Quantitative Information Flow (QIF) **DCC/ICEx/UFMG**
Prof. Mário S. Alvim **2021/1**

PROBLEM SET

Posterior vulnerability and leakage
(Chapter 05)

Necessary reading for this assignment:

- *The Science of Quantitative Information Flow* (Alvim, Chatzikokolakis, McIver, Morgan, Palamidessi, and Smith):
  - Chapter 5: *Posterior vulnerability and leakage*
    - Chapter 5.1: *Posterior g-vulnerability and its basic properties*
    - Chapter 5.2: *Multiplicative and additive g-leakage*
    - Chapter 5.3: *A closer look at posterior Bayes vulnerability and Bayes leakage*
    - Chapter 5.4: *Measuring leakage with Shannon entropy*
    - Chapter 5.5: *More properties of posterior g-vulnerability and g-leakage*
    - Chapter 5.6: *Example channels and their leakage*
    - Chapter 5.7: *Max-case posterior g-vulnerability*

Review questions.

1. State the formulation of *posterior g-vulnerability* given by Definition 5.2, by Theorem 5.6, and by Theorem 5.7. Explain the purpose of each of these different (even if all equivalent) formulations.

2. State the property of *monotonicity* relating prior- and posterior g-vulnerability. Explain informally what it means in terms of the information gained by an adversary by observing the behavior of a channel processing a secret value.

3. State the definition of *additive and multiplicative g-leakage*.

Exercises.

4. (Exercise 5.1) Prove Theorem 5.7.

5. (Exercise 5.3) It is noted before Definition 4.14 that *noninterference* is a traditional name for the "no leakage" property. Show that noninterference indeed implies "no g-leakage" — that is, show that if C satisfies noninterference, then for any prior π and gain function g, we have $\mathcal{L}_g^{\times}(\pi, C) = 1$ and $\mathcal{L}_g^{+}(\pi, C) = 0$.

6. (Exercise 5.4) The *Monty Hall problem* is a famous brain teaser, based loosely on the old game show *Let's Make a Deal*, whose host was Monty Hall. In 1990, the problem appeared in the "Ask Marilyn" column of *Parade* magazine, formulated as follows:

   > Suppose you're on a game show, and you're given the choice of three doors: behind one door is a car; behind the other two are goats. You pick a door, say Door 1, and the host, who knows what's behind the doors, opens another door, say Door 3, which has a goat. He then says to you, "Do you want to pick Door 2 instead?"
   >
   > Is it to your advantage to switch your choice?

This formulation is actually a bit imprecise about Monty Hall's behavior. What is intended is that Monty *always* opens a door that you did not choose and that contains a goat. (Since there are two doors containing goats, it is always possible for him to do that.) Also, he *always* gives you the option of switching. [1]

To solve this puzzle, let us formulate Monty Hall as a probabilistic channel M (for "Monty") whose secret input X is the door containing the car, and whose output Y is the door that Monty opens after your initial choice, which we assume is Door 1. In the case when the car is behind Door 1, note that *both* Doors 2 and 3 contain goats, giving Monty a choice of which door to open. Here we assume that he makes this choice by flipping a fair coin, opening Door 2 if he gets *heads* and Door 3 if he gets *tails*. Hence the channel matrix is as follows:

M	2	3
1	$1/2$	$1/2$
2	0	1
3	1	0

.

Also, we assume a uniform prior $\boldsymbol{\vartheta} = (1/3, 1/3, 1/3)$, so that the car is equally likely to be behind each of the doors.

(a) Calculate the hyper-distribution $[\boldsymbol{\vartheta} \triangleright \mathsf{M}]$.

(b) Calculate the posterior Bayes vulnerability $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{M}]$ and the multiplicative Bayes leakage $\mathcal{L}_1^{\times}(\boldsymbol{\vartheta}, \mathsf{M})$. Based on your calculations, should you stick with Door 1 or should you switch?

(c) Now assume that when the car is behind Door 1 (more generally, behind the door you choose), Monty uses a *biased* coin that gives *heads* with probability p and *tails* with probability $1-p$, for some p such that $0 \leq p \leq 1$. This changes the channel matrix to the following:

M	2	3
1	p	$1-p$
2	0	1
3	1	0

.

How does that change the results?

7. (Exercise 5.7) Recall Exercise 4.2, which considers a password checker implementation with a timing side channel. Here we continue that topic in the more realistic setting of a 4-digit PIN ranging from 0000 to 9999. As before, an ideal password checker is modeled by a family of channels C^{g} with output set $\{\text{reject}, \text{accept}\}$. And a flawed implementation that compares the guess with the correct PIN digit by digit, rejecting as soon as a mismatch is found, is modeled by a family of channels D^{g} with output set $\{(\text{reject}, 1), (\text{reject}, 2), (\text{reject}, 3), (\text{reject}, 4), \text{accept}\}$.

(a) Assuming a uniform prior $\boldsymbol{\vartheta} = (1/10{,}000, 1/10{,}000, \ldots, 1/10{,}000)$, show that, for any guess g, the posterior Bayes vulnerabilities under C^{g} and D^{g} are $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{C}^{\mathsf{g}}] = 1/5000$ and $V_1[\boldsymbol{\vartheta} \triangleright \mathsf{D}^{\mathsf{g}}] = 1/2000$.

(b) Given that, one might be tempted to conclude that the difference in the posterior Bayes vulnerabilities is small enough that the side channel in D^{g} is not worth bothering about.

But consider that in a typical PIN scenario, the adversary can enter a *sequence* of guesses $\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_k$, thus running channels $\mathsf{D}^{\mathsf{g}_1}, \mathsf{D}^{\mathsf{g}_2}, \ldots, \mathsf{D}^{\mathsf{g}_k}$ for some moderate value of k. (Too many

[1] A notable advantage of setting Quantitative Information Flow on a rigorous foundation is that it allows tricky problems (such as this one, Monty Hall) to be solved more or less *mechanically* — that is, without the need for deep thought. It calls to mind a wonderful quote from Alfred Whitehead:

> It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle — they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.

consecutive incorrect guesses might get her locked out.) Note that she can choose her guesses *adaptively*, which means that she can choose each guess g_i based on the outputs from the previous guesses $\mathsf{g}_1, \mathsf{g}_2, \ldots, \mathsf{g}_{i-1}$. How many adaptively chosen guesses to D^{g} does she need to determine the correct PIN in the worst case? In the average case?

(c) In contrast, how many adaptively chosen guesses to the ideal password checker C^{g} does she need in the worst case? In the average case?