On the Power of Multiple Anonymous Messages

Badih Ghazi Noah Golowich Ravi Kumar
Rasmus Pagh Ameya Velingker

Outline

- Review of central and local models of DP.
- Shuffled model of DP [Bittau et al., ‘16].
- Lower bounds for frequency estimation and selection in *single-message* shuffled model [this paper].
  - ”Privacy amplification by shuffling” is optimal.
- Upper bounds for frequency estimation in *multi-message* shuffled model [this paper].
  - Communication-efficient protocols.

Differential privacy – **central** model

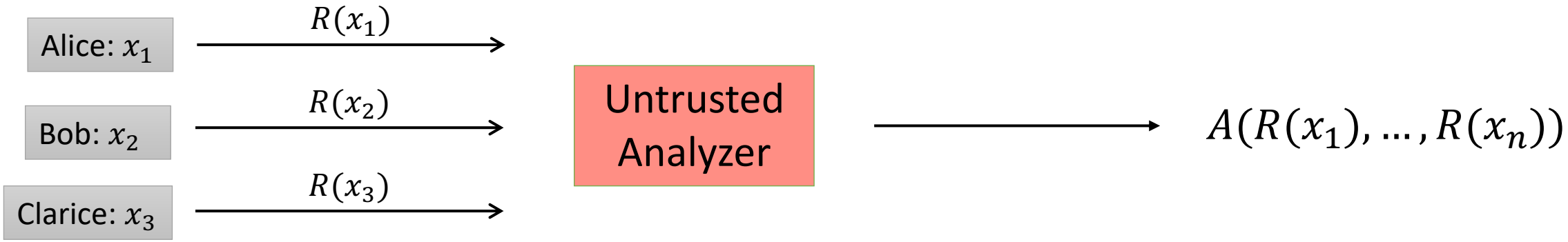
- **Universe** U, n users, dataset $X = (x_1, \dots, x_n) \in U^n$.
- *Trusted* analyzer adds noise to preserve privacy:

Definition: Algorithm A is **(ϵ, δ)-differentially private (DP)** if for all events T, for all *neighboring datasets* $X = (x_1, \dots, x_i, \dots, x_n), X' = (x_1, \dots, x_i', \dots, x_n)$,

$$\Pr_A[A(X) \in T] \leq e^{\epsilon} \cdot \Pr_A[A(X') \in T] + \delta$$

Differential privacy – **local** model

- Analyzer *untrusted* – users must add privacy preserving noise:
- R is **local randomizer**; users send $R(x_i)$ to analyzer.



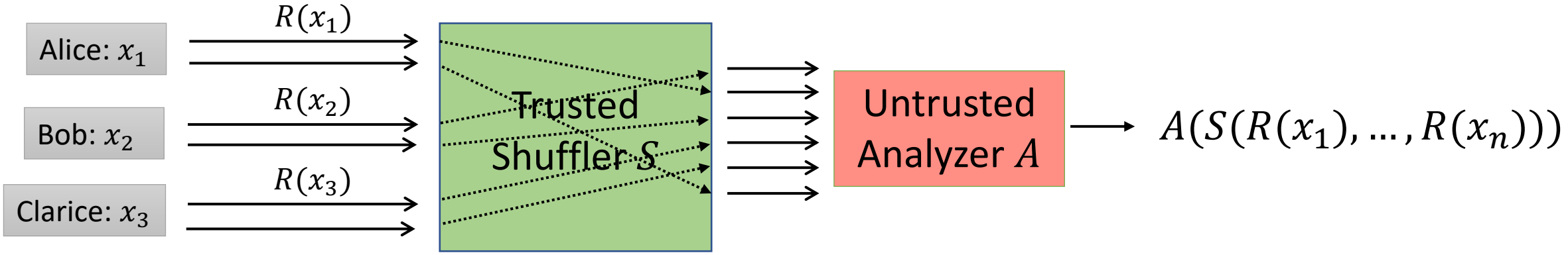
Definition: Algorithm is **(ϵ, δ)-differentially private (DP) in local model** if

$$x \mapsto R(x)$$

is (ϵ, δ)-differentially private.

Differential privacy – **shuffled** model [Bittau et al., '16]

- Local model necessitates large noise, so large error (often $\Omega(\sqrt{n})$).
- Solution: randomly permute messages (i.e., make them anonymous).



Definition: Algorithm (R, S) is **(ϵ, δ)-differentially private (DP) in shuffled model** if

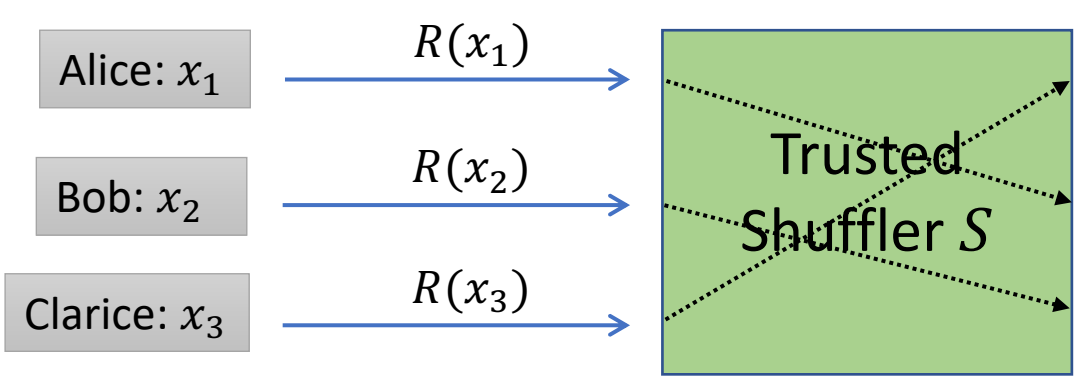
$$(x_1, \dots, x_n) \mapsto S\big(R(x_1), \dots, R(x_n)\big)$$

is (ϵ, δ)-differentially private.

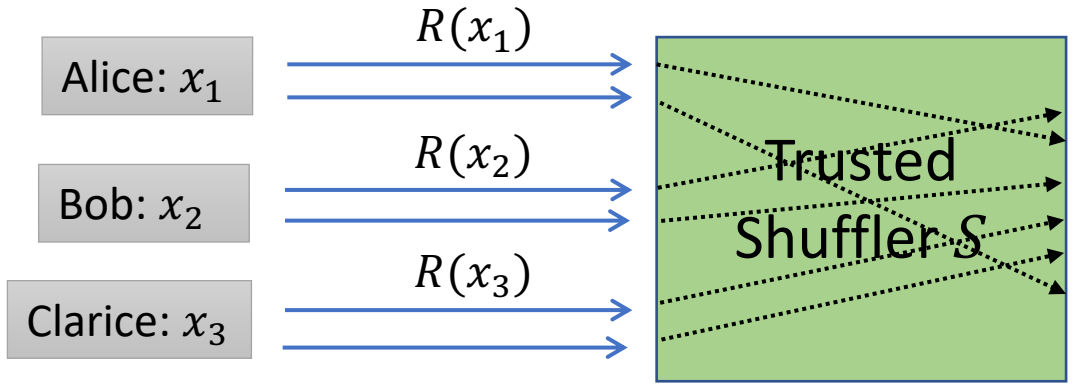
Single vs multi-message shuffled model

- Formally: $R(x_i)$ outputs m messages, S applies random permutation to $m \cdot n$ messages:

Single-message shuffled model: $m = 1$



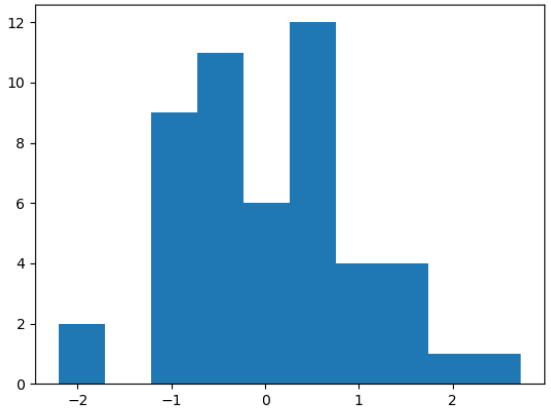
Multi-message shuffled model: $m > 1$



The power of a **single** anonymous message

- "Privacy amplification by shuffling" [Erlingsson et al., '18; Balle et al., '19]:
  - If R is $(\epsilon_L, 0)$-differentially private in *local model*, then:
  - (S, R) is (ϵ_S, δ)-differentially private in *shuffled model* for $\epsilon_S \ll \epsilon_L$.
- Application: frequency estimation:
  - $x_1, \ldots, x_n \in U := \{1, 2, \ldots, B\}$, goal is:
  - Compute *frequency* of each $j \in \{1, \ldots, B\}$,
    i.e., $\#\{i : x_i = j\}$.
- Measure error by *additive error:*

  Error := $\max_j$ |True freq(j) – Estimated freq(j)| $\leq n$.



Optimality of amplification by shuffling

- Using single-message randomizers w/ amplification by shuffling: can perform frequency estimation on $\{1, \dots, B\}$ with error:

$$\min \{\underbrace{\tilde{O}(n^{1/4})}_{R = \text{RAPPOR}}, \underbrace{\tilde{O}(B^{1/2})}_{R = \text{Randomized Resp.}}\}$$

[Erlingsson et al., '14] [Warner, '65]

- We show this is optimal:

Theorem:
Any $(1, o(1/n))$-DP *single-message* shuffled protocol for frequency estimation has additive error $\min\{\widetilde{\Omega}(n^{1/4}), \widetilde{\Omega}(B^{1/2})\}$.

Lower bound for **low-privacy** local DP frequency estimation

- By [Cheu et al., '18]: (ϵ_S, δ)-DP in 1-msg shuffled model $\Rightarrow$ $(\epsilon_S + \ln n, \delta)$-DP in local model; so, main component of proof is:

Theorem

If $\frac{2 \ln n}{3} \lesssim \epsilon_L \lesssim 2 \cdot \min\{\ln n, \ln B\}$ and $\delta_L < o(1/n \ln n)$, then the error of any (ϵ_L, δ_L)-locally DP frequency estimation protocol is:

$$\widetilde{\Omega}\left(\frac{1}{\sqrt{n} \cdot e^{\epsilon_L/4}}\right)$$

- Prior work [Duchi et al., '18]: lower bound of $\Omega\left(\frac{1}{\sqrt{n} \cdot e^{\epsilon_L}}\right)$.
  - Also, only applied to pure DP, i.e., $\delta_L = 0$.

Proof idea of local DP lower bound

- Fano's method: for $\alpha \approx$ {desired lower bound on error}:

$$V \sim Unif([B]) \xrightarrow{\text{perturbed}} X = \begin{cases} V \text{ with prob. } \alpha. \\ Unif([B]) \text{ with prob. } 1-\alpha \end{cases}$$

- We show:

$$I\big(V; R(X)\big) \leq \tilde{O}(\alpha^4 n\, e^{\epsilon_L})$$

- Above is actually **false** for general randomizers R; we have to use that R leads to a protocol with error $\lesssim \alpha$.

Outline

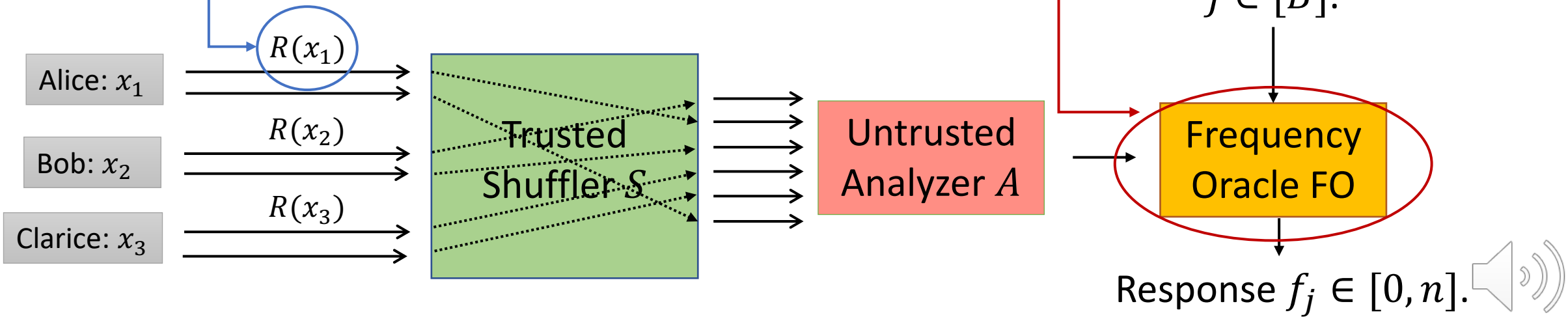
- Review of central and local models of DP.
- Shuffled model of DP [Bittau et al., ‘16].
- Lower bounds for frequency estimation and selection in *single-message* shuffled model [this paper].
  - ”Amplification by shuffling” is optimal.
- **Upper bounds for frequency estimation in *multi-message* shuffled model [this paper].**
  - **Communication-efficient protocols.**

Efficiency in the multi-message shuffled model

- 2 measures of efficiency: communication & computation.
- To define computation: issue that $B \gg n$, so infeasible to compute all frequencies of all $j \in [B]$.
  - Solution: have analyzer output **frequency oracle** FO, measure query time.

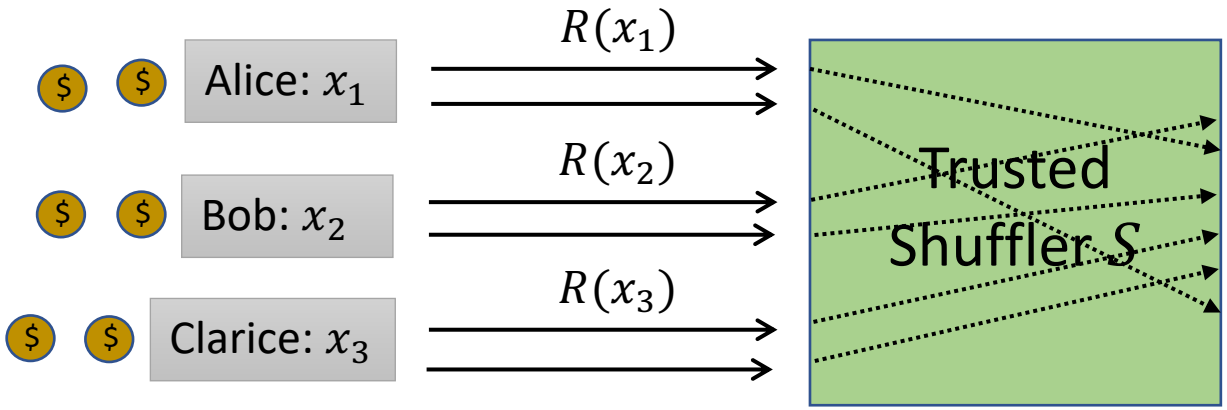
Communication: total length of all m messages output by 1 user.

Query time: time taken by a single query FO(j) = ?



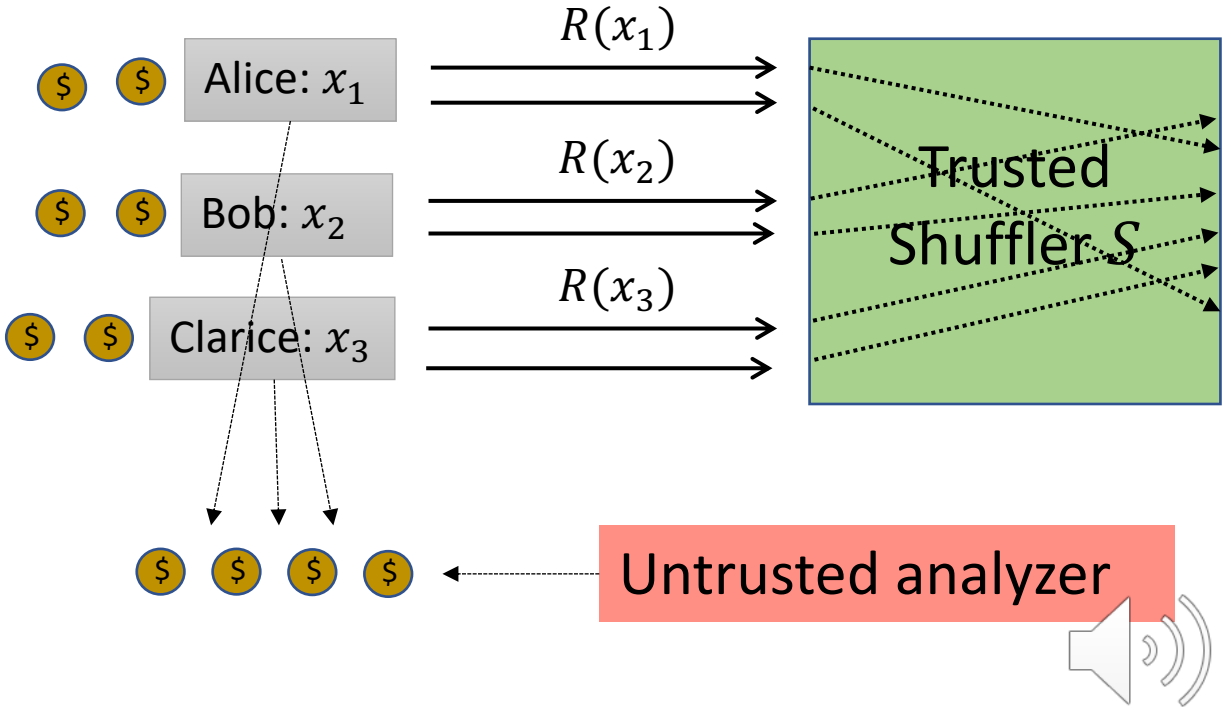
One more resource: public coins

Private coins:



Public coins:

(users additionally get common public coins that adversary can see too)



Efficient frequency estimation in multi-message shuffled model

Theorem

There exist multi-message shuffled model protocols satisfying (ϵ, δ)-differential privacy with the below properties:

Technique	Error	Communication	Query time	Public coins
Count-min	$\tilde{O}(1)$	$\tilde{O}(1)$	$\tilde{O}(1)$	Yes
Hadamard Response	$\tilde{O}(1)$	$\tilde{O}(1)$	$\tilde{O}(n)$	No

Idea for multi-message upper bounds

1. View frequency estimation as B parallel aggregation problems
   - I.e., estimate number of users holding each $j \in [B]$.
   - Exists local randomizer R_{agg} that perform aggregation with error $\tilde{O}(1)$ in shuffled model [Cheu et al., '18].

$$R(3) = \underset{j=1}{R_{agg}(0)} \quad \underset{j=2}{R_{agg}(0)} \quad \underset{j=3}{R_{agg}(1)} \quad \underset{j=4}{R_{agg}(0)} \quad \ldots$$

2. Use count-min sketch or Hadamard response to avoid having to send B separate messages (one for each j).

Hadamard response [Kairous et al., ‘16]; [Nguyen et al., ‘16]; [Acharya et al., ‘18]

- Prior work using Hadamard response for local DP:
- Let $H \in \{-1, 1\}^{B \times (B+1)}$ be Hadamard matrix w/ first row removed
- For $x \in [B]$, $R(x) \in [B+1]$ distributed as:

$$R(x) = \begin{cases} \text{Uniform over } \{k \text{ s.t. } H_{x,k} = 1\} \text{ with probability } \frac{e^{\epsilon}}{1+e^{\epsilon}} \\ \text{Uniform over } \{k \text{ s.t. } H_{x,k} = -1\} \text{ with probability } \frac{1}{1+e^{\epsilon}} \end{cases}$$

- Gets error $O\left(\frac{\sqrt{n \log B}}{\epsilon}\right)$ in local model.

Hadamard response for shuffled model

- Given input $x \in [B]$:
- 1 "signal" msg: concatenate $\log n$ indices of columns of H:

$$m = \left(m_1, \dots, m_{\log n}\right): \; m_t \sim Unif\left(\left\{k : H_{x,k} = 1\right\}\right)$$

- $\rho = O(\log(1/\delta)/\epsilon^2)$ "noise" messages:

$$\widetilde{m^{(1)}}, \dots, \widetilde{m^{(\rho)}} \sim Unif\left([B+1]^{\log n}\right)$$

- Local randomizer: $R(x) = \left(m, \widetilde{m^{(1)}}, \dots, \widetilde{m^{(\rho)}}\right)$.
- Privacy intuition: number of noise messages $\widetilde{m^{(l)}}$ consistent with any row x of H is Binom$(n\rho, \frac{1}{n})$ which is "smooth".

Additional results in our paper

- Tight (up to log factor) lower bound of $\Omega(d)$ on sample complexity for selection on d elements in the single-message shuffled model.
- Corollary of Hadamard response protocol for efficiently implementing ”sparse” non-adaptive statistical query algorithms in shuffled model.
- Applications of upper bounds for frequency estimation to range queries, quantile estimation.
- https://arxiv.org/pdf/1908.11358.pdf

Open problems

- Can one decrease the query time of $\tilde{O}(n)$ (achieved by Hadamard response) for private-coin frequency estimation in multi-message shuffled model?
- Selection on d elements in the multi-message model:
  - There is a multi-message protocol with sample complexity $O(\sqrt{d})$ [Cheu et al., '18].
  - (We showed in single-message shuffled model sample complexity is $\Omega(d)$.)
  - Can one do better? (In central model can get $O(\log d)$ sample complexity.)

Thank you!