18734: Foundations of Privacy

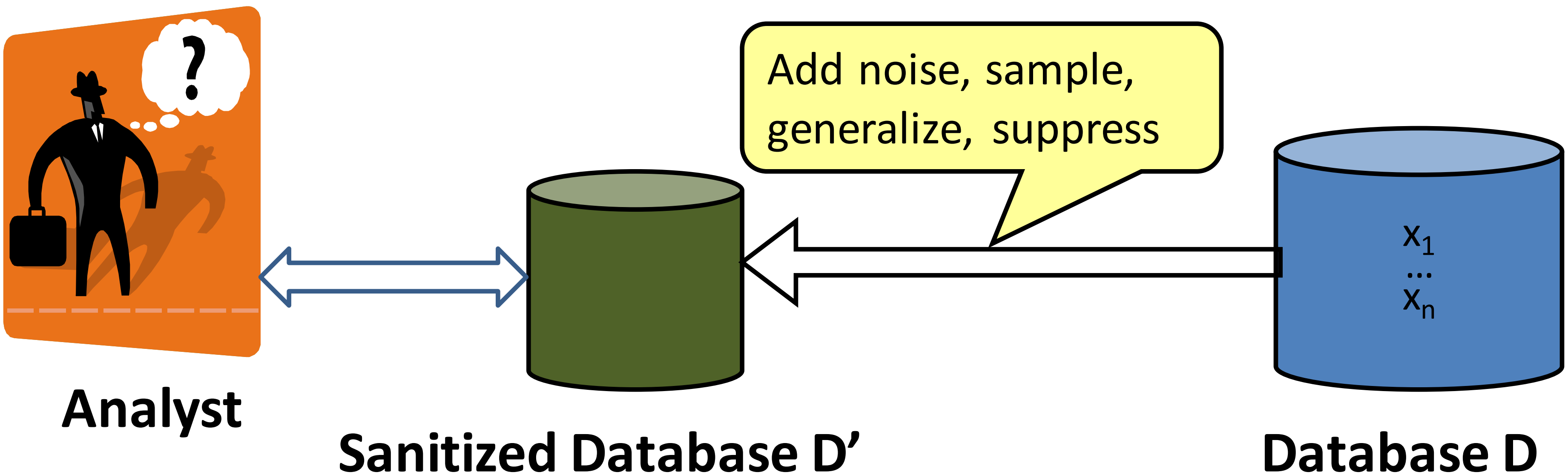
Differential Privacy: The Non-Interactive Setting

Anupam Datta

CMU

Fall 2016

Privacy-Preserving Statistics: Non-Interactive Setting



Goals:

- Release sanitized database that answers all queries from a class accurately
- Preserve differential privacy

Example Class: Interval Queries

- Database D of n points in [0,1] discretized to b bits of precision
- Interval indicator function

$$I_{a_1,a_2}(x) = \begin{cases} 1, & a_1 \leq x \leq a_2; \\ 0, & \text{otherwise}. \end{cases}$$

- Interval query

$$Q_{[a_1,a_2]}(D) = \sum_{x \in D} \frac{I_{a_1,a_2}(x)}{|D|}.$$

- Example: grade distribution in a course

Goal

- Design efficient mechanism to release sanitized dataset D' from D s.t. it is
  - Useful, i.e. answers all interval queries accurately with high probability
  - Preserves differential privacy

DEFINITION 2.6 (USEFULNESS DEFINITION 1). *A database mechanism A is (ϵ, δ)*-useful *for queries in class C if with probability $1 - \delta$, for every $Q \in C$ and every database D, for $\widehat{D} = A(D)$, $|Q(\widehat{D}) - Q(D)| \leq \epsilon$.*

Mechanism(roughly)

- Given D, perform a number of α'-differentially private interval queries to partition [0, 1] into sub-intervals containing probability mass in the range $[\epsilon_1/2 - \epsilon_2, \epsilon_1/2 + \epsilon_2]$

- Output dataset that has $(\epsilon_1/2) \cdot n$ points in each of these intervals

Theorem

THEOREM 4.2. *With* $\alpha' = (\epsilon\alpha)/4b$, $\epsilon_1 = (\epsilon/2)$ *and* $\epsilon_2 = (\epsilon^2/8)$, *the above mechanism preserves* α*-differential privacy while being* (ϵ, δ)*-useful for the class of interval queries given a database of size:*

$$|D| \geq O\left(\frac{b(\log b + \log(1/\epsilon\delta))}{\alpha\epsilon^3}\right)$$

More general results

- A Learning Theory Approach to Non-Interactive Database Privacy. Avrim Blum, Katrina Ligett, Aaron Roth. In the proceedings of STOC 2008: The 40th ACM Symposium on the Theory of Computing.