Framework for Differentially Private Data Analysis with Multiple Accuracy Requirements

Karl Knopf [kknopf@uwaterloo.ca]
University of Waterloo

MOTIVATION

Multiple analysts may want to access to sensitive datasets but they can have competing requirements

Multiple Research Teams with Different Error Tolerances

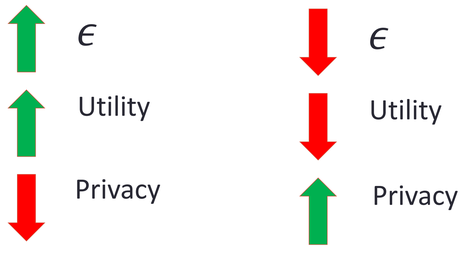
Multiple Data Products with Different Utility Requirements

DIFFERENTIAL PRIVACY

Differential privacy (DP) [2] is defined as:

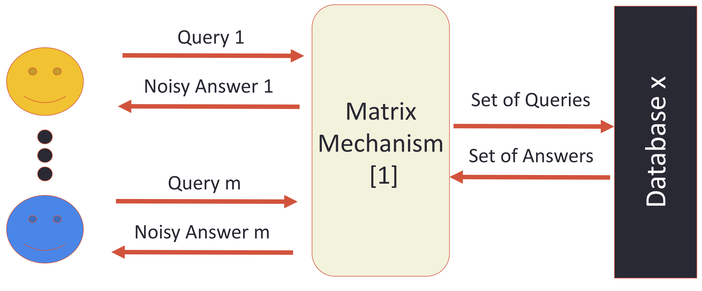
A randomized mechanism $\mathcal{M}$ is (ϵ, δ) −differentially private if for all $\mathcal{S} \in \text{Range}(\mathcal{M})$ and any pair of neighboring databases $(\mathcal{D}, \mathcal{D}')$ that differ in a single element:

$$[\mathcal{M}(\mathcal{D}) \in \mathcal{S}] \leq e^{\epsilon}\Pr[\mathcal{M}(\mathcal{D}') \in \mathcal{S}] + \delta$$



PRIOR MECHANISMS

Only one utility metric!



FRAMEWORK

How can we optimally and equitably support analyses with different utility requirements on a common sensitive dataset given a fixed privacy parameter?

An optimization problem to find the best DP mechanism A:

The Objective: Maximize equitability criterion — $\text{maximize}_{A,C}\, f(c_i)$

Constraint 1: Mechanism answers the queries — subject to $c_i w_i = c_i w_i A^+ A$,

Constraint 2: Accuracy is bounded — $\frac{2c_i}{\epsilon^2} \|A\|^2 \|w_i A^+\|_F^2 \leq \beta_i$

EQUITABILITY CRITERIA

1. Equality between Queries: Maximize the number of answered queries

$$f_1(c) = \sum_{i=1}^{n} c_i$$

2. Equality between Groups: Maximize the number of groups of queries that are answered completely

$$\mathrm{f}_2(c) = \max\left(\sum_{g_j \in G} I_1\left(\sum_{c_i \in g_j} \frac{c_i}{|g_j|}\right)\right)$$

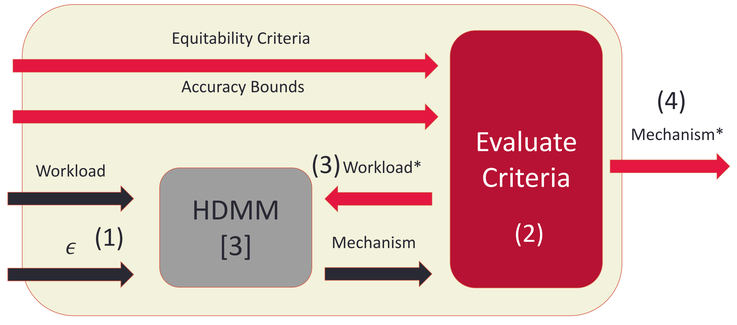
$$\text{where } I_1(x) = \begin{cases} 1 \; if \; x = 1, \\ 0 \; if \; x \neq 1 \end{cases}$$

3. Equitable rate of accurate queries per Group: Maximize the minimum rate of any set of queries

Define the rate for g_j as: $\sum_{c_i \in g_j} \frac{c_i}{|g_j|}$.

Then $\mathrm{f}_3(c) = \max\left(\min_{j=1,\ldots,k} \sum_{c_i \in g_j} \frac{c_i}{|g_j|}\right)$

OUR SOLUTION



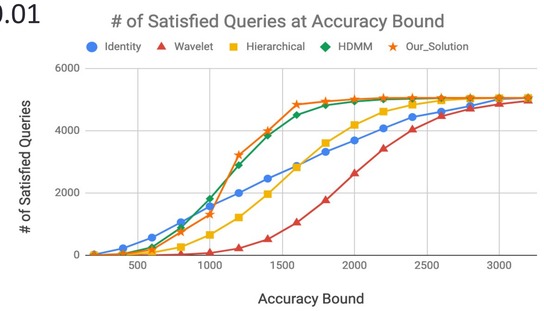
EXPERIMENTS

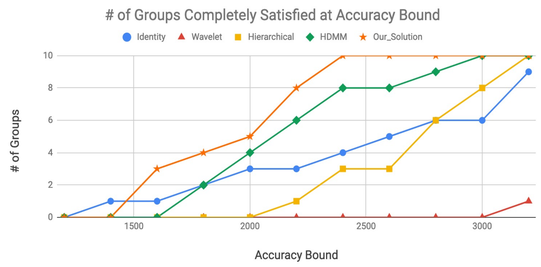
We compare our solution on all three equitability criteria to prior mechanisms, with the following parameters:

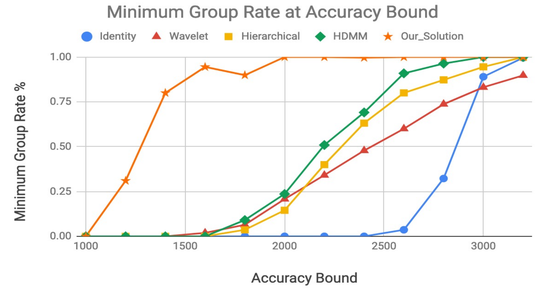
W = Full Range [0,100]

G = Ranges of [0,10),[10,20), … [90,100]

ϵ = 0.01







CONCLUSION

It is unclear which prior solution offers the best utility when individual query accuracy bounds are considered!

Our solution can find a more equitable mechanism under our criteria for universal query accuracy bounds.

References

[1] Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. The matrix mechanism: optimizing linear counting queries under differential privacy. The VLDB journal24, 6 (2015), 757–781.

[2] Cynthia Dwork, Aaron Roth, et al.2014. The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science9, 3–4(2014), 211–407

[3] Ryan McKenna, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala.2018. Optimizing error of high-dimensional statistical queries under differential privacy. Proceedings of the VLDB Endowment11, 10 (2018), 1206–1219