Privacy Enhancing Technologies FS2026
Exercise Sheet 7 (version 1)

Florian Tramèr

Problem 1: Conceptual Questions. For each of the following statements, say whether it is TRUE or FALSE. Write at most one sentence to justify your answer.

(a) For a database $D \in \{0,1\}^n$, the curator replies to each counting query S with $\texttt{Ans}(S) + E$, where E is ± 2 with probability $1/2$. An adversary can reconstruct the entire database with n subset queries.

(b) You and your friends want to determine which one of you has the lowest salary. You design and run a protocol, at the end of which all your friends learn that their "Big 4" salaries are higher than yours. This blatant invasion of your privacy could have been avoided if you had used a proper maliciously-secure MPC protocol for the function $f(x_1, \ldots, x_n) = \arg\min_i x_i$.

(c) Differential privacy and k-anonymity are both properties of an algorithm applied to a database.

(d) A mechanism that responds to any query with the value x_1 of the first individual in the database is blatantly non-private.

Problem 2: Data de-anonymization from auxiliary information. A credit-card company has a database that contains the transaction history of its customers. The company wants to release this data to the public to enable data-driven research. To prevent privacy leaks, they "anonymize" the data by removing any identifiers (names, addresses, etc.) and release a database $D \in \mathcal{X}^{n \times m}$ where each row contains m transactions made by one individual.

The adversary wants to de-anonymize one individual in the database, and learn something about their spending habits. Assume the adversary gets access to auxiliary information $\texttt{aux}(r) \in \mathcal{X}^k$ for one row $r \leftarrow D$ (chosen uniformly at random). This auxiliary information consists of information about k transactions selected uniformly at random from the row r. Namely, we have that $r_i = \texttt{aux}_i$ for $i \in \text{supp}(\texttt{aux})$.

(a) Let $r, r' \in \mathcal{X}^m$ be rows that differ in an α-fraction of the indices. Give a tight upper bound $f(\alpha, k)$ on the probability that r' is consistent with $\texttt{aux}(r)$, i.e.,

$$\Pr[r'_i = \texttt{aux}_i \text{ for all } i \in \text{supp}(\texttt{aux})] \leq f(\alpha, k) \ ,$$

where the probability is taken over the random choice of $\texttt{aux}$.

(b) Let $k > \frac{\log(\alpha/n)}{\log(1-\alpha)}$. Provide an algorithm that, given $\texttt{aux}(r)$ for a random $r \in D$, outputs a guess $\hat{r} \in \mathcal{X}^m$ such that $\hat{r}$ and r agree in at least a $(1-\alpha)$-fraction of the indices, with probability at least $1 - \alpha$, where the probability is taken over the random choice of r and $\texttt{aux}$, and the randomness (if any) of your algorithm.

(c) We say that a database D is (α, β)-*sparse* if at most for a β-fraction of the rows r of D, there exists another row $r' \in D$ that agrees with r in at least a $(1-\alpha)$-fraction of the indices.

Let D be (α, β)-sparse, and let r be a uniformly random row in D. Let $k > \frac{\log(\alpha/n)}{\log(1-\alpha)}$ as in part (b). Provide an algorithm that, given $\texttt{aux}(r)$ (with a uniformly random index set I), outputs a guess $\hat{r} \in \mathcal{X}^m$ such that $\hat{r} = r$, with probability at least $1 - \alpha - \beta$. This probability is taken over the random choice of r and I, and the randomness (if any) of your algorithm.

Problem 3: Solving noisy linear systems. Suppose we are given a matrix $A \in \mathbb{R}^{m \times n}$ (with $m \geq n$) and vector $y \in \mathbb{R}^m$ such that $y = As + e$, where $s \in \mathbb{R}^n$ is an unknown vector, and $e \in \mathbb{R}^m$ is a noise vector.

A natural way to estimate s is through least-squares: $\tilde{s} = \arg\min_s \|As - y\|_2$.

(a) Show how to cast the attacks of Dinur and Nissim in this setting. What is s? What is the matrix A? What properties does the error vector e have?

(b) To solve a noisy linear system, we can use the following algorithm:

- Compute the Singular Value Decomposition (SVD) of $A = P\Sigma Q^\top$. Here, P is an orthogonal $m \times m$ matrix, Σ is a diagonal $m \times n$ matrix of singular values, and Q is an orthogonal $n \times n$ matrix.
- Compute the *pseudo-inverse* $A^\dagger = Q\Sigma^\dagger P^\top$, where $\Sigma^\dagger$ is the diagonal matrix with the reciprocal of the non-zero singular values of Σ, i.e., $\Sigma^\dagger_{i,i} = 1/\Sigma_{i,i}$ for $\Sigma_{i,i} \neq 0$ and 0 otherwise.
- Let $\hat{s} = A^\dagger y$.

Assuming that $A^\dagger A = I_n$, show that $\tilde{s} = s + A^\dagger e$.

(c) Show that $\|\tilde{s} - s\|_2 \leq \frac{\|e\|_2}{\sigma_n(A)}$, where $\sigma_n(A)$ is the smallest singular value of A.

(d) Assume that $m = n$ and that if A is a random binary matrix (with 1/2 probability for each entry), then $\sigma_n(A) = \Theta(1/\sqrt{n})$ with high probability.

Use a least-squares approach to derive Theorem 2 from Dinur and Nissim.

Problem 4: Sensitivities. (This problem is due to Adam Smith and Jonathan Ullman).

For all of the following cases, assume we have a dataset $D = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and a function $f : \mathcal{X}^n \to \mathbb{R}^d$ (for some d that varies between the different parts of the question). For each of the following functions f and data domains $\mathcal{X}$:

- Give as tight a bound as you can on the ℓ_1-sensitivity of the function f. If the sensitivity is not bounded, answer ∞.
- Give as tight a bound as you can on the diameter of the range of f : specifically, what is the maximum difference (in ℓ_1 norm) between any two points in the range of f?

(a) The high-dimensional mean $f(D) = \frac{1}{n}\sum_{i=1}^n x_i$, when $\mathcal{X} = \{v \in \mathbb{R}^d : \|v\|_1 \leq 1\}$.

(b) The median $f(D) = \text{median}(x_1, \ldots, x_n)$, when $\mathcal{X} = [0, 1]$. You can assume that $n = 2k + 1$ is odd and so the median is the middle element of the sorted list.

(c) The Swiss government uses a specific (deterministic, non-constant) algorithm M to decide, based on state populations, how many of the seats in the parliament will go to each state. Think of the collected dataset as recording each person's canton of residence, so $\mathcal{X} = \{1, \ldots, 26\}$. Let $f(D)$ be the function that reports the number of people whose

residence would need to change in order to change the assignment of seats to cantons $M(D)$. In other words, $f(D)$ is the Hamming distance to the nearest data set D' with $M(D') \neq M(D)$.

(d) Suppose we have a fixed set of vertices V (independent of the data set). Our data set is a list of edges: each x_i is a pair of vertices (u, v) (so that $\mathcal{X} = V \times V$). Let $f(D)$ be the length of the shortest path between two specific vertices (let's say vertices number 1 and 2, assuming they are numbered 1 through $|V|$). (If no path exists, the length of the shortest path is ∞).