Accuracy Limits on Private Query Answering

CompSci 590.03
Instructor: Ashwin Machanavajjhala
Outline

• Baseline for Privacy: Blatant Non-Privacy

• Exponential Time Adversaries

• Polynomial Time Adversaries

• Feasibility result
Query Answering

Database

Query

True Answer

Perturbed Answer

Researcher
Model

• Database of bits: $d \in \{0,1\}^n$

• Queries: Subset sums
  
  – Consider $q \subseteq [n]$

  – $a_q = \sum_{i \in q} d_i$

• Perturbed Answer returned by a private algorithm: $A(q)$

  – Error: $\mathcal{E} = \max_q |A(q) - a_q|$
Blatant Non-Privacy

**Definition 3 (Non-Privacy).** A database $D = (d, A)$ is $t(n)$-non-private if for every constant $\varepsilon > 0$ there exists a probabilistic Turing Machine $M$ with time complexity $t(n)$ so that

$$\Pr[M^A(1^n) \text{ outputs } c \text{ s.t. } \text{dist}(c, d) < \varepsilon n] \geq 1 - \text{neg}(n).$$

- $\text{dist}(c,d)$ = Hamming distance
  = number of positions where databases $c$ and $d$ differ.

- $\text{neg}(n)$: $\forall c, \exists n_0, \forall n > n_0 \text{ neg}(n) < 1/n^c$

- Meaning of the definition:
  A database $d$ along with a perturbed access mechanism $A$ is $t(n)$-non-private if an attacker can “decode” the database with high probability using query-(perturbed) answer pairs in $t(n)$ time.
Outline

• Baseline for Privacy: Blatant Non-Privacy

• Exponential Time Adversaries

• Polynomial Time Adversaries

• Feasibility result
**Exponential Time Adversary**

**Theorem 2.** Let $D = (d, A)$ be a database where $A$ is within $o(n)$ perturbation. Then $D$ is $\exp(n)$-non-private.

[QUERY PHASE]
For all $q \subseteq [n]$: let $\tilde{a}_q \leftarrow A(q)$.

[WEEDING PHASE]
For all $c \in \{0, 1\}^n$: if $|\sum_{i \in q} c_i - \tilde{a}_q| \leq \mathcal{E}$ for all $q \subseteq [n]$ then output $c$ and halt.

$\mathcal{E} = o(n)$
Exponential Time Adversary

Attack always terminates (why?)

• Algorithm considers all database in the weeding phase.
• Original database $d$ is never weeded out.
Exponential Time Adversary

\[ \text{dist}(d, c) \leq 4\varepsilon = o(n) \]

Suppose \( \text{dist}(c, d) > 4\varepsilon \).
Let \( q_0 = \{i \mid d_i = 1, c_i = 0\} \), and \( q_1 = \{i \mid d_i = 0, c_i = 1\} \)
\[ |q_0| + |q_1| > 4\varepsilon. \] Thus, wlog \( |q_1| > 2\varepsilon \)

\[ \sum_{i \in q_1} d_i = 0 \implies A(q_1) < \varepsilon \]

But, \( \sum_{i \in q_1} c_i = |q_1| > 2\varepsilon \)

\[ \left| \sum_{i \in q_1} c_i - A(q_1) \right| > \varepsilon \]

Database c would not have passed the weeding phase
Exponential Time Adversary

**Theorem 2.** Let $D = (d, A)$ be a database where $A$ is within $o(n)$ perturbation. Then $D$ is $\exp(n)$-non-private.

[**QUERY PHASE**]
For all $q \subseteq [n]$: let $\tilde{a}_q \leftarrow A(q)$.

[**Weeding Phase**]
For all $c \in \{0, 1\}^n$: if $|\sum_{i \in q} c_i - \tilde{a}_q| \leq \mathcal{E}$ for all $q \subseteq [n]$ then output $c$ and halt.

With an exponential number of queries, an adversary can reconstruct the entire database **even if error in each query is $o(n)$**
Exponential Time Adversary

• What about $\Theta(n)$ error?

• Error = $n/2$
  – Trivial ...
  – Always answer $n/2$
  – No utility

• Error = $n/40$
  – Hint: Using the proof of the theorem ...
  – Can reconstruct $9/10$ of the database!
Summary of Exponential Adversary

• An adversary who can ask all queries can reconstruct a large fraction of the database with probability 1.

• What if the adversary is only allowed to asked a small set of queries?
Outline

• Baseline for Privacy: Blatant Non-Privacy

• Exponential Time Adversaries

• Polynomial Time Adversaries

• Feasibility Result
Polynomial Time Adversaries

**Theorem 3.** Let $\mathcal{D} = (d, A)$ be a database where $A$ is within $o(\sqrt{n})$ perturbation then $\mathcal{D}$ is $\text{poly}(n)$-non-private.

[**QUERY PHASE**]
Let $t = n(\log n)^2$. For $1 \leq j \leq t$ choose uniformly at random $q_j \subseteq_R [n]$, and set $\tilde{a}_{q_j} \leftarrow A(q_j)$.

[**WEEDING PHASE**]
Solve the following linear program with unknowns $c_1, \ldots, c_n$:

\[
\tilde{a}_{q_j} - \mathcal{E} \leq \sum_{i \in q_j} c_i \leq \tilde{a}_{q_j} + \mathcal{E} \quad \text{for } 1 \leq j \leq t
\]

\[
0 \leq c_i \leq 1 \quad \text{for } 1 \leq i \leq n
\]

[**ROUNDING PHASE**]
Let $c'_i = 1$ if $c_i > 1/2$ and $c'_i = 0$ otherwise. Output $c'$. 

Lecture 7 : 590.03 Fall 13
Polynomial Time Adversaries

**Theorem 3.** Let \( D = (d, A) \) be a database where \( A \) is within \( o(\sqrt{n}) \) perturbation then \( D \) is \( \text{poly}(n) \)-non-private.

**[QUERY PHASE]**
Let \( t = n(\log n)^2 \). For \( 1 \leq j \leq t \) choose uniformly at random \( q_j \subseteq_R [n] \), and set \( \tilde{a}_{q_j} \leftarrow A(q_j) \).

**[WEEDING PHASE]**
Solve the following linear program with unknowns \( c_1, \ldots, c_n \):

\[
\tilde{a}_{q_j} - \mathcal{E} \leq \sum_{i \in q_j} c_i \leq \tilde{a}_{q_j} + \mathcal{E} \quad \text{for } 1 \leq j \leq t \\
0 \leq c_i \leq 1 \quad \text{for } 1 \leq i \leq n
\]  

(1)

**[ROUNDING PHASE]**
Let \( c'_i = 1 \) if \( c_i > 1/2 \) and \( c'_i = 0 \) otherwise. Output \( c' \).

With \( n \log^2 n \) queries, an adversary can reconstruct the entire database **even if error in each query is** \( o(\sqrt{n}) \)
Summary of negative results

• Attackers can ask multiple questions to the database to learn sensitive information, even when each query answer is perturbed.

• General result
  – Perturbation need not be independent for each query (no assumption on how noise is infused)
  – Subset sum queries are quite general. Just use a random set of queries ...
  – Both exponential time and polynomial time attacks

• Need to think of privacy as a budget-constrained problem
  – Given a perturbation level, there is an upper bound on the number of queries that can be answered.
  – Once the limit is reached, no more queries can be answered.
Outline

• Baseline for Privacy: Blatant Non-Privacy

• Exponential Time Adversaries

• Polynomial Time Adversaries

• Feasibility Result
Tightness of the $o(\sqrt{n})$ bound

- There exists a mechanism that is not blatant non-private, and which can answer $\text{polylog}(T(n))$ queries with $\sqrt{T(n)}$ noise per query.
Not “Blatant non-private”

- Suppose database is drawn uniformly at random from \{0, 1\}^n.

- Consider 2 Turing machines with time complexity \(T(n)\)
  - \(M^A_1\) outputs pairs of queries and perturbed answers using \(A\), and an index \(i\)
  - \(M_2\) takes index \(i\) and all the other values in \(d (d^{-i})\) and outputs \(d_i\).

- We have \((T(n), \delta)\)-privacy if:

\[
\Pr \left[ \begin{array}{c} M^A_1(1^n) \text{ outputs } (i, \text{view}) ; \\ M_2(\text{view}, d^{-i}) \text{ outputs } d_i \end{array} \right] < \frac{1}{2} + \delta
\]

- ... a precursor to differential privacy (next class)
Feasibility Result

**Theorem 5.** Let $T(n) > \text{polylog}(n)$, and let $\delta > 0$. Let $DB$ be the uniform distribution over $\{0,1\}^n$, and $d \in_R DB$. There exists a $\tilde{O}(\sqrt{T(n)})$-perturbation algorithm $A$ such that $D = (d,A)$ is $(T(n),\delta)$-private.

1. Let $a_q = \sum_{i \in q} d_i$.

2. Generate a perturbation value: Let $(e_1, \ldots, e_R) \in_R \{0,1\}^R$ and $\mathcal{E} \leftarrow \sum_{i=1}^R e_i - R/2$.

3. Return $a_q + \mathcal{E}$. 
Proof Highlights

• A is a polylog($\sqrt{T(n)}$)-perturbation mechanism

Chernoff Bounds: $X_1, \ldots, X_n$ independent random vars 
$X_i \in [0,1], E(X_i) = p$, then

$$\Pr[X_1 + \cdots + X_n > np + x] < e^{-\frac{x^2}{2np(1-p)}}$$

$$\Pr[|\xi| > \log^2 n\sqrt{R}] < 2e^{-\frac{\log^4 n \cdot R}{R/2}} < \text{neg}(n)$$
Proof Highlights

To Show:
Probability that $d_i = 1$ given query answer pairs, and all the bits other than $d_i$ is bounded

\[
p_\ell = \Pr[d_i = 1|a_1, \ldots, a_\ell] < \frac{1}{2} + \delta
\]

\[
p_\ell = p_{\ell-1} \cdot \frac{\Pr[a_\ell|d_i = 1] \cdot \Pr[a_1, \ldots, a_{\ell-1}]}{\Pr[a_1, \ldots, a_\ell]}
\]

\[
1 - p_\ell = (1 - p_{\ell-1}) \cdot \frac{\Pr[a_\ell|d_i = 0] \cdot \Pr[a_1, \ldots, a_{\ell-1}]}{\Pr[a_1, \ldots, a_\ell]}
\]
Proof Highlights

• Adversary’s confidence in $d_i = 1$ after $L$ queries ...

$$\text{conf}_\ell \overset{\text{def}}{=} \log \left( \frac{p_\ell}{1 - p_\ell} \right)$$

• Adversary’s confidence starts at 0, and $\text{conf}_\ell = \text{conf}_{\ell - 1}$, when $i \notin q_i$

• For privacy, we want to show that

$$|\text{conf}_\ell| < \delta' = \log \left( \frac{\frac{1}{2} + \delta}{1/2 - \delta} \right) \text{ for all } 0 < \ell \leq t$$
Proof Highlights

• Confidence depends on all the prior queries. Maybe hard to compute.

\[ \text{step}_\ell \overset{\text{def}}{=} \text{conf}_\ell - \text{conf}_{\ell - 1} = \log \left( \frac{\Pr[a_\ell | d_i = 1]}{\Pr[a_\ell | d_i = 0]} \right) \]

• The sequence \(0 = \text{conf}_1, \text{conf}_2, \ldots, \text{conf}_t\) defines a random walk on a line, defined by random variable \(\text{step}_i\).

• We are done if we show that the random walk needs more than \(t\) steps to reach \(\delta’\) ...
Proof Highlights

• Consider two cases when $d_i = 1$ and $d_i = 0$. To get answer $a_l$ in both cases requires different noises $k$ and $k+1$.

$$\text{step}_l = \frac{\Pr[a_l|d_i = 1]}{\Pr[a_l|d_i = 0]} = \frac{\Pr[\mathcal{E} = k]}{\Pr[\mathcal{E} = k + 1]}$$

$$\Pr\left[\text{step}_l = \log \frac{k + 1}{R - k}\right] = \binom{R}{k}/2^k$$

• We can show expectation and absolute value of each step is small.

$$E\left[\sum_l \text{step}_l\right] \leq O\left(\frac{1}{\log^\mu n}\right)$$

$$|\text{step}_l| \leq O\left(\log^2 n / \sqrt{R}\right)$$
Proof Highlights

• Proof can be completed using the Hoeffdings inequality

If $X_1, X_2, \ldots, X_n$ are independent random variables
$s. t. \Pr[|X_i| \leq a] = 1.$

Let $S = X_1 + X_2 + \cdots + X_n$

$$\Pr[S - E(S) > t] < e^{-\frac{t^2}{2na^2}}$$

• The step random variables satisfy all these conditions.
Summary

• Showing feasibility requires defining privacy.

• Privacy defined in terms of adversary’s posterior knowledge

• Algorithm uses additive randomization and maintains no state about previous queries
  – No need for query auditing
  – However there is a bound on the number of queries allowable.

• Precursor to differential privacy
Next class

• Differential Privacy

References:
• Dinur, Nissim, “Revealing information while preserving privacy”, PODS 2003