MPRI 2.3.2, Foundations of Privacy
Final exam

The exam consists of several questions. For each of them, the percentage between parentheses indicates the percentage by which a correct answer contributes to the maximum score (20). The sum of all these percentages is intentionally more than 100%. This means that you do not have to answer all the questions to obtain the maximum score.

**Question 1 (15%)**

Only one of the following statements is true. Please say which one, and motivate your answer.

1. If a mechanism satisfies $\epsilon$-differential privacy, then the side knowledge (prior) does not help the adversary to discover the private information of the members of the database.

2. If a mechanism satisfies $\epsilon$-differential privacy, then the prior does not influence the answer of the query.

3. If a mechanism satisfies $\epsilon$-differential privacy for a certain prior, then it does satisfy $\epsilon$-differential privacy for any prior.

**Question 2 (15%)**

Only one of the following statements is true. Please say which one, and motivate your answer.

1. If two mechanisms satisfy $\epsilon$-differential privacy, then their composition satisfy $\epsilon$-differential privacy.

2. If two mechanisms satisfy $\epsilon$-differential privacy, then their composition satisfy $3\epsilon$-differential privacy.

3. If two mechanisms satisfy $\epsilon_1$-differential privacy and $\epsilon_2$-differential privacy respectively, then their composition satisfy $\epsilon_3$-differential privacy, where $\epsilon_3 = \max\{\epsilon_1, \epsilon_2\}$.

**Question 3 (30%)**

Consider the query $f(x) = \text{average height of the people in the database } x$

Assume that the database contains at least 100 people, and that the height of a person ranges between 50 and 200 centimeters. Consider a mechanism obtained by adding Laplacian noise to the answer of the query, according to the following distribution (where $y = f(x)$ is the answer of the query, $z$ is the reported answer, and $c$ is a normalization factor):

$$d_y(z) = ce^{-|z-y|}$$

Does the mechanism satisfy $\epsilon$-differential privacy, for some $\epsilon$? If the answer is yes, please give the minimum such $\epsilon$ (under the above assumptions on the dimension of the database and the range of the height). If the answer is no, please find a counterexample.

**Question 4 (40%)**

Compute the utility of the mechanism of Question 3, assuming that the prior distribution on the result $y$ of the query is uniform, and that the gain function is the binary one (i.e., $g(w, y) = 1$ if $w = y$, and 0 otherwise).

**Question 5 (30%)**

Let $C$ be a channel from $\mathcal{X}$ to $\mathcal{Y}$.

5.1 Show that for any prior $\pi$ and gain function $g$:

$$L^x_g(\pi, C) \leq |\mathcal{Y}| \quad \text{and} \quad L^y_g(\pi, C) \leq |\mathcal{X}|$$
5.2 Let $\pi_u$ be the uniform prior. Show that

$$(\forall g : L^\pi_g (\pi_u, C) = 1)$$

if and only if $C$ is non-interfering.

**Question 6 (40%)**

6.1 Consider an instance of the Dining Cryptographers protocol with 3 cryptographers on a line:

Crypt$_1$ —– Crypt$_2$ —– Crypt$_3$

That is, there is a coin between Crypt$_1$/Crypt$_2$ and Crypt$_2$/Crypt$_3$, but not between Crypt$_1$/Crypt$_3$.

Model the system as a channel. Is it non-interfering? Compute its multiplicative Bayes-capacity.

6.2 Now consider the usual instance of 3 Dining Cryptographers on a ring, but assume that the coin shared between Crypt$_1$/Crypt$_3$ is observable (i.e. visible to the adversary).

Repeat the question 2.1 for this variant.

6.3 Show that the channel of question 2.2 can be obtained by post-processing the channel of question 2.1.

How can we compute the multiplicative Bayes-capacity in question 2.2 using this fact?

6.4 Consider again 3 Dining Cryptographers on a ring, and assume that this time the coin shared between Crypt$_1$/Crypt$_3$ is hidden but biased (gives heads with probability other than one half). All the other coins are still fair.

What is the multiplicative Bayes-capacity of this system?

**Question 7 (30%)**

In the Crowds protocol, due to the probabilistic routing, each request could pass through corrupted users multiple times before arriving to the server, as shown in the figure below. However, in the security analysis, we only considered as “detected” the first user who forwards the request to a corrupted one.

To perform a more precise analysis, let us consider the first two detected users, instead of only the first one. Let $n, m$ be the number of honest and total users respectively. The set of secrets is still $X = \{1, \ldots, n\}$ (we are only interested in the privacy of honest users).

On the other hand, the information available to the adversary is now more detailed. Observations are of the form $y = (d_1, d_2)$ where $d_1 \in \{1, \ldots, n, \bot\}$ (the first detected user, similarly to the original analysis) and $d_2 \in \{1, \ldots, m, \bot\}$ (the second detected user, who might be corrupted himself).

Show that this extra information is in fact useless to the adversary. More precisely, show that for any prior $\pi$ and gain function $g$:

$$V_g (\pi, C^1) = V_g (\pi, C^2)$$

where $C^2$ is the channel obtained by the detailed analysis, considering two detected users, and $C^1$ is the channel of the original analysis, considering a single detected user.