CSci 8271
Security and Privacy in Computing
Day 9: Local Differential Privacy with RAPPOR

Stephen McCamant
University of Minnesota

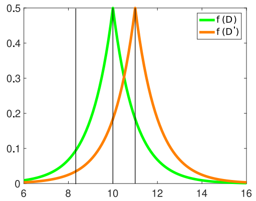
Protecting database privacy

- Can we protect sensitive collected information to allow statistics
  - By transforming data, and/or restricting queries
- A weakness of many early attempts was breaking depending on extra information an attacker might have

Centralized differential privacy

- If the result of an aggregation would be the same without my input, it can't be hurting my privacy
- Can't make it exactly the same, but ensure it's similar by adding randomness
- $P[A(V) \in R] \leq e^{\epsilon} P[A(V') \in R]$

Laplace mechanism



Distribution of adding Laplacian noise:
f(D) = 10 + noise, f(D') = 11 + noise

Epsilon and delta

- The parameter ϵ represents a privacy budget
  - Often little specific guidance on choosing it
  - In an interactive system, it can run out
- (ϵ, δ) differential privacy also allows a possibility of complete failure

Local differential privacy

- If no trusted third party, data owners must each add their own noise
- Allows more applications, but has a worse privacy/utility tradeoff

Randomized response intuition

- Earlier proposed for embarrassing survey questions
- Randomly choose to answer either randomly or honestly
- The effect of the random answers can be removed after aggregation
- But no one can tell for sure about any particular response

Permanent response

- Repeatedly adding different noise to the same honest value would give it away
- So, add one level of noise permanently, and save the result
- Still not enough to protect "what is your age in days today?"

Instantaneous response

- A second layer of randomization makes each repeated response different
  - Avoid tracking, and more protection against a weaker attacker
  - $\epsilon_1 < \epsilon_\infty$
- Paper proves formulas for the ϵ values in terms of other parameters

Some empirical results

- N responses let you learn at most $\sqrt{N}/10$ most common values
- In a sample distribution, detects mostly the most common elements
- Short case studies of malware binaries on Windows and Chrome user home pages