T-79.514 Special Course on Cryptology

Revealing Information while Preserving Privacy

Emilia Oikarinen
Helsinki University of Technology
emilia.oikarinen@hut.fi

October 29, 2003
Background

- Consider a hospital database consisting of medical history of a population.
  - The privacy of individual patients should be maintained.
  - Could the database be used to obtain some statistical information?
  - Why the removing of all identifying attributes from the database does not help?

Overview of the Lecture

- Model-Statistical Databases and Statistical Queries
- Database Privacy in Terms of Non-Privacy
- Impossibility Results – Exponential/Polynomial Adversary
- Privacy and Feasibility Results
- Conclusions
Notations

- **neg**\((n)\) — a function that is asymptotically smaller than any inverse polynomial, i.e. for all \(c > 0\) and for all sufficiently large \(n\), it holds that \(\text{neg}(n) < 1/n^c\).

- **dist**\((c, d)\) — the *Hamming distance* of two binary strings \(c, d \in \{0, 1\}^n\), i.e. \(\text{dist}(c, d) = |\{i \mid c_i \neq d_i\}|\).

- \(\tilde{O}(T(n)) = O(T(n) \log^k(n))\), for some \(k > 0\).

- \(\mathcal{M}\) is a Turing-machine. \(\mathcal{M}^A\) is an \(\mathcal{A}\)-oracle Turing-machine, where \(\mathcal{M}\) has an access to algorithm \(\mathcal{A}\) and each call to \(\mathcal{A}\) costs a unit time.
Model-Statistical Databases and Statistical Queries

- Let \( d = (d_1, \ldots, d_n) \in \{0, 1\}^n \). A (statistical) query is a subset \( q \subseteq \{1, \ldots, n\} \). The (exact) answer to a query \( q \) is the sum of all database entries in \( q \), i.e. \( a_q = \sum_{i \in q} d_i \).

- A (statistical) database \( D = (d, A) \) is a query-response mechanism. The response to a query \( q \) is \( A(q, d, \theta) \), where \( \theta \) is the internal state of the algorithm \( A \).

- We usually omit \( d \) and \( \theta \) and write \( A(q) \) for \( A(q, d, \theta) \).
Privacy Methods for Statistical Databases

(i) query restriction
(ii) data perturbation
(iii) output perturbation

The quality of a database algorithm $\mathcal{A}$ in terms of the magnitude of its perturbation:

- An answer $\mathcal{A}(q)$ is within $\mathcal{E}$ perturbation if $a_q - \mathcal{A}(q) \leq \mathcal{E}$.
- An algorithm $\mathcal{A}$ is within $\mathcal{E}$ perturbation if for all queries $q \subseteq \{1, \ldots, n\}$ the answer $\mathcal{A}(q)$ is within $\mathcal{E}$ perturbation.
Database Privacy

- Problem of finding a balance between private functions and information functions.

- A computational definition of privacy: it is computationally infeasible to retrieve private information from the database.

- Other measures of privacy used in previous works include e.g. variance of query answers and the estimator variance.

- Reversed order compared to cryptography.

- Before we define privacy, we consider the concept of non-privacy.
Non-Privacy

- A database $\mathcal{D} = (d, \mathcal{A})$ is $t(n)$-non-private, if for every constant $\varepsilon > 0$ there exists a probabilistic Turing-machine $\mathcal{M}$ with time-complexity $t(n)$ such that

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

where the probability is taken over coin tosses of $\mathcal{A}$ and $\mathcal{M}$.

- From now on, we will restrict the adversary by making the queries non-adaptive.
Impossibility Results – Exponential Adversary

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

- **Adversary’s algorithm.**
  Let $\mathcal{A}$ be within $\mathcal{E} = o(n)$ perturbation. Let $\mathcal{M}$ be the following.

(i) (Query phase)
   For all $q \subseteq \{1, \ldots, n\}$, let $\tilde{a}_q = \mathcal{A}(q)$.

(ii) (Weeding phase)
   For all $c \in \{0, 1\}^n$, if $|\sum_{i \in q} c_i - \tilde{a}_q| \leq \mathcal{E}$ for all $q \subseteq \{1, \ldots, n\}$,
   then output $c$ and halt.
Impossibility Results – Polynomial Adversary

Let us consider a more realistic scenario in which the adversary is polynomially bounded.

- **Theorem.** Let $\mathcal{D} = (d, \mathcal{A})$ be a database where $\mathcal{A}$ is within $o(\sqrt{n})$ perturbation. Then $\mathcal{D}$ is $\text{poly}(n)$-non-private.
• **Adversary’s algorithm.** (A within \( E = o(\sqrt{n}) \) perturbation):

  (i) (Query phase)
  Let \( t = n \log^2(n) \). For \( 1 \leq j \leq t \), choose uniformly at random \( q_j \subseteq \{1, \ldots, n\} \), and set \( \tilde{a}_{q_j} = A(q_j) \).

  (ii) (Weeding phase)
  Solve the following linear program (LP) with \( n \) unknowns \( c_1, \ldots, c_n \).

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

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

  (iii) (Rounding phase)
  Let \( c'_i = 1 \) if \( c_i > 1/2 \) and \( c'_i = 0 \) otherwise. Output \( c' \).
Tightness of the Impossibility Results

- A database algorithm that is within $\tilde{O}(\sqrt{n})$ perturbation and private against polynomial adversaries:

Let $d \in \{0, 1\}^n$ at random and set the perturbation magnitude $E = \sqrt{n}(\log n)^{1+\varepsilon} = \tilde{O}(\sqrt{n})$. Consider database $D = (d, A)$ with algorithm $A$ defined as follows,

(i) For an input query $q \subseteq \{1, \ldots, n\}$, compute $a_q = \sum_{i \in q} d_i$.

(ii) If $|a_q - \frac{|q|}{2}| < E$, return $\frac{|q|}{2}$.

(iii) Otherwise, return $a_q$.

- The above database is effectively useless.
We present now, a database algorithm that has some privacy combined with some usability.

We relax the requirements in definition of non-privacy and require that $A(q)$ is within $\mathcal{E}$ perturbation for most $q$, i.e.

$$\Pr_{q \in \{1, \ldots, n\}} [A(q) \text{ is within } \mathcal{E} \text{ perturbation}] = 1 - \text{neg}(n).$$

Let $\mathcal{DB}$ be the uniform distribution over $\{0, 1\}^n$ and select $d \in \mathcal{DB}$ at random.
Tightness of the Impossibility Results Cont’d

- The database algorithm \( \mathcal{A} \) will use an internal state \( \theta \) that is initialized upon the first call.

- \( \theta \) consists of \( n \) bits \( d' = (d'_1, \ldots, d'_n) \) where \( d'_i = d_i \) with probability \( 1/2 + \delta \) and \( d'_i = 1 - d_i \) otherwise. Thus \( \theta \) is a private version of the database.

- On an input query \( q \subseteq \{1, \ldots, n\} \) algorithm \( \mathcal{A} \) answers \( \tilde{a}_q = \sum_{i \in q} d'_i \).

- \( \mathcal{A} \) is within \( \tilde{O}(\sqrt{n}) \) perturbation and the database has some usability (Note that, the algorithm is essentially RRT).
Definition of Privacy

Let $\mathcal{DB}$ be a distribution over $\{0, 1\}^n$ and $d$ is drawn according to $\mathcal{DB}$. A database $\mathcal{D} = (d, A)$ is $(T(n), \delta)$-private, if for every pair of probabilistic Turing machines $\mathcal{M}_1$ and $\mathcal{M}_2$ having time-complexity $T(n)$, it holds that

$$\Pr \left[ \mathcal{M}_1(1^n) \text{ outputs } (i, \text{view}); \mathcal{M}_2(\text{view}, d^{-i}) \text{ outputs } d_i \right] < \frac{1}{2} + \delta,$$

where $d^{-i} = (d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_n)$. The probability is taken over the choice of $d$ from $\mathcal{DB}$ and the coin tosses of all machines involved.
Feasibility Results

- Assume that the adversary has no prior information about the database (modeled by drawing the database from the uniform distribution over $n$-bit strings)

- **Theorem.** Let $T(n) > \log^k(n)$ and $\delta > 0$. Let $DB$ be uniform distribution over $\{0, 1\}^n$, and select $d \in DB$ at random. There exists a $\tilde{O}(\sqrt{T(n)})$-perturbation algorithm $A$ such that $D = (d, A)$ is $(T(n), \delta)$-private.
Conclusions

- If some random noise of magnitude $\leq \mathcal{E}$ is added to a database to preserve privacy, there is a threshold phenomenon where a polynomially bounded adversary can reconstruct almost all the database entries if $\mathcal{E} \ll \sqrt{n}$, and if $\mathcal{E} \gg \sqrt{n}$ the adversary can reconstruct none of them.

- Privacy can be preserved with respect to an adversary having running time limited to $\mathcal{T}(n)$ for an arbitrary $\mathcal{T}$ when a perturbation magnitude of about $\sqrt{\mathcal{T}(n)}$ is used.