Quantitative notions of leakage for one-try attacks

Christelle Braun[a] Konstantinos Chatzikokolakis[b]
Catuscia Palamidessi[a]

[a] *INRIA and LIX, École Polytechnique*
Palaiseau, France
{braun,catuscia}@lix.polytechnique.fr

[b] *Technical University of Eindhoven*
Eindhoven, The Netherlands
{kostas}@chatzi.org

Abstract

Recent research in quantitative theories for information-hiding topics, such as Anonymity and Secure Information Flow, tend to converge towards the idea of modeling the system as a noisy channel in the information-theoretic sense. The notion of information leakage, or vulnerability of the system, has been related in some approaches to the concept of mutual information of the channel. A recent work of Smith has shown, however, that if the attack consists in one single try, then the mutual information and other concepts based on Shannon entropy are not suitable, and he has proposed to use Rényi's min-entropy instead. In this paper, we consider and compare two different possibilities of defining the leakage, based on the Bayes risk, a concept related to Rényi min-entropy.

Keywords: Information-hiding, hypothesis testing, probability of error, Rényi min-entropy.

1 Introduction

Information-hiding refers to a large class of problems including Secure Information Flow and Anonymity. There has been a growing interest in developing *quantitative* theories for this class of problems, because it has been recognized that non quantitative (i.e. possibilistic) approaches are in general too coarse, in the sense that they tend to consider as equivalent systems that have very different degrees of protection.

Concepts from Information Theory have revealed quite convenient in this domain. In particular, the notion of noisy channel has been used to model protocols for information-hiding, and the flow of information in programs. The idea is that the input of the channel represents the information to be kept secret, and the output represents the observable. The noise of the channel is generated by the efforts of the protocol to hide the link between the secrets and the observable, often achieved by using randomized mechanisms.

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

Correspondingly, there have been various attempts to define the degree of leakage by using concepts based on Shannon entropy, notably the mutual information [14,4,7,8] and the related notion of capacity [10,9,2].

In a recent work, however, Smith has shown that the concept of mutual information is not very suitable for modeling the information leakage in the situation in which the adversary attempts to guess the value of the secret in one single try [12]. He shows an example of two programs in which the mutual information is about the same, but the probability of making the right guess, after having observed the output, is much higher in one program than in the other. In a subsequent paper [13], Smith proposes to use a notion based on Rényi *min-entropy*.

We look at the problem from the point of view of the *probability of error*: the probability that an adversary makes the wrong guess. We propose to formalize the notion of leakage as the "difference" between the probability of error *a priori* (before observing the output) and *a posteriori* (using the output to infer the input via the so-called MAP rule). We argue that there are at least two natural ways of defining this difference: one, that we call *multiplicative*, corresponds to Smith's proposal. The other, which we call *additive*, is new. In both cases, we show that it is relatively easy to find the suprema, which is nice in that it allows us to consider the worst case of leakage. The worst case is also interesting because it abstracts from the input distribution, which is usually unknown, or (in the case of anonymity) may depend on the set of users.

2 Preliminaries

2.1 Noisy channels and Hypothesis Testing

In this section we briefly review some basic notions about noisy channels and hypothesis testing that will be used throughout the paper. We refer to [5] for more details.

A *channel* is a tuple $\langle X, Y, p(\cdot|\cdot)\rangle$ where X, Y are random variables representing, respectively, the input with possible values $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ (the *secrets* or *hypotheses*) and the output with possible values $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$ (the *observables*). The distribution on X, $\vec{\pi} = (\pi_1, \ldots, \pi_n)$ is called *a priori* input distribution. We will also use the notation $p(x_i)$ and $p(y_j)$ to indicate the probabilities of the input x_i (i.e. $p(x_i) = \pi_i$) and the output y_j, respectively. We will denote by $p(y_j|x_i)$ the conditional probability of observing the output y_j when the input is x_i. These conditional probabilities constitute what is called the *channel matrix*, where $p(y_j|x_i)$ is the element at the intersection of the i-th row and j-th column.

The *a posteriori* probability $p(x_i|y_j)$ is the probability that the input is x_i, given that we observe the output y_j. The a priori and the a posteriori probabilities of x_i are related by Bayes theorem:

$$p(x_i|y_j) = \frac{p(y_j|x_i)\, p(x_i)}{p(y_j)}$$

In hypothesis testing we try to infer the *true* hypothesis (i.e. the value that was really given in input) from the observed output value. In general, it is not

possible to determine the right hypothesis with certainty. We are interested, then, in minimizing the *probability of error*, i.e. the probability of making the wrong guess.

We assume that the process of guessing the hypothesis is represented by a *decision function* $f : \mathcal{Y} \to \mathcal{X}$, i.e. the function which gives, for every output y_j, the guessed input x_i.

The (average) probability of error associated to f is given by the sum of the probabilities of guessing a wrong hypothesis for each given output, averaged by the probabilities of the outputs. Since the probability of making the wrong guess, when the output is y_j, is given by $1 - p(f(y_j)|y_j)$, the average probability of error is:

$$\begin{aligned} PE_f &= \textstyle\sum_j p(y_j)(1 - p(f(y_j)|y_j)) \\ &= 1 - \textstyle\sum_j p(y_j)p(f(y_j)|y_j) \end{aligned}$$

It is easy to see that a decision function f_B minimizes the probability of error if and only if it satisfies the MAP (*Maximum A Posteriori probability*) criterion, namely, for each output y_j it chooses an input x_i for which $p(x_i|y_j)$ is maximum. Formally:

$$f_B(y_j) = x_i \;\Rightarrow\; \forall k \;\; p(x_i|y_j) \geq p(x_k|y_j)$$

It is easy to see that the probability of error associated to f_B is then given by

$$PE_B = 1 - \textstyle\sum_j p(y_j) \max_i p(x_i|y_j)$$

By using Bayes theorem, we can rewrite PE_B as:

$$PE_B = 1 - \textstyle\sum_j \max_i (p(y_j|x_i)\, \pi_i)$$

PE_B is also called *the Bayes risk*. Note that it is a function of $\vec{\pi}$, so we will also write $PE_B(\vec{\pi})$ when we need to emphasize its dependency on $\vec{\pi}$.

2.2 Rényi entropies, Shannon entropy, and mutual information

Rényi entropies [11] are a family of functions representing the uncertainty associated to a random variable. The Rényi entropy of order α, with $\alpha \geq 0$ and $\alpha \neq 1$, is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_i p(x_i)^\alpha \right)$$

In the case of a uniform distribution all the Rényi entropies of are equal to $\log n$. Otherwise the entropies are weakly decreasing as a function of α. The following are some particular cases:

$\alpha = 0$	$H_0(X) = \log \lvert X \rvert = \log n$	Hartley entropy
$\alpha \to 1$	$H_1(X) = -\sum_i p(x_i) \log p(x_i)$	Shannon entropy
$\alpha \to \infty$	$H_\infty(X) = -\log \max_i p(x_i)$	min-entropy

Shannon conditional entropy of X given Y represents the average residual entropy of X once the value of Y is known, and it is defined as

$$H_1(X|Y) = -\sum_y p(y) H_1(X|Y=y) = \sum_{ij} p(x_i, y_j) \log p(x_i|y_j) = H_1(X,Y) - H_1(Y)$$

where $H_1(X,Y)$ represents the entropy of the conjunction of X and Y.

The mutual information of X and Y represents the correlation of information between X and Y. It is defined as

$$I(X;Y) = H_1(X) - H_1(X|Y) = H_1(X) + H_1(Y) - H_1(X,Y)$$

It is possible to show that $I(X;Y) \geq 0$, with $I(X;Y) = 0$ iff X and Y are independent.

2.3 Convexity and corner points

We recall here some basic notions of convexity. Let $\mathbb{R}$ be the set of real numbers. The elements $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ constitute a set of *convex coefficients* if, for every $i \in \{1, \ldots, k\}$, $\lambda_i \geq 0$ and $\sum_k \lambda_k = 1$. Given a vector space V, a *convex combination* of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in V$ is any vector of the form $\sum_i \lambda_i \vec{v}_i$ where $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ are a set of convex coefficients.

Definition 2.1 (Convex set) A subset S of a vector space is *convex* if every convex combination of vectors in S is in S.

In the following we will denote by $D^{(n)}$ the domain of probability distributions of dimension n. It is easy to see that, for every n, $D^{(n)}$ is convex.

We give now the definition of *convex function*.

Definition 2.2 (Convex function) Given a convex subset S of a vector space V, and a function $f : S \to \mathbb{R}$, we say that f is *convex* if for any $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in S$ and any set of convex coefficients $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$, we have $f(\sum_i \lambda_i \vec{v}_i) \leq \sum_i \lambda_i f(\vec{v}_i)$. A function f is *concave* if its opposite $-f$ is convex.

We now introduce (with a slight abuse of terminology) the concept of *convex base*.

Given a subset S of V, the *convex hull* of S, which we will denote by $ch(S)$, is the smallest convex set containing S. Since the intersection of convex sets is convex, it is clear that $ch(S)$ always exists.

Definition 2.3 Given two vector sets S and U, we say that U is a convex base for S if $U \subseteq S$ and $S \subseteq ch(U)$.

In the following, for a given vector $\vec{v} = (v_1, v_2, \ldots, v_n)$, we will use the notation $(\vec{v}, f(\vec{v}))$ to denote the vector (with one additional dimension) $(v_1, v_2, \ldots, v_n, f(\vec{v}))$. Similarly, given a vector set S in a n-dimensional space, we will use the notation $(S, f(S))$ to represent the set of vectors $\{(\vec{v}, f(\vec{v})) \mid \vec{v} \in S\}$ in a $(n+1)$-dimensional space. The notation $f(S)$ represents the image of S under f, i.e. $f(S) = \{f(\vec{v}) \mid \vec{v} \in S\}$.

Definition 2.4 Given a vector set S, a convex base U of S, and a function $f : S \to \mathbb{R}$, we say that U is a set of *corner points* of f if $(U, f(U))$ is a convex base for $(S, f(S))$. We also say that f is *convexly generated* by $f(U)$.

In other words, if U is a set of corner points of f, then for every $\vec{v} \in D^{(n)}$, there are elements $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$ in U and $\lambda_1, \lambda_2, \ldots, \lambda_k$ in $\mathbb{R}$ such that $\vec{v} = \sum_i \lambda_i \vec{u}_i$ and $f(\vec{v}) = \sum_i \lambda_i f(\vec{u}_i)$.

3 Mutual Information and Capacity

In [12] Smith proved that the notion of mutual information (based on Shannon entropy) is not suitable to express the information leakage when the adversary tries to guess the value of the input in one single try. In fact, he showed an example of systems with the same mutual information, the same a priori uncertainty, and for which the probabilities of guessing the right input a posteriori (i.e. after observing the output) are very different.

The systems in Smith's example are deterministic, i.e. have the property that the input determines univocally the output. For such systems, it turns out that the discrepancy observed for the mutual information does not arise in the case of the capacity. Surprisingly, indeed, Smith showed in [13] that, under a uniform input distribution, the capacity is equal to the logarithm of the ratio between the a posteriori and the a priori probability of making a right guess, which coincides with his proposal for the notion of leakage. We will come back on this point in the next section.

Unfortunately, this coincidence does not carry out to the more general case of probabilistic channels, and, worse yet, the notion of capacity suffers (in the general case) from the same problem as the mutual information. The following example illustrates the situation.

Example 3.1 Consider the following channels:

	y_1	y_2	y_3
x_1	2/3	1/6	1/6
x_2	1/6	2/3	1/6
x_3	1/6	1/6	2/3

Fig. 1. Channel matrix C

It is easy to see that the Shannon capacity of C is 1/3, while the one of D is 2/3. However, under the uniform input distribution, the ratio between the a posteriori and the a priori probability of making the right guess is the same (2).

	y_1	y_2	y_3
x_1	2/3	1/3	0
x_2	0	2/3	1/3
x_3	1/3	0	2/3

Fig. 2. Channel matrix D

4 Towards a more suitable notion of leakage

In the following, we are interested in quantifying the *leakage* of a security protocol, i.e. the amount of information about the input that an adversary can learn by running the protocol and observing the resulting output.

4.1 Probabilities of a right guess

Before running the protocol, the probability that a given input x_i occurs depends only on the a priori distribution $\vec{\pi}$, and a rational adversary will therefore assume that the most probable input, called the *a priori probability of a right guess* $PR_i(\vec{\pi})$, will be the input having the maximum a priori probability, i.e.:

Definition 4.1 The a priori probability of a right guess is defined as

$$PR_i(\vec{\pi}) = \max_i \pi_i$$

After running the protocol and seeing the output, the adversary may revise his guess. An adversary applying the MAP rule, when observing output y_j, will choose as most probable input x_i the one for which the a posteriori probability $p(x_i|y_j)$ is the highest. The average of this value on all possible outputs gives the *a posteriori probability of a right guess* $PR_o(\vec{\pi})$, which is the complement of the Bayes risk.

Definition 4.2 The a posteriori probability of a right guess is defined as

$$PR_o(\vec{\pi}) = \sum_j \max_i (p(y_j|x_i)\pi_i)$$

In the rest of this paper, we will consider only adversaries applying the MAP rule since this is the rule that gives the best result (from the point of view of the adversary).

4.2 Leakage and uncertainty

Intuitively, the *leakage* is the amount of information learnt by the adversary by observing the output of the protocol. Following [13], it seems natural to define it as the difference between the uncertainty about the input before observing the output, and the remaining uncertainty afterwards:

$$\text{Information Leaked} = \text{Initial Uncertainty} - \text{Remaining Uncertainty} \quad (1)$$

Now, the question is how to measure information, and (correspondingly) what do we actually mean by uncertainty. We consider here two possibilities. The first

leads to a multiplicative notion of leakage, and it follows the proposal of Smith [13]. The second leads to an additive notion, and it is new.

4.3 Multiplicative leakage

In relation to Equation (1), Smith [13] measures the information in bits, and proposes to define the initial uncertainty as the *min-entropy* of X, $H_\infty(X)$, the instance of Rényi entropy [11] obtained for $\alpha = \infty$. As for the remaining uncertainty, it would be natural to use the conditional *min-entropy* of X given Y. Unfortunately there is no agreement on what Rényi's generalization of Shannon's conditional entropy should be, even though there seem to be a consensus towards $\sum_y p(y) H_\alpha(X|Y=y)$ [1]. Smith however proposes to use the definition of $H_\infty(X|Y)$ equivalent to the one given in [6], which is

$$H_\infty(X|Y) = -\log PR_o(\vec{\pi})$$

In this way, Smith obtains a definition of leakage similar to the definition of mutual information, except that Shannon entropy is replaced by H_∞:

$$L(X;Y) \;=\; H_\infty(X) - H_\infty(X|Y) \;=\; \log \frac{PR_o(\vec{\pi})}{PR_i(\vec{\pi})}$$

We consider a similar definition for leakage, namely the ratio between $PR_o(\vec{\pi})$ and $PR_i(\vec{\pi})$, which coincides with Smith's notion apart from the absence of the logarithm. Furthermore, in general we want to abstract from the a priori distribution, and consider the worst case, hence we are particularly interested in the supremum of such ratio.

Definition 4.3 We define the multiplicative leakage as

$$\mathcal{L}_\times(\vec{\pi}) \;=\; \frac{PR_o(\vec{\pi})}{PR_i(\vec{\pi})}$$

We will also use the notation $\mathcal{ML}_\times$ to represent the supremum of this quantity:

$$\mathcal{ML}_\times \;=\; \max_{\vec{\pi}}(\mathcal{L}_\times(\vec{\pi}))$$

Note that $PR_i(\vec{\pi}) > 0$ for every $\vec{\pi}$, hence $\mathcal{L}_\times(\vec{\pi})$ is always defined.

4.4 Additive leakage

Another possible interpretation for Equation (1) is to consider the uncertainty as the probability of guessing the wrong input. The leakage then expresses how much the knowledge of the observable helps decreasing such probability. This leads to define the leakage as the difference between the probabilities of error before and after observing the output. As usual, we are particularly interested in the supremum of this difference.

Definition 4.4 We define the additive leakage as

$$\mathcal{L}_+(\vec{\pi}) \;=\; PR_o(\vec{\pi}) - PR_i(\vec{\pi})$$

We will also use the notation $\mathcal{ML}_+$ to represent the supremum of this quantity:

$$\mathcal{ML}_+ \;=\; \max_{\vec{\pi}}(\mathcal{L}_+(\vec{\pi}))$$

5 Properties of the mutiplicative leakage

In this section we consider the multiplicative leakage and we study its supremum. It turns out that the supremum is very easy to compute. In fact, it coincides with the value of the leakage in the point of uniform distribution, and it is equal to the sum of the maxima of the columns. This property was also discovered independently by Geoffrey Smith and Ziyuan Meng (personal communication).

Proposition 5.1

$$\mathcal{ML}_\times = \mathcal{L}_\times(\pi_u) = \sum_j \max_i p(y_j|x_i)$$

where π_u is the uniform distribution.

Proof

$$\begin{aligned}
\mathcal{L}_\times(\pi_1,\ldots,\pi_n) &= \tfrac{1}{\max_i \pi_i} \textstyle\sum_j \max_i(p(y_j|x_i)\pi_i) \\
&\leq \tfrac{1}{\max_i \pi_i} \textstyle\sum_j (\max_i p(y_j|x_i))(\max_i \pi_i) \\
&= \textstyle\sum_j \max_i p(y_j|x_i) \\
&= n \textstyle\sum_j \max_i (p(y_j|x_i)\tfrac{1}{n}) \\
&= \mathcal{L}_\times(\tfrac{1}{n},\ldots,\tfrac{1}{n})
\end{aligned}$$

□

6 Properties of the additive leakage

We turn now our attention to the additive leakage. We will see that the supremum is not always in the point of uniform distribution. However, we prove that it is in one of the corner points of PR_i. Since PR_i has a finite set of corner points, and their form is known, also the additive leakage is relatively easy to compute.

First we prove a general property concerning the relation between suprema, convexity, and corner points:

Proposition 6.1 *Consider two functions $f, g : D^{(n)} \to \mathbb{R}$ where f has a set of corner points U, and g is convex. Define $h : D^{(n)} \to \mathbb{R}$ as $h(\vec{v}) = f(\vec{v}) + g(\vec{v})$. Let $\vec{w} \in D^{(n)}$ be a point in which h has a maximum, i.e. for every $\vec{v} \in D^{(n)}$, $h(\vec{v}) \leq h(\vec{w})$. Then there exists $\vec{u} \in U$ such that $h(\vec{w}) = h(\vec{u})$. Namely, the maximum value of h is in a corner point of f.*

Proof
By contradiction. Suppose that for every $\vec{u} \in U$, $h(\vec{u}) < h(\vec{w})$. Since $\vec{w} \in D^{(n)}$, there are elements $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ in U and $c_1, c_2, \ldots, c_k$ in $\mathbb{R}$ such that $\vec{w} = \sum_i c_i \vec{v}_i$

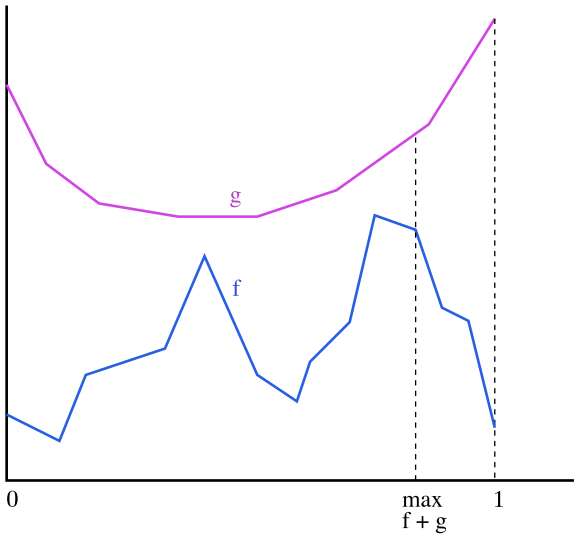

Fig. 3. An illustration of Proposition 6.1

and $f(\vec{x}) = \sum_i c_i f(\vec{v}_i)$. Then

$$\begin{aligned}
h(\vec{w}) &= h(\textstyle\sum_i c_i \vec{v}_i) \\
&= f(\textstyle\sum_i c_i \vec{v}_i) + g(\textstyle\sum_i c_i \vec{v}_i) \\
&= \textstyle\sum_i c_i f(\vec{v}_i) + g(\textstyle\sum_i c_i \vec{v}_i) \\
&\leq \textstyle\sum_i c_i f(\vec{v}_i) + \textstyle\sum_i c_i g(\vec{v}_i) && \text{since } g \text{ is convex} \\
&= \textstyle\sum_i c_i h(\vec{v}_i) \\
&< \textstyle\sum_i c_i h(\vec{w}) && \text{since } \vec{v}_i \in U \\
&= h(\vec{w}) && \text{since } \textstyle\sum c_i = 1
\end{aligned}$$

□

An example of Proposition 6.1 is illustrated in Figure 3.

We now show that $-PR_i$ and PR_o satisfy the hypotheses of Proposition 6.1. The necessary property for $-PR_i$ comes from a result in [3].

Proposition 6.2 ([3], Proposition 3.9) *The function PR_i on $D^{(n)}$ is convexly generated by $(U, f(U))$ with $U = U_1 \cup U_2 \cup \ldots \cup U_n$ where, for each r, U_r is the set of all vectors that have value $1/r$ in exactly r components, and 0 everywhere else.*

Remark 6.3 The function $-PR_i$ has the same corner points as PR_i.

We now prove that PR_o satisfy the necessary property.

Proposition 6.4 *PR_o is convex.*

Proof

Let $\vec{z}$ be the convex combination $\sum_i \lambda_i \vec{z}_i$ where the dimension of $\vec{z}, \vec{z}_1, \ldots, \vec{z}_m$ corresponds to the number of input variables and $\vec{z}_1, \ldots, \vec{z}_m$ is a set of corner points.

The j^{th} component z_{k_j} of any corner point $\vec{z_k}$ corresponds to the input variable x_{k_j} chosen according to the MAP rule when the output variable y_j is obtained.

$$\begin{aligned} PR_o(\textstyle\sum_i \lambda_i \vec{z_i}) &= \textstyle\sum_j \max_k \{p(y_j|x_k)(\sum_i \lambda_i \vec{z_i})_k\} \\ &= \textstyle\sum_j p(y_j|x_{k_j})(\sum_i \lambda_i \vec{z_i})_{k_j} \\ &= \textstyle\sum_j p(y_j|x_{k_j})(\sum_i \lambda_i z_{i,k_j}) \\ &= \textstyle\sum_j \sum_i \lambda_i p(y_j|x_{k_j}) z_{i,k_j} \\ &= \textstyle\sum_i \lambda_i \sum_j p(y_j|x_{k_j}) z_{i,k_j} \\ &\leq \textstyle\sum_i \lambda_i \sum_j \max_k p(y_j|x_k) z_{i,k} \\ &= \textstyle\sum_i \lambda_i PR_o(\vec{z_i}) \end{aligned}$$

□

Corollary 6.5 *$\mathcal{ML}_+$ is reached on one of the corner points of PR_i.*

Proof
Since $\mathcal{L}_+ = PR_o - PR_i$ and PR_o is convex, Proposition 6.1 shows that if $\mathcal{ML}_+$ exists, it must be reached on a corner point of $-PR_i$, which correspond to the corner points of PR_i. □

Remark 6.6 In general $\mathcal{ML}_+$ is not reached on the point of uniform distribution.

Example 6.7 Consider the channel whose matrix is given in Figure 4.

	y_1	y_2	y_3
x_1	1	0	0
x_2	0	$1-e$	e
x_3	0	$1-2e$	$2e$

Fig. 4. Channel matrix ($e \in [0, 1/2]$)

The calculation of $\mathcal{L}_+$ on the distributions corresponding to the corner points gives:

Corner points	PR_o	PR_i	$\mathcal{L}_+$
$(1,0,0),(0,1,0),(0,0,1)$	1	1	0
$(1/2,1/2,0),(1/2,0,1/2)$	1	$1/2$	$1/2$
$(0,1/2,1/2)$	$(e+1)/2$	$1/2$	$e/2$
$(1/3,1/3,1/3)$	$(e+2)/3$	$1/3$	$(e+1)/3$

We have for every $e \in [0, 1/2[$,

$$0 = \mathcal{L}_+(1,0,0) \leq \mathcal{L}_+(0,1/2,1/2) < \mathcal{L}_+(1/3,1/3,1/3) < \mathcal{L}_+(1/2,1/2,0) = 1/2$$

and $\mathcal{L}_+(1/3,1/3,1/3) = \mathcal{L}_+(1/2,1/2,0) = 1/2$ for $e = 1/2$. Therefore if $e < 1/2$, $\mathcal{ML}_+ = 1/2$, reached on distributions that are different from the uniform distribution $(1/3,1/3,1/3)$.

Moreover, this remark holds also for symmetric matrices:

Remark 6.8 Even in case of symmetric matrices, in general $\mathcal{ML}_+$ is not reached on the point of uniform distribution.

Example 6.9 Consider the channel whose matrix is given in Figure 5.

	y_1	y_2	y_3	$\ldots$	y_{10}	y_{11}
x_1	0	1/10	1/10	$\ldots$	1/10	1/10
x_2	1/10	0	1/10	$\ldots$	1/10	1/10
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_{10}	1/10	1/10	1/10	$\ldots$	0	1/10
x_{11}	1/10	1/10	1/10	$\ldots$	1/10	0

Fig. 5.

Let $\vec{\pi} = (1/r, 1/r, \ldots, 1/r, 0, \ldots, 0)$ be the a priori distribution with an equal probability of $1/r$ for the r first inputs. This distribution is a corner point of the matrix, and since the matrix is symmetric, any other corner point corresponding to a distribution containing r non-null probabilities of $1/r$ will give the same results for PR_i, PR_o and $\mathcal{L}_+$.

$$PR_i(\vec{\pi}) = 1/r$$

$$\begin{aligned} PR_o(\vec{\pi}) &= \textstyle\sum_{1,\ldots,11}(1/r)(1/10) \\ &= 11/(10r) \end{aligned}$$

$$\begin{aligned} \mathcal{L}_+(\vec{\pi}) &= PR_o(\vec{\pi}) - PR_i(\vec{\pi}) \\ &= 1/(10r) \end{aligned}$$

Therefore $\mathcal{ML}_+$ is reached when r has the smallest value, i.e. when $r = 2$. This corresponds to the distribution $(1/2, 1/2, 0, \ldots, 0)$ and gives $\mathcal{ML}_+ = 1/20$, while $\mathcal{L}_+(1/11, \ldots, 1/11) = 1/110$.

7 Comparison

In this section, we compare the two notions of leakage. We first compare them with respect to a specific distribution, and then we consider the comparison of their

worst cases.

If we consider a specific distribution, it comes out that the two notions are equivalent, in the sense that a program is better with respect to the additive notion if and only if it is better with respect to the multiplicative notion.

Proposition 7.1 *Consider two programs P and P', and let $\mathcal{L}_+$ and $\mathcal{L}_+{}'$ be the additive measures of leakage for P and P', respectively. Analogously, let $\mathcal{L}_\times$ and $\mathcal{L}_\times{}'$ be the multiplicative measures of leakage for P and P', respectively. We have that, for every $\vec{\pi}$*

$$\mathcal{L}_+(\vec{\pi}) \;\leq\; \mathcal{L}_+{}'(\vec{\pi}) \quad\Leftrightarrow\quad \mathcal{L}_\times(\vec{\pi}) \;\leq\; \mathcal{L}_\times{}'(\vec{\pi})$$

Proof
Let PR_o and $PR_o{}'$ be the a posteriori probability of a right guess for P and P', respectively. Analogously, let PR_i and $PR_i{}'$ be the a priori probability of a right guess for P and P', respectively. Observe that $PR_i(\vec{\pi}) = \max_i \pi_i = PR_i{}'(\vec{\pi})$. Hence

$$\begin{aligned}
\mathcal{L}_+(\vec{\pi}) \;\leq\; \mathcal{L}_+{}'(\vec{\pi}) \quad&\Leftrightarrow\quad PR_o(\vec{\pi}) - PR_i(\vec{\pi}) \;\leq\; PR_o{}'(\vec{\pi}) - PR_i{}'(\vec{\pi}) \\
&\Leftrightarrow\quad \frac{PR_o(\vec{\pi}) - PR_i(\vec{\pi})}{\max_i \pi_i} \;\leq\; \frac{PR_o{}'(\vec{\pi}) - PR_i{}'(\vec{\pi})}{\max_i \pi_i} \\
&\Leftrightarrow\quad \frac{PR_o(\vec{\pi}) - PR_i(\vec{\pi})}{PR_i(\vec{\pi})} \;\leq\; \frac{PR_o{}'(\vec{\pi}) - PR_i{}'(\vec{\pi})}{PR_i{}'(\vec{\pi})} \\
&\Leftrightarrow\quad \frac{PR_o(\vec{\pi})}{PR_i(\vec{\pi})} - 1 \;\leq\; \frac{PR_o{}'(\vec{\pi})}{PR_i{}'(\vec{\pi})} - 1 \\
&\Leftrightarrow\quad \mathcal{L}_\times(\vec{\pi}) \;\leq\; \mathcal{L}_\times{}'(\vec{\pi})
\end{aligned}$$

□

Another criterion of comparison is the worst case. We consider the two notions on some examples.

Example 7.2 Let us consider a $2^k \times 2^k$ channel with the 2^k first natural numbers as inputs and outputs, i.e. $\mathcal{X} = \mathcal{Y} = \{0, \ldots, 2^k - 1\}$. Consider a random input variable X with values ranging in $\{0, 2^k - 1\}$.

Consider the following program:

Program P(X)

```
1  ▷ Input X
2  if X = 0 or X = 1
3      then Output X
4      else Output one of the values {2, ..., 2^k − 1} chosen randomly
            according to the uniform distribution
```

This program corresponds to a channel whose matrix is given in Figure 6.

	0	1	2	...	2^k-1
0	1	0	0	...	0
1	0	1	0	...	0
2	0	0	p	...	p
...	...	...	...	...	...
2^k-1	0	0	p	...	p

Fig. 6. Channel matrix ($p = 1/(2^k-2) = 1/2^{k-1}$)

Let us consider first $\mathcal{ML}_+$. Because of Corollary 6.5, we know that $\mathcal{ML}_+$ is reached on a corner point, i.e. a distribution of the form $(q_1, \ldots, q_{2^k})$ where each q_i is either 0 or $1/r$, and there are r elements with value $1/r$ in the distribution.

For every corner point $\vec{\pi}$ we have $PR_i(\vec{\pi}) = 1/r$, thus maximizing $\mathcal{L}_+$ for a given r is equivalent to maximizing PR_o. From the channel matrix, one can see that the maximum value of PR_o is reached on an input distribution where the two first elements are as high as possible.

Therefore, we can restrict our study to distributions of the form $(1/r, \ldots, 1/r, 0 \ldots, 0)$, i.e. distributions where the elements with value $1/r$ are the r first elements.

For $r = 1$, we have:

$$PR_i(1, 0, \ldots, 0) = 1$$

$$PR_o(1, 0, \ldots, 0) = 1$$

Thus:

$$\mathcal{L}_+(1, 0 \ldots, 0) = 0$$

$$\mathcal{L}_\times(1, 0, \ldots, 0) = 1$$

For $r = 2$, we have:

$$PR_i(1/2, 1/2, 0 \ldots, 0) = 1/2$$

$$PR_o(1/2, 1/2, 0, \ldots, 0) = 1$$

Thus:

$$\mathcal{L}_+(1/2, 1/2, 0 \ldots, 0) = 1/2$$

$$\mathcal{L}_\times(1/2, 1/2, 0, \ldots, 0) = 2$$

For $r \geq 3$, we have:

$$PR_i(1/r, 1/r, 1/r, 0, \ldots, 0) = 1/r$$

$$\begin{aligned} PR_o(1/r, 1/r, 1/r, 0, \ldots, 0) &= 1/r + 1/r + (2^k - 2) * (1/r) * p \\ &= 3/r \end{aligned}$$

Thus:

$$\mathcal{L}_\times(1/r, 1/r, 1/r, 0, \ldots, 0) = 3$$

$$\mathcal{L}_+(1/r, 1/r, 1/r, 0, \ldots, 0) = 2/r$$

We observe that for $r \geq 3$, the value of $\mathcal{L}_+$ decreases when r increases. Since $\mathcal{L}_+(1/3, 1/3, 1/3, 0, \ldots, 0) = 2/3 > \mathcal{L}_+(1/2, 1/2, 0, \ldots, 0) = 1/2 > \mathcal{L}_+(1, 0, \ldots, 0) = 0$, we have $\mathcal{ML}_+ = 2/3$ reached for $r = 3$.

In particular, $\mathcal{ML}_+ > \mathcal{L}_+(1/2^k, \ldots, 1/2^k) = 1/2^{k-1}$ for all $k > 1$.

Concerning $\mathcal{L}_\times$, we have that, for $r \geq 3$, $\mathcal{L}_\times(1/r, 1/r, 1/r, 0, \ldots, 0) = 3 > \mathcal{L}_\times(1/2, 1/2, 0, \ldots, 0) = 2 > \mathcal{L}_\times(1, 0, \ldots, 0) = 1$, thus $\mathcal{ML}_\times = 3$, reached on any distribution $(1/r, 1/r, 1/r, 0, \ldots, 0)$ with $r \geq 3$, and in particular on the uniform distribution, which confirms Proposition 5.1.

Example 7.3 Let us consider the following program:

Program P$'(X)$

1 ▷ Input X
2 with probability $3/2^k$ Output X
3 with probability $1 - 3/2^k$ Output a value in $\{0, 2^k - 1\}\backslash\{X\}$ chosen randomly according to the uniform distribution

This program corresponds to a channel whose matrix is given in Figure 7.

	0	1	2	...	$2^k - 1$
0	p_1	p_2	p_2	...	p_2
1	p_2	p_1	p_2	...	p_2
2	p_2	p_2	p_1	...	p_2
...	...	...	...	...	...
$2^k - 1$	p_2	p_2	p_2	...	p_1

Fig. 7. Channel matrix ($p_1 = 3/2^k$ and $p_2 = (1 - (3/2^k))/(2^k - 1)$)

The symmetry of the matrix implies that we can restrict the study to the a priori distribution $\vec{\pi} = (1/r, 1/r, \ldots, 1/r, 0, \ldots, 0)$, where the r elements with value $1/r$ are the first elements in the distribution.

In this case, for $r \geq 1$:

$$PR_i(\vec{\pi}) = 1/r$$

$$\begin{aligned} PR_o(\vec{\pi}) &= r(p_1/r) + (2^k - r)(p_2/r) \\ &= p_1 + [(2^k/r) - 1]p_2 \end{aligned}$$

Finally:

$$\mathcal{L}_+(\vec{\pi}) = p_1 - p_2 - \tfrac{2}{r(2^k-1)}$$

Thus $\mathcal{L}_+$ increases when r increases, and $\mathcal{ML}_+ = 1/2^{k-1}$ is reached for $r = 2^k$ (on the uniform distribution).

$$\begin{aligned}\mathcal{L}_\times(\vec{\pi}) &= rp_1 + (2^k - r)p_2 \\ &= r(p_1 - p_2) + 2^k p_2\end{aligned}$$

Since $p_1 > p_2$, $\mathcal{L}_\times$ increases when r increases, and thus $\mathcal{ML}_\times = 3$ is obtained for $r = 2^k$ (on the uniform distribution, which confirms Proposition 5.1).

The programs P and P' have therefore the same worst-case multiplicative measures of leakage $\mathcal{ML}_\times = 3$, but the worst case of the additive measure of leakage $\mathcal{ML}_+$ is equal to $2/3$ for P and equal to $1/2^{k-1}$ for P'.

8 Conclusion

We have considered two notions of leakage related to the Bayes risk. One of them, which we call multiplicative, corresponds to the notion recently proposed by Smith based on Renyi min-entropy. The other, which we call additive, is new. We have shown that the two notions are equivalent in all distributions. If we consider the distributions that give the worst case for the leakage, however, then the two notions are different. In particular, the multiplicative one has the worst case always in correspondence of the uniform distribution, while this is not the case for the additive one. So we can consider the new notion as a criterion, in addition to the one of Smith, to help assessing the degree of protection offered by a protocol or a program.

Acknowledgement

We are grateful to Pasquale Malacaria and to Geoffrey Smith for the many fruitful discussions.

References

[1] Cachin, C., "Entropy Measures and Unconditional Security in Cryptography," Ph.D. thesis, Swiss Federal Institute of Technology, Zürich, Switzerland (1997).
URL ftp://ftp.inf.ethz.ch/pub/publications/dissertations/th12187.ps.gz

[2] Chatzikokolakis, K., C. Palamidessi and P. Panangaden, *Anonymity protocols as noisy channels*, Information and Computation **206** (2008), pp. 378–401, http://www.lix.polytechnique.fr/~catuscia/papers/Anonymity/Channels/full.pdf.
URL http://hal.inria.fr/inria-00349225/en/

[3] Chatzikokolakis, K., C. Palamidessi and P. Panangaden, *On the bayes risk in information-hiding protocols*, Journal of Computer Security **16** (2008), pp. 531–571.
URL http://hal.inria.fr/inria-00349224/en/

[4] Clark, D., S. Hunt and P. Malacaria, *Quantitative information flow, relations and polymorphic types*, Journal of Logic and Computation, Special Issue on Lambda-calculus, type theory and natural language **18** (2005), pp. 181–199.

[5] Cover, T. M. and J. A. Thomas, "Elements of Information Theory," John Wiley & Sons, Inc., 1991.

[6] Dodis, Y., R. Ostrovsky, L. Reyzin and A. Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM J. Comput **38** (2008), pp. 97–139.
URL http://dx.doi.org/10.1137/060651380

[7] Malacaria, P., *Assessing security threats of looping constructs*, in: M. Hofmann and M. Felleisen, editors, *Proceedings of the 34th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2007, Nice, France, January 17-19, 2007* (2007), pp. 225–235.
URL http://doi.acm.org/10.1145/1190216.1190251

[8] Malacaria, P. and H. Chen, *Lagrange multipliers and maximum information leakage in different observational models*, in: Úlfar Erlingsson and Marco Pistoia, editor, *Proceedings of the 2008 Workshop on Programming Languages and Analysis for Security (PLAS 2008)* (2008), pp. 135–146.

[9] Moskowitz, I. S., R. E. Newman, D. P. Crepeau and A. R. Miller, *Covert channels and anonymizing networks.*, in: S. Jajodia, P. Samarati and P. F. Syverson, editors, *WPES* (2003), pp. 79–88.

[10] Moskowitz, I. S., R. E. Newman and P. F. Syverson, *Quasi-anonymous channels*, in: *IASTED CNIS*, 2003, pp. 126–131.

[11] Rény, A., *On measures of entropy and information*, in: *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, 1960, pp. 547–561.

[12] Smith, G., *Adversaries and information leaks (tutorial)*, in: G. Barthe and C. Fournet, editors, *Proceedings of the Third Symposium on Trustworthy Global Computing*, Lecture Notes in Computer Science **4912** (2007), pp. 383–400.
URL http://dx.doi.org/10.1007/978-3-540-78663-4_25

[13] Smith, G., *On the foundations of quantitative information flow*, in: L. De Alfaro, editor, *Proceedings of the 12th International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2009)*, Lecture Notes in Computer Science **5504** (2009), pp. 288–302.

[14] Zhu, Y. and R. Bettati, *Anonymity vs. information leakage in anonymity systems*, in: *Proc. of ICDCS* (2005), pp. 514–524.