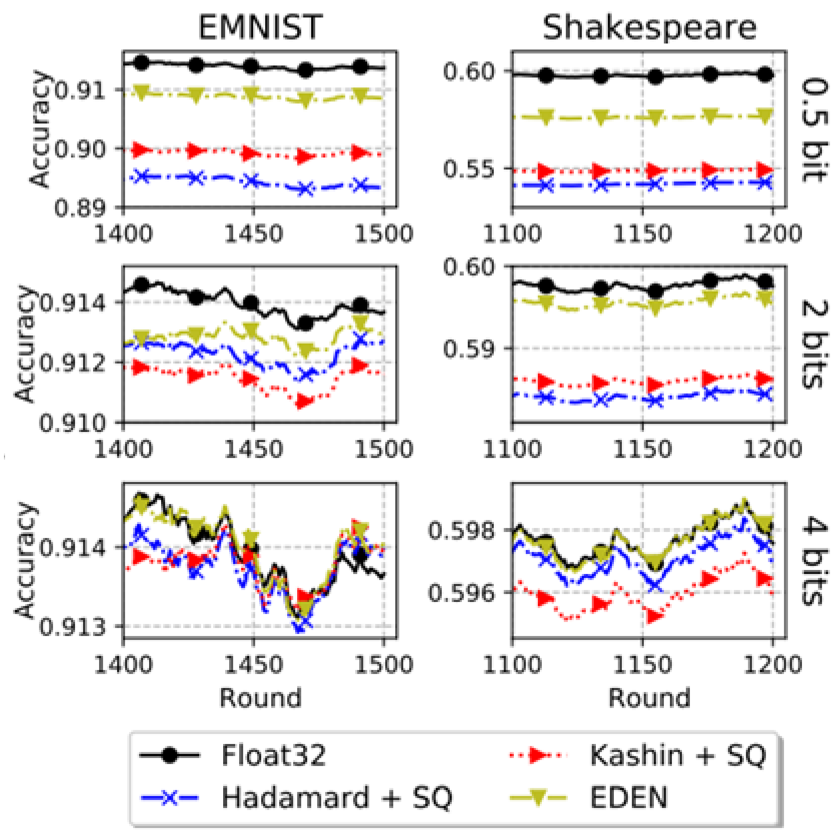
Efficient DME for diverse Networks (EDEN)

<u>Two principal quantization strategies:</u>

2) When considering **variable length coding**, we use Shannon's source coding theorem to determine the **expected** number of bits (i.e., the *entropy*).
Given a bit budget, b, we search for the smallest Δ that achieves entropy ≤ b.



Federated learning benchmarks



(Additional evaluation in the paper)

Efficient DME for diverse Networks (EDEN)

More in the paper:

- Heterogeneous sender communication budgets.
- Handling packet losses.
- Support non-integer number of bits per coordinate.
- The Hadamard transform as a rotation.
- Full source code is available at: https://github.com/amitport/EDEN-Distributed-Mean-Estimation.

Thank You!