Quantifying Leakage in the Presence of Unreliable Sources of Information

Sardaouna Hamadou, Catuscia Palamidessi, Vladimiro Sassone

Quantifying Leakage in the Presence of Unreliable Sources of Information

Sardaouna Hamadou[a], Catuscia Palamidessi[a,1], Vladimiro Sassone[b,2]

[a] *INRIA and LIX, Ecole Polytechnique*
[b] *University of Southampton*

Abstract

Belief and min-entropy leakage are two well-known approaches to quantify information flow in security systems. Both concepts stand as alternatives to the traditional approaches founded on Shannon entropy and mutual information, which were shown to provide inadequate security guarantees. In this paper we unify the two concepts in one model so as to cope with the frequent (potentially inaccurate, misleading or outdated) attackers' side information about individuals on social networks, online forums, blogs and other forms of online communication and information sharing. To this end we propose a new metric based on min-entropy that takes into account the adversary's beliefs.

Keywords: Information hiding, quantitative information flow, belief combination, probabilistic models, uncertainty, accuracy

1. Introduction

Protecting *sensitive* and *confidential* data is becoming increasingly important in many fields of human activities, such as electronic communication, auction, payment and voting. Many protocols for protecting confidential information have been proposed in the literature. In recent years the frameworks for reasoning, designing, and verifying these protocols have considered

[1]The work of Catuscia Palamidessi was partially supported by the INRIA Large Scale Initiative CAPPRIS (Collaborative Action for the Protection of Privacy Rights in the Information Society).

[2]The work of Vladimiro Sassone was partially supported by the project Horizon 2020: SUNFISH

probabilistic aspects and techniques for two reasons. First, the data to be protected often range in domains naturally subject to statistical considerations. Second and more important, the protocols often use randomised primitives to obfuscate the link between the information to be protected and the observable outcomes. This is the case, e.g., of the DCNets [1], Crowds [2], Onion Routing [3], and Freenet [4].

From the formal point of view, the *degree of protection* is the converse of the *leakage*, i.e. the amount of information about the secrets that can be deduced from the observables. Early approaches to information hiding in literature were the so-called *possibilistic approaches*, in which the probabilistic aspects were abstracted away and replaced by non-determinism. Some examples of these approaches are those based on *epistemic logic* [5, 6], on *function views* [7], and on *process calculi* [8, 9]. Subsequently, however, it has been recognised that the possibilistic view is too coarse, in that it tends to consider as equivalent randomized obfuscation methods that have very different degrees of protection.

The *probabilistic approaches* became therefore increasingly more popular. At the beginning they were investigated mainly at their strongest form of protection, namely to express the property that the observables reveal no (quantitative) information about the secrets (*strong anonymity*, *no interference*) [1, 6, 10]. Such strong property, however, is almost never achievable in practice. Hence, weaker notions of protection started to be considered. We mention in particular Rubin and Reiter's concepts of *possible innocence* and of *probable innocence* [2] and their variants explored in [11]. These are, however, still true-or-false properties. The need to express in a quantitative way the degree of protection has then lead naturally to explore suitable notions within the well-established fields of *Information Theory* and of *Statistics*.

Concepts from Information Theory [12] have proved quite useful 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 $s \in \mathcal{S}$ of the channel represents the information to be kept secret, and the output $o \in \mathcal{O}$ 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, usually by means of randomised mechanisms. Consequently, an input s may generate several different outputs o, according to a conditional probability distribution $p(o \mid s)$. These probabilities constitute the *channel matrix* $\mathcal{C}$. Similarly, for each output there may be several different corresponding inputs, according to the con-

verse conditional probability $p(s \mid o)$ which is linked to the above by the Bayes law: $p(s \mid o) = p(o \mid s)\, p(s)/p(o)$. The probability $p(s)$ is the *a priori* probability of s, while $p(s \mid o)$ is the *a posteriori* probability of s, after we know that the output is o. These probability distributions determine the *entropy* and the *conditional entropy* of the input, respectively. They represent the uncertainty about the input, before and after observing the output. The difference between entropy and conditional entropy is called the *mutual information* and expresses how much information is carried by the channel, i.e. how much uncertainty about the input we lose by observing the output (i.e., equivalently, how much information about the input we gain by observing the output).

Even though several notions of entropy have been proposed in Information Theory, Shannon's is by far the most famous of them, due to its relation with the *channel's rate*, i.e., the speed by which information can be transmitted accurately on a channel. Consequently, there have been various attempts to define the degree of protection by using concepts based on Shannon entropy, notably mutual information [13, 14, 15, 16] and the related notion of capacity, which is the supremum of the mutual information over all possible input distributions, and which therefore represents the worst case from the point of view of security [17, 18, 19].

A refinement of the above approaches came from the ideas of integrating the notions of extra knowledge and belief [20, 21, 22]. The idea is that the gain obtained by looking at the output should be relative to the possible initial knowledge or belief that an attacker may have about the secret. For instance, assume that in a parliament composed by m Labourists and n Conservatives, m members voted against a proposal to eliminate the minimum wage. Without any additional knowledge it is reasonable to believe that all Labourists voted against. If however we came to know that exactly one Conservative voted against, then it is more reasonable to believe that the most liberally-inclined Conservative voted against, and the least liberally-inclined Labourist voted in favour. In this case, the a posteriori belief is likely to be much more accurate than the a priori one, and the gain obtained using the knowledge about MPs' relative positioning on the left-to-right scale is much larger than the one computed as difference of entropies. Consequently, [22] proposed to define the protection of a system in terms of the difference (expressed in terms of Kullback-Leibler divergence) between the accuracy of the a posteriori belief and the accuracy of the a priori one.

Another criticism to the Shannon-entropy-based approach came from

Smith, who argued that it is not very suitable to model information leakage in the typical scenario of protocol attacks, where the adversary has only a limited number of tries to guess the value of the secret [23]. In such a scenario, the natural measure of the threat is the *probability* that the adversary guesses the right value. The case of "one-try only" was dubbed by Smith *vulnerability of the secret*. Shannon entropy, on the other hand, represents the expected number of attempts that an adversary has to make to discover the secret, assuming that there is no limit to such number, and that the adversary can narrow down the value by probing properties of the secret. Smith gave an example of two programs whose Shannon's mutual information is about the same, yet 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 [24], Smith proposed to define the leakage in terms of a notion of mutual information based on Rényi *min-entropy* (the logarithm of the vulnerability), which captures the case of an adversary disposing of one single try. Subsequent approaches going under the name of g-leakage have extended the analysis to multiple tries, and to the case in which each guess is associated with a gain (or loss) which depends on the level of approximation [25, 26, 27]. The min-entropy approach remains however the canonical framework, not only for its simplicity, but also because the worst-case min-leakage (aka min-capacity) has been proved to be an upper bound to the g-leakage [25].

In [28] the authors extended the vulnerability model of [24] in the context of the Crowds protocol for anonymous message posting to encompass the frequent situation where attackers have extra knowledge. They pointed out that in Crowds the adversary indeed has extra information (viz., the target servers) and assumed that she knows the correlation between that and the secret (viz., the users' preferences for servers). They proved that in such scenarios anonymity is more difficult to achieve.

In our opinion, a fundamental issue remains wide open: the need to measure and account for the *accuracy* of the adversary's extra knowledge. Indeed, [28] assumes that the adversary's extra information is accurate, and such an assumption is generally not warranted. Inaccuracy can indeed arise, e.g. from people giving deliberately wrong information, or simply from outdated data. As already noticed in [22] there is no reason in general to assume that the probability distributions the attacker uses are correct, and therefore they must be treated as *beliefs*.

This paper fills this gap by generalising the model on Rényi min-entropy

to cope with the presence of the attacker's beliefs. To this end we propose a new metric based on the concept of vulnerability that takes into account the adversary's beliefs. The idea is that the attacker does not know the *actual* probability distributions (i.e., the a priori distribution of the protocol's hidden input and its correlation with the extra information), and is assuming them. The *belief-vulnerability* is then the expected probability of guessing the value of the hidden input in *one try* given the adversary's belief. Informally, the adversary chooses the value of the secret input which has the maximum a posteriori probability according to her belief. Then the vulnerability of the secret input is expressed in terms of the actual a posteriori probabilities of the adversary's possible choices. We show the strength of our definitions both in terms of their theoretical properties and their utility by applying them to various threat scenarios and comparing the results to the previous approaches. Among its several advantages, our model allows to identify the levels of accuracy for the adversary's beliefs which are compatible with the security of a given program or protocol.

This paper revises and expends an earlier version [29]. First, we have simplified the model of the adversary's belief. Indeed, in the previous model, an attacker's belief consists of a pair: an assumed prior distribution over the set of secret inputs and an external channel leaking some (potentially incorrect) information about these secret inputs to the adversary. In this exposition, we simply model the belief by a subjective probability distribution, which summarises the aggregated information about the secret initially collected by the attacker from different and potentially conflicting sources of evidence. This approach is more in line with existing models of beliefs in the literature and simplifies significantly the exposition of our results and their comparison to existing work. Our approach is motivated further in §Appendix A. The flexibility of the model is illustrated in §6, where we apply our approach to different programs under various attack scenarios. We show in particular that a program which performs better than another in one threat scenario might become worse when the threat scenario evolves. Finally, exploiting notions and techniques from Dempster-Shafer Theory [30, 31], in §Appendix B we propose a new technique to estimate the reliability of an adversary's belief. This allows us to obtain a sound method of updating an arbitrary belief applying the *Bayes' rule* for updating a hypothesis.

The rest of the paper is organised as follows: in §2 we fix some basic notations and recall some fundamental notions of Information Theory; in §3 we briefly revise previous approaches to quantitative information follow;

§4 and §5 deliver our core technical contribution by extending the model on Rényi min entropy to the case of attacker's beliefs and investigating its theoretical properties; in §6 we apply our approach to various threat scenarios and compare it to the previous approaches; in §7 we discuss the related work whilst §8 contains our concluding remarks.

2. Preliminaries

In this section we briefly revise the elements of Information Theory which underpin the work in this paper, and illustrate our conceptual framework.

2.1. Some notions of Information Theory

Being in a purely probabilistic setting gives us the ability to use tools from information theory to reason about the uncertainty of a random variable and the inaccuracy of assuming a distribution for a random variable. In particular we are interested in the following notions: *entropy, mutual information, relative entropy* and *min-entropy.* We refer the reader to [32, 12] for more details.

We use capital letters X, Y to denote discrete random variables and the corresponding small letters x, y and calligraphic letters $\mathcal{X}$, $\mathcal{Y}$ for their values and set of values respectively. We denote by $p(x), p(y)$ the probability of x and y respectively and by $p(x, y)$ their joint probability.

Let X, Y be random variables. The *(Shannon) entropy* $H(X)$ of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{1}$$

The entropy measures the *uncertainty* of a random variable. It takes its maximum value $\log |\mathcal{X}|$ when X is uniformly distributed and its minimum value 0 when X is a constant. We take the logarithm with a base 2 and thus measure entropy in *bits*. The *conditional entropy*

$$H(X|Y) = -\sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log p(x|y) \tag{2}$$

measures the amount of uncertainty of X when Y is known. It can be shown that $0 \leq H(X|Y) \leq H(X)$ with the leftmost equality holding when Y completely determines the value of X and the rightmost one when Y reveals no information about X, i.e., X and Y are independent random variables.

Comparing $H(X)$ and $H(X|Y)$ give us the notion of *mutual information*, denoted $I(X;Y)$ and defined by

$$I(X;Y) = H(X) - H(X|Y). \tag{3}$$

It is non-negative, symmetric and bounded by $H(X)$. In other words,

$$0 \leq I(X;Y) = I(Y;X) \leq H(X).$$

The *relative entropy* or *Kullback-Leibler distance* between two probability distribution p and q on the same set $\mathcal{X}$, denoted $D(p \parallel q)$, is defined as

$$D(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. \tag{4}$$

It is non-negative (but not symmetric) and it is 0 if and only if $p = q$. The relative entropy measures the *inaccuracy* or *information divergence* of assuming that the distribution is q when the true distribution is p.

The *guessing entropy* $G(X)$ is the expected number of tries required to guess the value of X optimally. The optimal strategy is to guess the values of X in decreasing order of probability. Thus if we assume that $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and x_i's are arranged in decreasing order of probabilities, i.e., $p(x_1) \geq p(x_2) \geq \cdots \geq q(x_n)$, then

$$G(X) = \sum_{1 \leq i \leq n} ip(x_i). \tag{5}$$

The min-entropy $H_\infty(X)$ of a random variable is given by

$$H_\infty(X) = -\log \max_{x \in \mathcal{X}} p(x) \tag{6}$$

and measures the difficulty for an attacker to correctly guess the value of X in *one* try (obviously using the optimal strategy above). It can be shown that $H_\infty(X) \leq H(X)$ with equality when X is uniformly distributed. In general, $H(X)$ can be arbitrary higher than $H_\infty(X)$, since it can be arbitrary high even if X assumes a given value with probability close to 1.

2.2. Framework

In this paper we consider a framework similar to the probabilistic approaches to anonymity and information flow used e.g. in [6, 33, 34], and [24]. We restrict ourselves to *total* protocols and programs with one *high level* input A, a random variable over a finite set $\mathcal{A}$, and one *low level* output (observable) O, a random variable over a finite set $\mathcal{O}$. We represent a protocol/program by the matrix of the conditional probabilities $p(o_j \mid a_i)$ that the low output is o_j given that the high input is a_i. An adversary or eavesdropper can see the output of a protocol, but not its input, and she is interested in deriving the value of the input from the observed output in *one single* try.

In this paper we shall assume that the high input is generated according to an *a priori* probabilistic distribution $p(a_i)$ *unknown to the adversary*, and we also denote by $p_\beta(a_i)$ the subjective probability modeling the adversary's *initial belief*, as explained in the introduction[3]. In other words, $p_\beta(a_i)$ is her assumed *a priori* distribution of A.

Example 1. Let A be a random variable with an unknown (to the adversary) a priori distribution over $\mathcal{A} = \{0, 1, 2, 3\}$. Suppose that A is the high input of the deterministic program `C1` below, whose low output is

$$O = \begin{cases} 1 & \text{if } a \in \{0, 1\} \\ 2 & \text{otherwise.} \end{cases}$$

```
PROG C1:
  BEGIN
    O := ⌊ log(A + 2) ⌋
  END
```

Now suppose that the adversary, for some reason, initially believes that A is an odd number and that 1 is more likely than 3. For example $p_\beta(1) = 0.9$ and $p_\beta(3) = 0.1$. In the case of wrong belief, i.e., when A is actually an even number, her low observation of `PROG C1` does not allow her to realize that her initial belief is wrong. She will not be able to correct it and will therefore pick again the wrong value. Indeed, both observations 1 and 2 are compatible with the odd numbers. However, whereas observing $O = 1$

[3]See §Appendix A for the rationale for representing the adversary's belief by a subjective probability distribution.

$p(o \mid a)$	o_0	o_1	o_2
a_0	$1-\lambda$	λ	0
a_1	1	0	0
a_2	λ	0	$1-\lambda$
a_3	1	0	0

Table 1: Conditional probabilistic matrix of PROG C2

would strengthen her belief that A is 1, seeing a low output of 2 might raise some doubts if the adversary does not fully trust the source of her belief, as the unlikely 3 (viz., $p_\beta(3) = 0.1$) has happened. In Section Appendix B we present a novel technique allowing the adversary to estimate the reliability of her initial belief and to discount it accordingly.

Now suppose that A is the high input of the probabilistic program C2 below, with low output $O \in \{-1, 0, 2\}$ and conditional probabilistic matrix as in Table 1.

```
PROG C2:
  BEGIN
     R 'sampled from {0,2} with p(0) = λ and p(2) = 1 − λ';
     If A = R
       Then O := A
       Else O := −1
  END
```

Contrary to the PROG C1, the low output of PROG C2 may allow the adversary to realize that her initial belief about the parity of A is wrong. For example if she initially believes that A is odd and then she observes that O is either 0 or 2, then she knows that her belief is wrong. But the observation $O = -1$ does not help her to correct her inaccurate belief, as it can be induced by both even and odd numbers.

3. Uncertainty vs accuracy

This section reviews the existing definitions for quantifying information leakage. We begin by quantitative approaches to information flow based on Shannon entropy and mutual information, and recall why they fail generally to give good security guarantees. We then present alternative approaches based on the adversary's beliefs initially proposed by Clarkson, Myers and

Schneider [22]. We conclude the section by presenting more recent alternative approaches based on the concept of vulnerability [24] and Rényi min-entropy.

3.1. Shannon entropy approach

There seems to be a general consensus in the literature for using Shannon entropy to measure uncertainty and mutual information to quantify information leakage [35, 36, 14, 37, 33]. We remind the reader that these approaches aim at quantifying information flow as a reduction of the adversary uncertainty about the high input and take no account of the adversary's initial belief. Shannon entropy $H(A)$ as a measure of the uncertainty of A seems adequate to express the adversary's initial uncertainty about A. Similarly, as the conditional entropy $H(A \mid O)$ measures the remaining amount of uncertainty of A when O is known, it seems appropriate to express the adversary's remaining uncertainty. We thus have the following definitions.

- *initial uncertainty* (IU): $H(A)$
- *remaining uncertainty* (RU): $H(A \mid O)$
- *information leakage* (IL): $IU - RU = H(A) - H(A \mid O) = I(A; O)$

Nevertheless, recent work by Smith [24] suggests that these notions do not support security guarantees satisfactorily. In particular the remaining uncertainty is generally of little value in characterising the real threat that the adversary could guess the value of A given her low observations. Smith uses the following example to prove that.

Example 2. Consider the following programs `C3` and `C4`, where A is a uniformly distributed $8k$-bit integer, $k \geq 2$, & denotes bitwise 'AND', and $0^{7k-1}1^{k+1}$ a binary constant.

```
PROG C3:
  BEGIN
     If A mod 8 = 0
       Then O := A
       Else O := 1
  END
```

```
PROG C4:
  BEGIN
       O := A & 0^{7k-1}1^{k+1}
  END
```

`PROG C3` reveals completely the high input when A is a multiple of 8 while it reveals nothing about A otherwise (except of course the very fact that it is

not a multiple of 8). On the contrary, `PROG C4` reveals *always and only* the last $k+1$ bits of A.

According to Shannon entropy-based metrics, we have $IU = 8k$, $RU = 7k - 0.169$ and $IL = k + 0.169$ for `PROG C3`, and $IU = 8k$, $RU = 7k - 1$ and $IL = k + 1$ for `PROG C4` (the reader is referred to [24] for the detailed calculations). So, under such definitions, `PROG C4` appears actually *worse* than `PROG C3`, as $7k - 1 < 7k - 0.169$, even though intuitively `C3` leaves A highly vulnerable to being guessed (e.g., when it is a multiple of 8) while `C4` does not, at least for large k.

3.2. Belief approach

Clarkson *et al.* [21, 22] showed that the Shannon entropy approach is inadequate for measuring information flow when the adversary makes assumptions about the high-level secret and such assumptions might be incorrect. Based on the conviction that it is unavoidable that the attacker makes such (potentially inaccurate) assumptions, they proposed a new metric. They formalised the idea of an adversary's belief as a distribution of A assumed by the adversary: information flow is then expressed as an improvement in the *accuracy* of such belief. The *initial accuracy* is the Kullback-Leibler distance between the adversary's initial *belief* and the *actual* value of A; similarly the *remaining accuracy* is the Kullback-Leibler distance between the Bayesian-updated belief of the adversary after her low observation, and the actual value of A.

As already noticed by Smith [24], when the adversary's belief coincides with the a priori distribution of A, then the belief approach reduces again to the inadequate standard approach illustrated above.

3.3. Vulnerability approach

Having observed that both the consensus and the belief approaches fail in general to give good security guarantees, Smith [24] proposes a new metric for quantitative information flow based on the notions of *vulnerability* and *min-entropy*. We briefly revise these concepts here.

The vulnerability of a random variable A is the worse-case probability that an adversary could guess the value of A correctly in *one try*. The vulnerability of A, denoted $V(A)$, is thus formally defined as follows.

Definition 1. $V(A) = \max_{a \in \mathcal{A}} p(a)$.

The *conditional vulnerability* of a A given O measures the expected probability of guessing A in one try given O. It is denoted $V(A \mid O)$ and defined as follows.

Definition 2. $V(A \mid O) = \sum_{o \in \mathcal{O}} p(o) V(A \mid o)$, where $V(A \mid o)$ is $\max_{a \in \mathcal{A}} p(a \mid o)$.

The initial uncertainty about A is then defined as the negative logarithm of $V(A)$, which turns out to be the min-entropy of the random variable A – cf. (6) above. The remaining uncertainty about A after observing O is defined as the min-entropy of A given O. Thus we have the following vulnerability-based definitions:

- IU : $\quad H_\infty(A) = -\log V(A)$
- RU : $\quad H_\infty(A \mid O) = -\log V(A \mid O)$
- IL : $\quad IU - RU = H_\infty(A) - H_\infty(A \mid O)$

Security guarantees of the vulnerability-based approach: by applying these definitions to the programs of Example 2, we have $IU = 8k$, $RU = 8k - 3$ and $IL = 3$ for `PROG C3`, and $IU = 8k$, $RU = 7k - 1$ and $IL = k + 1$ for `PROG C4`. While these quantities remain the same as in the consensus approach for `PROG C4`, the new metric increases the leakage ascribed to `PROG C3` reflecting the fact that the low observations of `PROG C3` leave the high input very vulnerable to being guessed.

4. Unifying Belief and Vulnerability

We now propose an alternative approach based on the vulnerability concept that takes into account the adversary's belief.

4.1. Belief-vulnerability

Let A be a random high level variable. The *belief-vulnerability* of A is the expected probability of guessing A in one try given the adversary's belief. The adversary will choose a value having the maximal probability according to her belief, that is a value $a' \in \Gamma$, where $\Gamma = \operatorname{argmax}_{a \in \mathcal{A}} \; p_\beta(a)$. The vulnerability of A is then the *real* probability that the adversary's choice is correct, that is the probability $p(a')$. As there might be many values of A in Γ with the maximal probability, the attacker will pick uniformly at random one of them. Hence we have the following definition.

Definition 3. Let A be a random variable and $\Gamma = \text{argmax}_{a \in \mathcal{A}}\ p_\beta(a)$. The belief-vulnerability of A, denoted $V_\beta(A)$, is defined as

$$V_\beta(A) = \frac{1}{|\Gamma|} \sum_{a' \in \Gamma} p(a'). \tag{7}$$

The *initial threat* is the *belief-uncertainty* of A (viz. the *belief min-entropy*[4] of A) given by the following definition.

Definition 4. Let A be a random variable. The initial threat to A, denoted $H_\infty^\beta(A)$, is defined as

$$H_\infty^\beta(A) = \log\Big(\frac{1}{V_\beta(A)}\Big). \tag{8}$$

Note that from the above definition, the initial uncertainty may be infinite when for all a in Γ, $p(a) = 0$, modelling the impossibility of a correct guess in one single try when the attacker is choosing a value having zero prior probability. However it might be very difficult to quantify the decrease in uncertainty when initially it is infinite. Therefore we will assume throughout the rest of this paper that initially every value is possible, that is $p(x) > 0$ for all x in $\mathcal{X}$.

Example 3. Suppose that A is distributed over $\{0, 1, 2, 3\}$ and the adversary's belief is about the parity of A.

Table 2 summarizes the initial uncertainty about A when the actual a priori distribution and the adversary's prebelief are p_1 and p_2, and $p_{\beta 1}$ and $p_{\beta 2}$ respectively. The pair (p, p_β) means that the actual a priori distribution of A is p while the adversary prebelief is p_β. Thus in $(p_1, p_{\beta 1})$ the adversary believes that p_1 always uniformly produces an even number while, though it usually produces effectively an even number, with probability 0.02 it also produces an odd number that fools the attacker. However in a one-time guessing attack, this slightly "inaccurate" belief does not affect the vulnerability of A as knowing p_1, the actual a priori distribution, would not change the attacker choice. In $(p_2, p_{\beta 1})$, on the contrary, p_2 usually fools the attacker by producing an odd number while the attacker is always expecting an even

[4]We call it belief min-entropy since when the belief coincides with the actual a priori distribution, then the belief min-entropy reduces to min-entropy.

	$(p_1, p_{\beta 1})$	$(p_1, p_{\beta 2})$	$(p_2, p_{\beta 1})$	$(p_2, p_{\beta 2})$
$V_\beta(A)$	0.49	0.01	0.02	0.47
$H_\infty^\beta(A)$	1.03	6.64	5.64	1.09

$p_{\rho 1}$	p_1	p_2	$p_{\beta 1}$	$p_{\beta 2}$
a_0	0.49	0.03	0.5	0
a_1	0.01	0.47	0	0.95
a_2	0.49	0.01	0.5	0
a_3	0.01	0.49	0	0.05

Table 2: Initial uncertainty in presence of belief

number. This decreases very much the vulnerability of A as it is almost impossible for the attacker to guess the value of the secret in one try when her initial belief is so wrong. Hence, the increase in her initial uncertainty.

Next we show that the lower bound of the belief-uncertainty is obtained with a full knowledge of the actual a priori distribution (viz., a true and justifiable belief).

Theorem 1. $H_\infty(A) \leq H_\infty^\beta(A)$.

Proof. Let $\Gamma = \text{argmax}_{a \in \mathcal{A}}\ p_\beta(a)$ be the set of the adversary's possible choices.

$$V_\beta(A) = \frac{1}{|\Gamma|} \sum_{a \in \Gamma}\ p(a) \leq \frac{1}{|\Gamma|} \sum_{a \in \Gamma} \max_{a' \in \mathcal{A}} p(a') \quad (\text{ since } \Gamma \subseteq \mathcal{A})$$

Therefore, $V_\beta(A) \leq \frac{1}{|\Gamma|} \sum_{a \in \Gamma} V(A) = \frac{1}{|\Gamma|}\ |\Gamma| V(A) = V(A)$. Hence $H_\infty^\beta(A) \geq H_\infty(A)$. □

The lower bound is reached if and only if all the adversary's possible choices have the actual maximum a priori probability.

Proposition 1. $H_\infty^\beta(A) = H_\infty(A)$ *iff* $\Gamma \subseteq argmax_{a \in \mathcal{A}}\, p(a)$.

Proof. The proof in the right direction is similar to the proof of Theorem 1 where the inequality is replaced by the equality. To see why the opposite holds, it is sufficient to observe that when there exists an element of Γ that has not the maximum a priori probability then $V_\beta(A)$ is less than $V(A)$. Hence the uncertainty is greater than the lower-bound. □

In particular when the prebelief coincides with the actual a priori distribution so are the belief min-entropy and the min-entropy.

Corollary 1. *If $p(a) = p_\beta(a)$ for all a in $\mathcal{A}$, then $H^\beta_\infty(A) = H_\infty(A)$.*

In the vulnerability model where the adversary is supposed to know the a priori probability distribution of the high input, the minimum vulnerability, or equivalently the maximum uncertainty, is obtained using a uniform distribution. However in the presence of belief, the uncertainty can be arbitrary high. In fact we have the following upper-bound of the belief min-entropy.

Theorem 2. $H^\beta_\infty(A) \leq -\log\Big(\min_{a\in\mathcal{A}} p(a)\Big)$.

Proof. Similar to the proof of Theorem 1 with max replaced by min and the inequality is reversed. □

In particular the belief uncertainty hits the upper-bound when all the adversary's possible choices have the actual minimal prior probability.

Proposition 2. $H^\beta_\infty(A) = -\log\Big(\min_{a\in\mathcal{A}} p(a)\Big)$ *iff* $\Gamma \subseteq \mathit{argmin}_{a\in\mathcal{A}}\ p(a)$.

The proof in the right direction is again similar to the proof of Theorem 1 where the inequality is replaced by the equality and max by min. To see why the other implication holds, it is sufficient to observe that when there exists an element of Γ that has not the minimum a priori probability then $V_\beta(A)$ is greater than $\min_{a\in\mathcal{A}} p(a)$. Hence the uncertainty is less than the upper-bound.

We conclude this section by showing two interesting properties of the belief uncertainty. First, we show that beliefs have no effect on the initial uncertainty of A when its actual a priori probability distribution is uniform. In that case, any choice of the attacker results in the same actual vulnerability and she has the same probability $\frac{1}{|\mathcal{A}|}$ of a correct guess.

Proposition 3. *If A is uniformly distributed then for all initial belief p_β, we have*

$$H^\beta_\infty(A) = H_\infty(A) = \log\Big(|\mathcal{A}|\Big).$$

Proof. Let $\Gamma = \mathrm{argmax}_{a\in\mathcal{A}}\ p_\beta(a)$.

$$V_\beta(A) = \frac{1}{|\Gamma|}\sum_{a\in\Gamma} p(a) = \frac{1}{|\Gamma|}\sum_{a\in\Gamma}\frac{1}{|\mathcal{A}|} = \frac{1}{|\Gamma|}\sum_{a\in\Gamma} V(A) = \frac{1}{|\Gamma|}\ |\Gamma| V(A) = V(A).$$

Hence $H^\beta_\infty(A) = H_\infty(A) = \log\Big(|\mathcal{A}|\Big)$. □

Conversely, for a vacuous belief, the initial belief-uncertainty is independent of the actual a priori distribution of the high input.

Proposition 4. *If the prebelief is vacuous then*

$$H_\infty^\beta(A) = \log\Big(|\mathcal{A}|\Big).$$

Proof. If the prebelief is vacuous then its pignistic transformation p_β (Lemma 4) is the uniform probability distribution. Thus $\Gamma = \operatorname{argmax}_{a\in\mathcal{A}}\ p_\beta(a) = \mathcal{A}$. Therefore,

$$V_\beta(A) = \frac{1}{|\Gamma|}\sum_{a\in\Gamma} p(a) = \frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} p(a) = \frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} 1 = \frac{1}{|\mathcal{A}|}$$

Hence $H_\infty^\beta(A) = \log\Big(|\mathcal{A}|\Big)$. □

4.2. A posteriori belief-vulnerability

The belief-vulnerability of A given the evidence O is the expected probability of guessing A in one try given the evidence. Given evidence $O = o$, the adversary will choose a value having the maximal conditional probability according to her belief, that is a value $a' \in \Gamma_o$, where $\Gamma_o = \operatorname{argmax}_{a\in\mathcal{A}}\ p_\beta(a \mid o)$. But unlike the a priori belief-vulnerability, here the evidence O could contradict the attacker prebelief. For example assume that the adversary initially believes that the high input value is a_2, that is $p_\beta(a_2) = 1$ then she observes the output o_1 of `PROG C5` of which we only give the channel matrix Table 3. The evidence o_1 contradict her prebelief since only a_0 or a_3 could produce o_1. Hence her prebelief becomes vacuous and cannot be used as an a priori probability to update her belief.

$p(o \mid a)$	o_0	o_1	o_2
a_0	0.99	0.01	0
a_1	1	0	0
a_2	0.01	0	0.99
a_3	0	1	0

Table 3: Channel matrix of `PROG C5`

To avoid such completely wrong belief, we take the point of view that the adversary initial beliefs satisfy an *admissibility restriction* (see e.g. [22, 38]).

In particular, we consider an admissibility restriction we deem ϵ-*admissible* beliefs[5], where a belief never differs by more that a factor of ϵ from a uniform distribution, that is, $p_\beta(a) \geq \frac{\epsilon}{|\mathcal{A}|}$ for all a in $\mathcal{A}$. Note that the evidence o and the adversary prebelief could still be highly conflicting but never fully contradicting each other. Indeed, consider again the program `PROG C5` above. If we assume that the adversary prebelief is such that $p_\beta(a_0) = 0.90$ and $p_\beta(a_3) = 0.01$ then her belief highly supports a_0. Now observing o_1 would conflict with this belief as the evidence o_1 alone highly supports a_3 rather than a_0: a_0 would more likely have produced o_0. In this section, we shall assume that the adversary updates her belief using the Bayes' rule. Thus, the a posteriori belief is

$$\begin{aligned} p_\beta(a \,|\, o) &= \frac{p(o \,|\, a) p_\beta(a)}{p_\beta(o)} \\ &= \frac{p(o \,|\, a) p_\beta(a)}{\sum_{a' \in \mathcal{A}} p(o \,|\, a') p_\beta(a')} \end{aligned}$$

which is well defined thanks to the admissibility requirement. Appendix Appendix B introduces an advanced technique that takes into account the level of conflict to estimate the admissibility factor of an arbitrary belief given the evidence (i.e. the observables) induced by the channel.

The vulnerability of A given o is then the *real* probability that the adversary's choice (viz. a value $a' \in \Gamma_o = \mathrm{argmax}_{a \in \mathcal{A}}\ p_\beta(a \,|\, o)$ is the correct one, that is the objective conditional probability $p(a' \,|\, o)$. Again, as there might be many values of A with the maximal conditional probability, the attacker will pick uniformly at random one element in Γ_o. Hence we have the following definition.

Definition 5. The belief-vulnerability of A given O is defined as

$$V_\beta(A \,|\, O) = \sum_{o \in \mathcal{O}} p(o) V_\beta(A \,|\, o), \text{ where } V_\beta(A \,|\, o) = \frac{1}{|\Gamma_0|} \sum_{a \in \Gamma_o} p(a \,|\, o).$$

Next, we show how to compute $V_\beta(A \,|\, O)$ from the channel matrix $p(o \,|\, a)$

[5] Also considered in [21].

and the actual a priori probability $p(a)$.

$$\begin{aligned} V_\beta(A \,|\, O) &= \sum_{o \in \mathcal{O}} p(o) V_\beta(A \,|\, o) \\ &= \sum_{o \in \mathcal{O}} p(o) \Big(\frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) \Big) \\ &= \sum_{o \in \mathcal{O}} \frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) p(o). \end{aligned}$$

By Bayes theorem, we have $V_\beta(A \,|\, O) = \sum_{o \in \mathcal{O}} \frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(o \,|\, a) p(a)$. Therefore, the a posteriori belief-vulnerability can be easily computed as follows.

Proposition 5. *Let* $\Gamma_o = argmax_{a \in \mathcal{A}} \; p_\beta(a \,|\, o)$ *then*

$$V_\beta(A \,|\, O) = \sum_{o \in \mathcal{O}} \frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(o \,|\, a) p(a).$$

We then define the *remaining uncertainty* as the belief min-entropy of $A \,|\, O$.

Definition 6. Let A be a high level input of a channel and O its low output. The remaining threat to A after observing O, denoted $H^\beta_\infty(A \,|\, O)$, is defined as

$$H^\beta_\infty(A \,|\, O) = \log\Big(\frac{1}{V_\beta(A \,|\, O)}\Big).$$

Example 4. Again suppose that A is uniformly distributed over $\{0, 1, 2, 3\}$ and the adversary's initial belief is about the parity of A. Table 4 summarizes the remaining uncertainty about A after observing O, the output of the program `PROG C5` (see Table 3). Here, the actual a priori distributions are p_1 and p_2 as defined in Example 3. The prebeliefs $p_{\beta'1}$ and $p_{\beta'2}$ are slight modifications of $p_{\beta 1}$ and $p_{\beta 2}$ of Example 3 to cope with the admissibility requirement. For instance, consider the 0.04-admissible belief $p_{\beta'1}$. Then the adversary's a posteriori belief is as follows:

$p_{\beta'1}(a \,\vert\, o)$	a_0	a_1	a_2	a_3
o_0	0.9702	0.02	0.0098	0
o_1	0.3289	0	0	0.6711
o_2	0	0	1	0

	$(p_1, p_{\beta' 1})$	$(p_1, p_{\beta' 2})$	$(p_2, p_{\beta' 1})$	$(p_2, p_{\beta' 2})$
$V_\beta(A \mid O)$	0.9802	0.5051	0.5296	0.9699
$H^\beta_\infty(A \mid O)$	0.03	0.99	0.92	0.04

	p_1	p_2	$p_{\beta' 1}$	$p_{\beta' 2}$
a_0	0.49	0.03	0.49	0.01
a_1	0.01	0.47	0.01	0.97
a_2	0.49	0.01	0.49	0.01
a_3	0.01	0.49	0.01	0.01

Table 4: Remaining uncertainty of program `PROG C5`

Hence, $\Gamma_{o_0} = \{a_0\}$, $\Gamma_{o_1} = \{a_3\}$ and $\Gamma_{o_2} = \{a_2\}$. Therefore,

$$\begin{aligned} V_\beta(A \mid O) &= \sum_{o \in \mathcal{O}} \frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(o \mid a) p(a) \\ &= \sum_{j \in \{0,1,2\}} \frac{1}{|\Gamma_{o_j}|} \sum_{a \in \Gamma_{o_j}} p(o_j \mid a) p(a) \\ &= p(o_0 \mid a_0) p(a_0) + p(o_1 \mid a_3) p(a_3) + p(o_2 \mid a_2) p(a_2) \\ &= 0.99 \times p(a_0) + 1 \times p(a_3) + 0.99 \times p(a_2) \\ &= 0.99 \times \Big(p(a_0) + p(a_2) \Big) + p(a_3). \end{aligned}$$

In other words, if the actual a priori distribution does not support a_1 highly then program `PROG C5` will leave A highly vulnerable. This is the case for both p_1 and p_2 even though p_2 is highly conflicting with $p_{\beta' 1}$. The reason is that the a posteriori belief $p_{\beta' 1}(a \mid o)$ never supports a_1. So putting almost all the mass on a_1 will always fool the attacker. Similarly, for $p_{\beta' 2}$, we obtain

$$V_\beta(A \mid O) = p(a_1) + 0.99 \times p(a_2) + p(a_3).$$

In order to minimize the a posteriori belief-vulnerability of A, the actual a priori distribution p must put almost all the mass on a_0. Again, this is not the case for both p_1 and p_2 which explains why $V_\beta(A \mid O)$ is always greater that $\frac{1}{2}$.

Now we establish both lower and upper bounds for the remaining uncertainty in the presence of beliefs. We start with the lower bound and show

that, as in the case of the initial uncertainty, it is obtained with a full knowledge of the actual a priori distribution (viz., a true and justifiable belief).

Theorem 3. $H_\infty(A\,|\,O) \leq H^\beta_\infty(A\,|\,O)$.

Proof. Let $\Gamma_o = \mathrm{argmax}_{a\in\mathcal{A}}\ p_\beta(a\,|\,o)$.

$$\begin{aligned}
V_\beta(A\,|\,O) &= \sum_{o\in\mathcal{O}} p(o) V_\beta(A\,|\,o) \\
&= \sum_{o\in\mathcal{O}} p(o)\Big(\frac{1}{|\Gamma_o|}\sum_{a\in\Gamma_o}\ p(a\,|\,o)\Big) \\
&\leq \sum_{o\in\mathcal{O}} p(o)\Big(\frac{1}{|\Gamma_o|}\sum_{a\in\Gamma_o}\ \max_{a'\in\mathcal{A}} p(a'\,|\,o)\Big) \quad (\text{since } \Gamma_o \subseteq \mathcal{A}) \\
&\leq \sum_{o\in\mathcal{O}} p(o)\Big(\frac{1}{|\Gamma_o|}\sum_{a\in\Gamma_o}\ V(A\,|\,o)\Big) \\
&\leq \sum_{o\in\mathcal{O}} p(o)\Big(\frac{|\Gamma_o|}{|\Gamma_o|}\ V(A\,|\,o)\Big) \\
&\leq V(A\,|\,O).
\end{aligned}$$

Hence, $H_\infty(A\,|\,O) \leq H^\beta_\infty(A\,|\,O)$. □

Again, the lower bound is reached if and only if all the adversary's possible choices have the actual maximal a posteriori probability.

Proposition 6. $H^\beta_\infty(A\,|\,O) = H_\infty(A\,|\,O)$ *iff* $\forall o \in \mathcal{O}$, $\Gamma_o \subseteq \mathit{argmax}_{a\in\mathcal{A}}\, p(a\,|\,o)$.

Proof. The proof in the right direction is similar to the proof of Theorem 3 where the inequality is replaced by the equality since $\Gamma_o \subseteq \mathrm{argmax}_{a\in\mathcal{A}}\, p(a\,|\,o)$. To see why the opposite holds, it is sufficient observe that when there exists an element of Γ_o that has not the maximal a posteriori probability then $V_\beta(A\,|\,o) \leq V(A\,|\,o)$. Hence the uncertainty is greater than the lower-bound. □

In particular when the initial belief coincides with the actual a priori distribution then the belief-vulnerability remaining uncertainty reduces to the vulnerability remaining uncertainty.

Corollary 2. *If* $p(a) = p_\beta(a)$ *for all* a *in* $\mathcal{A}$, *then* $H^\beta_\infty(A\,|\,O) = H_\infty(A\,|\,O)$.

Now, onto the upper-bound.

Theorem 4. *Let A be a high input of a channel and O its low output. Then*

$$H_\infty^\beta(A \,|\, O) \leq \log\Big(\frac{1}{\zeta}\Big), \text{ where } \zeta = \min_{o \in \mathcal{O}}\big(\frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) \big).$$

Note that ζ in the above theorem is strictly greater than zero since we consider admissible beliefs for some positive ϵ. Hence the upper bound is well defined. yet, ζ could be very small, meaning that the presence of belief could add a huge amount of uncertainty and hence reduce a lot the vulnerability of A.

Proof. Let $\Gamma_o = \text{argmax}_{a \in \mathcal{A}}\ p_\beta(a \,|\, o)$.

$$\begin{aligned} V_\beta(A \,|\, O) &= \sum_{o \in \mathcal{O}} p(o) V_\beta(A \,|\, o) \\ &= \sum_{o \in \mathcal{O}} p(o) \Big(\frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) \Big) \\ &\geq \sum_{o \in \mathcal{O}} p(o) \min_{o \in \mathcal{O}} \Big(\frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) \Big) \\ &\geq \sum_{o \in \mathcal{O}} p(o) \zeta = \zeta. \end{aligned}$$

Hence, $H_\infty^\beta(A \,|\, O) \leq \log\Big(\frac{1}{\zeta}\Big)$. □

5. Belief-leakage

As usual, we define the information leakage of a channel as the reduction of uncertainty.

Definition 7. Let A be a high input of a channel C and O its low output. Given p, the actual prior distribution of A and p_β the belief of the adversary about A, the information leakage of C is

$$IL_\beta(p, C : p_\beta) = H_\infty^\beta(A) - H_\infty^\beta(A \,|\, O).$$

	$(p_1, p_{\beta' 1})$	$(p_1, p_{\beta' 2})$	$(p_2, p_{\beta' 1})$	$(p_2, p_{\beta' 2})$
$IL_\beta(\texttt{C5})$	1	5.65	4.72	1.05

Table 5: Information leakage of program `PROG C5`

Example 5. Let's consider again the parameters in Example 4. Since $p_{\beta' 1}$ (resp. $p_{\beta' 2}$), the slight modification of $p_{\beta 1}$ (resp. $p_{\beta 2}$), does not affect the adversary's choice, it induces the same initial uncertainty as in Example 3. From Tables 2 and 4 we obtain the information leakage of the program `PROG C5` as in Table 5.

We remark that when the initial belief and the actual a priori distribution are not highly conflicting (viz. $(p_1, p_{\beta' 1})$ and $(p_2, p_{\beta' 2})$) the leakage is about 1 bit of information. Moreover, the leakage is almost equal to the actual leakage of the program. However, even though the leakage is high (more than 4 bits of information), when they are highly conflicting, the actual leakage cannot exceed two bits of information. Hence much of the leakage is due to the correction of the misinformation induced by the wrong belief.

We note that while the remaining uncertainty $H^\beta_\infty(A \,|\, O)$ is a good characterization of the vulnerability of A after observing O, the information leakage alone, as usual, does not tell us much about the security of A. Moreover it might correct a lot the misinformation induced by a wrong belief about A. It is therefore important to characterize the the amount of uncertainty induced by a wrong belief.

5.1. Accuracy

From Theorems 1 and 3 we learned that wrong beliefs increase uncertainty. We introduce here a metric for the accuracy of a belief which quantifies the amount of uncertainty induced by the belief inaccuracy. Intuitively, the more the belief is inaccurate, the more the adversary's uncertainty should increase. We therefore quantify the inaccuracy of a belief by the divergence between the vulnerability induced by the belief and the actual vulnerability.

Definition 8. The belief divergence of p_β and p (aka the inaccuracy of p_β w.r.t. p) is $D_\beta(p \,\|\, p_\beta) = -log\Big(\frac{V_\beta(A)}{V(A)}\Big)$.

$D_\beta(p \,\|\, p_\beta)$ is always positive or null and accounts for the amount of increase in initial uncertainty due to the inaccuracy of p_β.

Lemma 1. $D_\beta(p \,\|\, p_\beta) \geq 0$

Proof. Follows directly from Theorem 1. □

Proposition 7. $H^\beta_\infty(A) = H_\infty(A) + D_\beta(p \,\|\, p_\beta)$

Proof. Follows directly from their respective definitions (4 and 8). □

Similarly we define the posterior inaccuracy as the divergence between the corresponding posterior vulnerabilities.

Definition 9. The posterior inaccuracy of p_β w.r.t. p is $D_\beta(p \,\|\, p_\beta : C) = -log\Big(\frac{V_\beta(A \,|\, O)}{V(A \,|\, O)}\Big)$.

Again, it is always positive or null (cf. Theorem 3) and accounts for the amount of increase in the remaining uncertainty due to the inaccuracy of p_β.

Lemma 2. $D_\beta(p \,\|\, p_\beta : C) \geq 0$

Proposition 8. $H^\beta_\infty(A \,|\, O) = H_\infty(A \,|\, O) + D_\beta(p \,\|\, p_\beta : C)$

5.2. Leakage

In the previous section we have seen that a wrong belief increases the prior and the posterior min-entropy by the terms $D_\beta(p \,\|\, p_\beta)$ and $D_\beta(p \,\|\, p_\beta : C)$, respectively. Note that $D_\beta(p \,\|\, p_\beta)$ and $D_\beta(p \,\|\, p_\beta : C)$ can be arbitrary high. But: how do belief leakage and min-entropy leakage compare? We would expect the min-entropy leakage never to exceed the belief leakage, since the channel may correct the initial wrong belief and hence leaks more information. Actually this is not always the case. Indeed, as shown by our analysis of program `PROG C2` (see Example 1) in Section 6, the belief leakage can be less, equal or greater than the min-entropy leakage. The reason is that the channel is not always reducing the inaccuracy of the belief. In fact a channel may increase the adversary's confidence over some hypotheses which are actually very misleading. Thus the change in inaccuracy can be positive, null or even negative.

From Propositions 7 and 8 we have that a wrong belief increases (resp. decreases) the leakage by a term $\Delta D_\beta(p \,\|\, p_\beta : C) = D_\beta(p \,\|\, p_\beta) - D_\beta(p \,\|\, p_\beta : C)$ equal to the decrease (resp. increase) in inaccuracy. Hence, the belief leakage $IL_\beta(p, C : p_\beta)$ is related to the min-leakage $IL_\infty(p, C)$ as follows.

Proposition 9. $IL_\beta(p, C : p_\beta) = IL_\infty(p, C) + \Delta D_\beta(p \,\|\, p_\beta : C)$

Interestingly, as in the case of the min-entropy leakage and the g-leakage, the belief leakage is also null when the low output and the high input are independent.

Theorem 5 (Zero-leakage)**.** $H^\beta_\infty(A) - H^\beta_\infty(A \,|\, O) = 0$ *if* A *and* O *are independent.*

Proof. Assume that A and O are independent. Then from Equation 9 we have that for all observable o, the posterior belief $p_\beta(a \,|\, o)$ is equal to the prior belief $p_\beta(a)$. Hence, $\Gamma_o = \Gamma$ for all observable o. In other words, the observables do not affect the adversary's choice. Therefore,

$$\begin{aligned} V_\beta(A \,|\, O) &= \sum_{o \in \mathcal{O}} p(o) V_\beta(A \,|\, o) = \sum_{o \in \mathcal{O}} p(o) \Big(\frac{1}{|\Gamma_o|} \sum_{a \in \Gamma_o} p(a \,|\, o) \Big) \\ &= \sum_{o \in \mathcal{O}} p(o) \Big(\frac{1}{|\Gamma|} \sum_{a \in \Gamma} p(a \,|\, o) \Big) \quad (\text{ Independence of } A \text{ and } O) \\ &= \sum_{o \in \mathcal{O}} p(o) \Big(\frac{1}{|\Gamma|} \sum_{a \in \Gamma} p(a) \Big) = \sum_{o \in \mathcal{O}} p(o) V_\beta(A) = V_\beta(A). \end{aligned}$$

Hence $H^\beta_\infty(A \,|\, O) = H^\beta_\infty(A)$. □

Finally, we may wonder whether the belief leakage of a channel C can ever be negative, as C may increase the inaccuracy of a belief. However, as in the case of the min-entropy leakage and the g-leakage [25], the belief leakage is always positive or null as it is defined in terms of expectation.

Theorem 6. $H^\beta_\infty(A) - H^\beta_\infty(A \,|\, O) \geq 0$.

6. On the Applicability of the Belief-vulnerability Approach

The previous section establishes our definitions in terms of their theoretical properties. Now we show the utility of our approach by applying it to various threat scenarios and comparing the results to the previous approaches. This is done through an extensive analysis of the deterministic program `PROG C1` and the probabilistic program `PROG C2` of Example 1. In order to simplify the reading, we reproduce them here.

```
PROG C2:
  BEGIN
     R 'sampled from {0,2} with p(0) = λ and p(2) = 1 − λ';
     If A = R
       Then O := A
       Else O := −1
  END
```

```
PROG C1:
  BEGIN
    O := ⌊ log(A + 2) ⌋
  END
```

Each of these programs is analysed under the following hypothesis.

- The actual prior distribution of the high input A is such that

$$p(a_0) = p(a_2) = \frac{\omega}{2(1+\omega)} \text{ and } p(a_1) = p(a_3) = \frac{1}{2(1+\omega)},$$

  for some $0 < \omega \leq 1$.

- The adversary believes that A is most likely an even number, that is she assumes the following $\frac{1}{2}$-admissible belief:

$$p_\beta(a_0) = p_\beta(a_2) = \frac{3}{8} \text{ and } p_\beta(a_1) = p_\beta(a_3) = \frac{1}{8}.$$

  Thus $\Gamma = \{a_0, a_2\}$ and hence the probability that a guess of the attacker is correct is reduced by the factor ω compared to someone who knows the actual distribution.

We denote by IU_x the initial uncertainty computed using the approach $x \in \{s, \infty, \beta\}$ where s, ∞ and β denote the Shannon entropy, Rényi min-entropy and belief-based approaches respectively. Ditto for RU_x and IL_x. We start with `PROG C1`.

6.1. Analysis of `PROG C1`

Shannon entropy leakage. For the prior entropy we have:

$$\begin{aligned} IU_s = H(A) &= -\sum_{a\in\mathcal{A}} p(a)\log p(a) \\ &= -\Big(\frac{\omega}{2(1+\omega)}\log\frac{\omega}{2(1+\omega)} + \frac{1}{2(1+\omega)}\log\frac{1}{2(1+\omega)} \\ &\quad + \frac{\omega}{2(1+\omega)}\log\frac{\omega}{2(1+\omega)} + \frac{1}{2(1+\omega)}\log\frac{1}{2(1+\omega)}\Big) \\ &= -\Big(\frac{\omega}{1+\omega}\log\frac{\omega}{2(1+\omega)} + \frac{1}{1+\omega}\log\frac{1}{2(1+\omega)}\Big) \\ &= 1+\log(1+\omega) - \frac{\omega}{1+\omega}\log\omega. \end{aligned}$$

For the posterior entropy, we have $p(o_0) = p(o_1) = \frac{1}{2}$ since we have $p(o) = \sum_{a\in\mathcal{A}} p(o \mid a)p(a)$. The conditional probabilities $p(a \mid o) = \frac{p(o \mid a)p(a)}{p(o)}$ are given in Table 6. Thus,

$$\begin{aligned} RU_s = H(A \mid O) &= -\sum_{o\in\mathcal{O}} p(o)\sum_{a\in\mathcal{A}} p(a \mid o)\log p(a \mid o) \\ &= -\Big[\frac{1}{2}\Big(\frac{\omega}{1+\omega}\log\frac{\omega}{1+\omega} + \frac{1}{1+\omega}\log\frac{1}{1+\omega}\Big) \\ &\quad + \frac{1}{2}\Big(\frac{\omega}{1+\omega}\log\frac{\omega}{1+\omega} + \frac{1}{1+\omega}\log\frac{1}{1+\omega}\Big)\Big] \\ &= -\Big(\frac{\omega}{1+\omega}\log\frac{\omega}{1+\omega} + \frac{1}{1+\omega}\log\frac{1}{1+\omega}\Big) \\ &= \log(1+\omega) - \frac{\omega}{1+\omega}\log\omega. \end{aligned}$$

Hence the leakage is $IL_s = IU_s - RU_s = 1$. It is illustrated in Figure 1.

Min-entropy leakage. For the prior min-entropy, we have

$$\begin{aligned} IU_\infty = H_\infty(A) &= -\log V(A) = -\log(\max_{a\in\mathcal{A}} p(a)) = -\log\frac{1}{2(1+\omega)} \\ &= 1+\log(1+\omega). \end{aligned}$$

Similarly, for the posterior min-entropy, we have

$$\begin{aligned} RU_\infty = H_\infty(A \mid O) &= -\log V(A \mid O) = -\log\Big(\sum_{o\in\mathcal{O}} p(o)\max_{a\in\mathcal{A}} p(a \mid o)\Big) \\ &= -\log(\frac{1}{2}\cdot\frac{1}{1+\omega} + \frac{1}{2}\cdot\frac{1}{1+\omega}) = \log(1+\omega). \end{aligned}$$

$p(o \mid a)$	0_0	0_1
a_0	1	0
a_1	1	0
a_2	0	1
a_3	0	1

$p(a \mid o)$	a_0	a_1	a_2	a_3
o_0	$\frac{\omega}{1+\omega}$	$\frac{1}{1+\omega}$	0	0
o_1	0	0	$\frac{\omega}{1+\omega}$	$\frac{1}{1+\omega}$

$p_\beta(a \mid o)$	a_0	a_1	a_2	a_3
o_0	$\frac{3}{4}$	$\frac{1}{4}$	0	0
o_1	0	0	$\frac{3}{4}$	$\frac{1}{4}$

x	IU_x	RU_x	$IL_x(C1)$
s	$1+\log(\omega+1)-\frac{\omega}{\omega+1}\log\omega$	$\log(\omega+1)-\frac{\omega}{\omega+1}\log\omega$	1
∞	$1+\log(\omega+1)$	$\log(\omega+1)$	1
β	$1+\log(\omega+1)-\log\omega$	$\log(\omega+1)-\log\omega$	1

Table 6: Matrices of conditional probabilities and the leakages of `PROG C1`

Again, the leakage is 1 since $IL_\infty = IU_\infty - RU_\infty = 1$. It is illustrated in Figure 1.

Belief leakage. Since the adversary initially believes that A is most likely even, then $\Gamma = \{a_0, a_2\}$. Hence the belief vulnerability of A is

$$V_\beta(A) \;=\; \frac{1}{\Gamma}\sum_{a\in\Gamma} p(a) = \frac{1}{2}\sum_{a\in\{a_0,a_2\}} p(a)) = \frac{\omega}{2(1+\omega)}.$$

Thus $IU_\beta = H^\beta_\infty(A) = -\log V_\beta(A) = 1+\log(1+\omega)-\log(\omega)$.

For the posterior belief vulnerability, from Bayes' rule, we obtain the adversary's updated belief $p_\beta(a \mid o)$ as in Table 6, whence we obtain the following sets of adversary's possible choices $\Gamma_{o_0} = \{a_0\}$ and $\Gamma_{o_1} = \{a_2\}$. Therefore,

$$\begin{aligned} V_\beta(A \mid O) \;&=\; \sum_{o\in\mathcal{O}} \frac{1}{|\Gamma_o|}\sum_{a\in\Gamma_o} p(o \mid a)p(a) = \sum_{j\in\{0,1\}} \frac{1}{|\Gamma_{o_j}|}\sum_{a\in\Gamma_{o_j}} p(o_j \mid a)p(a) \\ &=\; p(o_0 \mid a_0)p(a_0) + p(o_1 \mid a_2)p(a_2) = \frac{\omega}{1+\omega}. \end{aligned}$$

Hence $RU_\beta = H^\beta_\infty(A \mid O) = -\log V_\beta(A \mid O) = \log(1+\omega)-\log(\omega)$, which results in a leakage of 1 again. Thus, the fact that the adversary initially

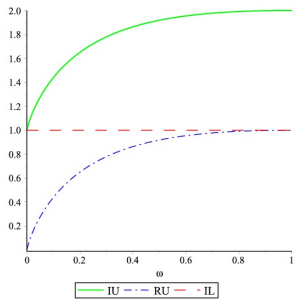

Figure 1: Shannon leakage of `C1`

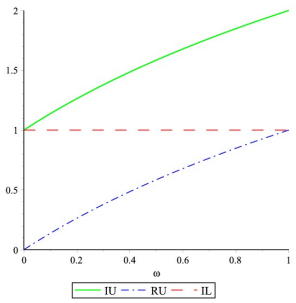

Figure 2: Min-leakage of `C1`

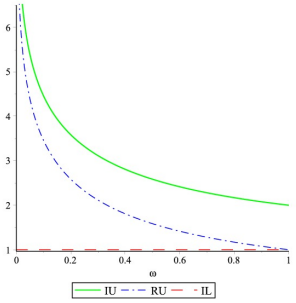

Figure 3: Belief leakage of `C1`

believes that A is even does not affect the quantity of information leaked by `PROG C1`. However, the real question is not how much information is leaked by this program, but what the remaining uncertainty represents in terms of security threat to the high input. Even though the adversary's belief does not affect the quantity of the information leakage, it dramatically affects both the initial and remaining uncertainty. Indeed, as illustrated by Figure 3, both IU_β and RU_β tend toward infinity as ω tends toward zero. On the other hand, IU_β and RU_β tend toward two and one, respectively, as ω tends toward one. In other words inaccurate beliefs strengthen the security of the program (by confusing the adversary). Thus, a deliberate leakage of a wrong side information biased toward the less likely parity of the high input is a good strategy to strengthen the security of this program.

6.2. Analysis of `PROG C2`

We continue our analysis with the probabilistic program `PROG C2`. The initial uncertainty remains the same as in the analysis of the program `PROG C1`. Following the same approach as for `PROG C1`, we obtain the matrices of conditional probabilities and the leakages for `PROG C2` shown in Table 7.

Proposition 10 (Shannon entropy leakage of `PROG C2`)**.**

$$\begin{aligned} IL_s(C1) &= 1 + \log(\omega+1) - \frac{\omega}{2(\omega+1)}\log\omega - \frac{\omega+2}{2(\omega+1)}\log(\omega+2) \\ &\quad + \frac{\omega}{2(\omega+1)}\Big(\lambda\log\lambda + (1-\lambda)\log(1-\lambda)\Big). \end{aligned}$$

$p(o \mid a)$	0_0	0_1	0_2
a_0	$1-\lambda$	λ	0
a_1	1	0	0
a_2	λ	0	$1-\lambda$
a_3	1	0	0

$p(a \mid o)$	a_0	a_1	a_2	a_3
o_0	$\frac{(1-\lambda)\omega}{2+\omega}$	$\frac{1}{2+\omega}$	$\frac{\lambda\omega}{2+\omega}$	$\frac{1}{2+\omega}$
o_1	1	0	0	0
o_2	0	0	1	0

$p_\beta(a \mid o)$	a_0	a_1	a_2	a_3
o_0	$\frac{3}{5}(1-\lambda)$	$\frac{1}{5}$	$\frac{3}{5}\lambda$	$\frac{1}{5}$
o_1	1	0	0	0
o_2	0	0	1	0

x	RU_x	$IL_x(C2)$
s	$\frac{\omega+2}{2(\omega+1)}\log(\omega+2) - \frac{\omega}{2(\omega+1)}\Big(\log\omega + \lambda\log\lambda + (1-\lambda)\log(1-\lambda)\Big)$	see Propostion 10
∞	1	$\log(\omega+1)$
β	$1 + \log(\omega+1) - \log\omega - \log(1+\max(\lambda, 1-\lambda))$	$\log(1+\max(\lambda, 1-\lambda))$

Table 7: Matrices of conditional probabilities and the leakages of `PROG C2`

The information flow ascribed by the Shannon entropy is illustrated in Figures 4 and 5, those of the belief-based approach in Figures 6 and 7. We note that the randomisation parameter λ of `PROG C2` has no effect on the min-entropy leakage. This is due to the fact that this metric focuses on the single probability that poses the greatest risk; the actual posterior probabilities of a_1 and a_3, the possible choices of the adversary, do not depend on λ. On the other hand, the accuracy factor ω of the adversary's prior belief has no effect on the belief leakage. This is due to the fact that she always chooses the even number a_0 or a_2 towards which λ is biased independently of ω. Since a_0 and a_2 play symmetric roles with respect to λ, any gain in uncertainty due to the value towards which λ is biased is offset by the loss due to the other value.

Finally, comparing the min-entropy leakage and the belief leakage of `PROG C2` (see Figure 8), we can assert that belief leakage is higher than min-entropy leakage except for highly accurate beliefs. In fact, we have the following result relating min-entropy and belief leakages depending on the randomisation parameter λ and ω, the accuracy factor of the adversary's prior belief.

Proposition 11. *The belief leakage of* `PROG C2` *is higher than its min-entropy leakage if and only if the randomisation parameter λ and ω, the accuracy factor of the adversary's prior belief, satisfy the following relation:*

$$\omega \leq \max(\lambda, 1-\lambda).$$

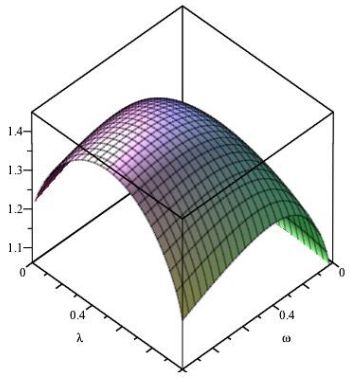

Figure 4: RU_s of PROG C2

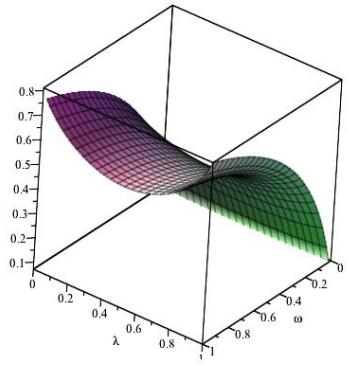

Figure 5: IL_s of PROG C2

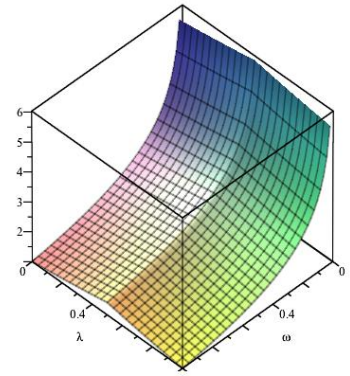

Figure 6: RU_β of PROG C2

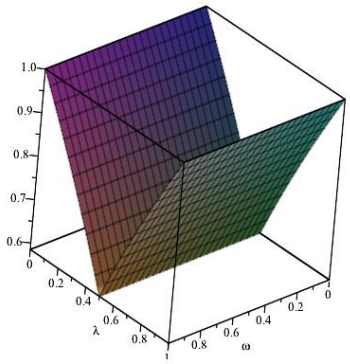

Figure 7: IL_β of PROG C2

6.3. Belief accuracy vs adversary's confidence

Our analysis so far confirmed that inaccurate beliefs tend to confuse the adversary by reducing the probability that she correctly guesses the hidden or protected data. However, it might be the case that the interaction with the security mechanism (wrongly) increased the adversary's confidence about the value of the hidden or protected data, that is, she wrongly believes that she has learned some useful information. Hence, she may end up being highly confident about her (wrong) result and may take action that results in serious harm to an innocent victim or to the whole society.

For instance, assume that $\mathcal{A}$ is the set of hidden identities of the users of the security system. Consider a 'dictator' adversary who takes action against a user a only when she believes that a is more likely to be the *culprit* (i.e.

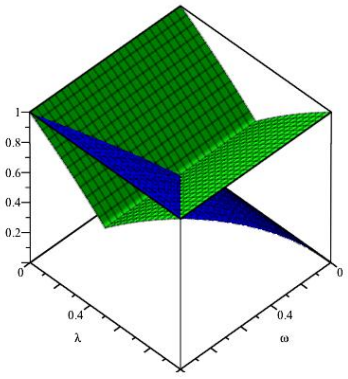

Figure 8: Belief leakage vs min-entropy leakage (PROG C2)

more likely to be the hidden input) than to not be the culprit. In other words, she only takes action against a when $p_\beta(a) > \frac{1}{2}$ a priori or $p_\beta(a \mid o) > \frac{1}{2}$ a posteriori for a given observable o.

Now let us see how our simple programs C1 and C2 perform against such adversary. First, we note that, a priori, there is no victim since the adversary's prior belief p_β does not assign a mass strictly higher than one-half to any value of A. However, after observing the outcome of C1 (see Table 6) we have that $\tilde{p}_\beta(a_0 \mid o_0) = \tilde{p}_\beta(a_2 \mid o_1) = \frac{3}{4}$. Therefore, the adversary will take action against a_0 (resp. a_2) when she observes o_0 (resp. o_1) even though a_0 (resp. a_2) is highly unlikely to be the input when ω, the accuracy factor of the adversary's prior belief, is very small. Let $Prob[Vic : p_\beta, p, C]$ (resp. $Prob[Inn : p_\beta, p, C]$) denote the probability that the adversary takes action against a user (resp. an innocent user) after observing the outcome of the channel C when her prior belief is p_β and the actual prior probability is p. Let p_β and p be as in the previous section. Then the followings hold.

Proposition 12. *Programs* C1 *and* C2 *perform against a dictator adversary*

as follows.

$$
\begin{aligned}
Prob[\,Vic : p_\beta, p, C1] &= 1 \\
Prob[Inn : p_\beta, p, C1] &= \frac{1}{1+\omega} \\
Prob[\,Vic : p_\beta, p, C2] &= \begin{cases} \frac{\omega}{2(1+\omega)} & \textit{if } \frac{1}{6} \leq \lambda \leq \frac{5}{6} \\ 1 & \textit{otherwise} \end{cases} \\
Prob[Inn : p_\beta, p, C2] &= \begin{cases} 0 & \textit{if } \frac{1}{6} \leq \lambda \leq \frac{5}{6} \\ \frac{2+\omega\times\min(\lambda,1-\lambda)}{2(1+\omega)} & \textit{otherwise} \end{cases}
\end{aligned}
$$

Proof. The probability of having a victim can be expressed as the sum of the probability $Prob[p_\beta(a \,|\, o) > \frac{1}{2}]$ that user a is a victim given the observable o weighted by the actual probability $p(o)$ of the observable. Hence

$$Prob[\,Vic : p_\beta, p, C] = \sum_{o\in\mathcal{O}} \sum_{a\in\mathcal{A}} p(o) \times Prob[p_\beta(a \,|\, o) > \frac{1}{2}].$$

Similarly, the probability of having an innocent victim can be expressed as the sum of the probability of having a victim a for each observable o weighted by the actual probability $\sum_{a'\neq a} p(a' \,|\, o)$ that somebody else could have been the actual input. Thus

$$Prob[\,Vic : p_\beta, p, C] = \sum_{o\in\mathcal{O}} \sum_{a\in\mathcal{A}} p(o) \times Prob[p_\beta(a \,|\, o) > \frac{1}{2}]\Big(\sum_{a'\neq a} p(a' \,|\, o)\Big).$$

Since $p(o) = \sum_{a\in\mathcal{A}} p(o \,|\, a)p(a)$, the proposition follows from Table 6 and Table 7. □

The results above show that in terms of protecting everybody (i.e. not having victims), `C1` ensures zero level of protection, since whatever the outcome of the program is the adversary will become confident enough about her guess to take action. Ditto for `C2` when the randomization factor λ is highly biased towards one value of A. However, when λ is not highly biased, then `C2` becomes extremely good, since the most likely outcome o_0 does not increase the adversary's confidence enough to trigger an action, specially when ω, the accuracy factor of her prior belief, is very small. Hence in terms of protecting everybody against such 'dictator' adversary, `C2` performs better than `C1` when λ is not highly biased. But they are equally bad when λ is highly biased.

In terms of protecting only the innocent victims, `C2` is perfect when λ is not highly biased, since in this case the adversary will never take action against an innocent victim. However when λ is highly biased, `C2` becomes slightly worse that `C1`. The reason is that in this case even when the input is an even number there is a small probability that the adversary's guess is wrong for `C2` when she observes o_0. This is never the case for `C1`.

The elementary examples in this paper illustrate the applicability of our metric to various threat scenarios. In particular, they show that confidence, i.e., the entropy of the adversary's belief, and accuracy are two orthogonal dimensions of security. We need to consider both of them in order to be able to model a wide range of threat scenarios. Typically, the adversary makes her decision according to her confidence in her beliefs, i.e., based on the uncertainty metric. On the other hand, the consequence of her decision, viz., the gain, usually depends on the accuracy metric. The 'dictator' example illustrates the limit case where the gain is independent of the accuracy (as the dictator always wins). It also shows the importance of the reliability of side information.

7. Related work

We have already mentioned in Sections 1 and 3 much of the related work. In this section we only discuss the work that is closely related to our framework and that was not already reported there.

As far as we know, [21, 22] have been the first papers to address the adversary's beliefs in quantifying information flow. This line of work, which inspired our formulation of the belief-vulnerability, is based on the Kullback-Leibler distance, a concept related to Shannon entropy which in general, as already mentioned in Section 3, fails to characterize many realistic attacks scenarios. Belief semantics has been explored also in the context of other security properties in [39, 40]. In a recent paper [41], Hussein proposed a generalization of the Bayesian-based framework of [22] based on Dempster-Shafer theory. The accuracy is expressed in terms of the generalized Jensen-Shannon divergence [42] between the belief and the actual value of the high input. Our metric is, on the contrary, based on the *min-entropy* leakage, where a leakage of k-bits means that on average the channel increases the *vulnerability* of the secret to being guessed correctly in *one single try* by the factor 2^k. Moreover, as shown by recent work [25], quantitative information flow metrics based on the min-entropy model can be extended by so-called

gain functions to account for a wider range of operational attack scenarios. While the framework in [25] is closely related to ours, as they both generalize the quantitative information flow model based on Rényi min-entropy, their objectives are orthogonal. Belief functions model the adversary's (potentially incorrect) side information; gain-functions model the operational threat scenarios, such as the adversary's benefit from guessing a value 'close' to the actual value of the secret, guessing a property of the secret, or guessing its value within some number of tries.

A related line of research has explored methods of statistical inference, in particular those from the *hypothesis testing* framework. The idea is that the adversary's best guess is that the true input is the one which has the maximum conditional probability (MAP rule) and that, therefore, the *a posteriori vulnerability* of the system is the complement of the *Bayes Risk*, which is the average probability of making the wrong guess when using the MAP rule [43]. This is always at least as high as the *a priori vulnerability*, which is the probability of making the right guess just based on the knowledge of the input distribution. It turns out that Smith's notion of leakage actually corresponds to the ratio between the a posteriori and the a priori vulnerabilities [24, 44].

Concerning the computation of the channel matrix and of the leakage for probabilistic systems, one of the first works to attack the problem was [45], in which the authors proposed various model-checking techniques. One of these is able to generate counterexamples, namely points in the execution where the channel exhibits an excessive amount of leakage. This method is therefore also useful to fix unsound protocols. A subsequent paper [46] focussed more on the efficiency and proposed the use of binary decision diagrams and symbolic model checking. For systems that are too large to apply model-checking, approximate techniques based on statistics have been proposed [47].

8. Conclusion and future work

This paper presents a new approach to quantitative information flow that incorporates the attacker's beliefs in the model on Rényi min entropy. We investigate the impact of such adversary's inaccurate, misleading or outdated side information on the security of the secret information. Our analysis reveals that inaccurate side information tends to confuse the adversary by decreasing the probability of a correct guess. However, due to the inaccuracy

of her side information, the interaction with the security mechanism may (wrongly) strengthen the adversary's confidence about her choice. While the attacker might not actually discover the true value of the secret, the fact that she wrongly believes to actually know it constitutes a security threat, since it can adversely influence how she views or interacts with the victim. We have shown the strength of our definitions both in terms of their theoretical properties and of their utility by applying them to various threat scenarios. As already stated in [21, 22], our results confirm, in particular, that the confidence of the attacker and the accuracy of her belief are two orthogonal dimensions of the security. Depending on the threat operational scenario, both could be equally important or one more relevant than the other. We observe that our approach can also be seen as moving away from the often criticised yet standard assumption in information flow that the adversary knows the true distribution of *secrets*. More realistically, here the focus is on the attacker's assumptions about such distributions (their a priori beliefs), and the consequences of making them.

As future work, we shall extend our model to consider a wide range of operational threat scenarios especially in the presence of wrong beliefs. To achieve this goal, we plan to build upon our belief framework and the gain-functions framework [25]. Secondly, we shall devise methods and techniques of collecting and aggregating the adversary's side information from different and potentially conflicting sources such as social networks, online forums and blogs, etc. We believe that the body of concepts and results related to the notion of belief combination will provide natural and solid tools for this goal. Finally, we also plan to implement our framework in a tool to assess the effectiveness of current security/privacy mechanisms in a real-world context and under various attack scenarios. Doing so, will help us to have a clear view of the real impact of side information collected from social networks, online forums, blogs and other forms of online communication and information sharing. It will also help us to accurately estimate the reliability of each of these sources, as well as providing guidelines and some mechanism design principles for privacy-preserving and security mechanisms.

[1] D. Chaum, The dining cryptographers problem: Unconditional sender and recipient untraceability, Journal of Cryptology 1 (1988) 65–75.

[2] M. K. Reiter, A. D. Rubin, Crowds: anonymity for Web transactions, ACM Transactions on Information and System Security 1 (1) (1998) 66–92.

[3] P. Syverson, D. Goldschlag, M. Reed, Anonymous connections and onion routing, in: IEEE Symposium on Security and Privacy, Oakland, California, 1997, pp. 44–54.

[4] I. Clarke, O. Sandberg, B. Wiley, T. W. Hong, Freenet: A distributed anonymous information storage and retrieval system., in: Designing Privacy Enhancing Technologies, International Workshop on Design Issues in Anonymity and Unobservability, Vol. 2009 of Lecture Notes in Computer Science, Springer, 2000, pp. 44–66.

[5] P. F. Syverson, S. G. Stubblebine, Group principals and the formalization of anonymity, in: World Congress on Formal Methods (1), 1999, pp. 814–833.

[6] J. Y. Halpern, K. R. O'Neill, Anonymity and information hiding in multiagent systems, Journal of Computer Security 13 (3) (2005) 483–512.

[7] D. Hughes, V. Shmatikov, Information hiding, anonymity and privacy: a modular approach, Journal of Computer Security 12 (1) (2004) 3–36.

[8] S. Schneider, A. Sidiropoulos, CSP and anonymity, in: Proc. of the European Symposium on Research in Computer Security (ESORICS), Vol. 1146 of LNCS, Springer, 1996, pp. 198–218.

[9] P. Y. Ryan, S. Schneider, Modelling and Analysis of Security Protocols, Addison-Wesley, 2001.

[10] M. Bhargava, C. Palamidessi, Probabilistic anonymity, in: M. Abadi, L. de Alfaro (Eds.), Proceedings of CONCUR, Vol. 3653 of Lecture Notes in Computer Science, Springer, 2005, pp. 171–185.

[11] K. Chatzikokolakis, C. Palamidessi, Probable innocence revisited, Theoretical Computer Science 367 (1-2) (2006) 123–138, http://www.lix.polytechnique.fr/~catuscia/papers/Anonymity/tcsPI.pdf.
URL http://hal.inria.fr/inria-00201072/en/

[12] T. M. Cover, J. A. Thomas, Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing), Wiley-Interscience, 2006.

[13] Y. Zhu, R. Bettati, Anonymity vs. information leakage in anonymity systems, in: Proc. of ICDCS, IEEE Computer Society, 2005, pp. 514–524.

[14] D. Clark, S. Hunt, 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 (2) (2005) 181–199.

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

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

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

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

[19] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, Anonymity protocols as noisy channels, Information and Computation 206 (2–4) (2008) 378–401. doi:10.1016/j.ic.2007.07.003.
URL http://hal.inria.fr/inria-00349225/en/

[20] M. Franz, B. Meyer, A. Pashalidis, Attacking unlinkability: The importance of context, in: N. Borisov, P. Golle (Eds.), Privacy Enhancing Technologies, 7th International Symposium, PET 2007 Ottawa, Canada, June 20-22, 2007, Revised Selected Papers, Vol. 4776 of Lecture Notes in Computer Science, Springer, 2007, pp. 1–16.
URL http://dx.doi.org/10.1007/978-3-540-75551-7_1

[21] M. R. Clarkson, A. C. Myers, F. B. Schneider, Belief in information flow, in: CSFW, IEEE Computer Society, 2005, pp. 31–45.

[22] M. R. Clarkson, A. C. Myers, F. B. Schneider, Quantifying information flow with beliefs, Journal of Computer Security 17 (5) (2009) 655–701.

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

[24] G. Smith, On the foundations of quantitative information flow, in: L. De Alfaro (Ed.), Proceedings of the Twelfth International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2009), Vol. 5504 of Lecture Notes in Computer Science, Springer, York, UK, 2009, pp. 288–302.

[25] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, G. Smith, Measuring information leakage using generalized gain functions, in: Proceedings of the 25th IEEE Computer Security Foundations Symposium (CSF), 2012, pp. 265–279. doi:http://doi.ieeecomputersociety.org/10.1109/CSF.2012.26.

[26] A. McIver, C. Morgan, G. Smith, B. Espinoza, L. Meinicke, Abstract channels and their robust information-leakage ordering, in: M. Abadi, S. Kremer (Eds.), Proceedings of the Third International Conference on Principles of Security and Trust (POST), Vol. 8414 of Lecture Notes in Computer Science, Springer, 2014, pp. 83–102.

[27] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, G. Smith, Additive and multiplicative notions of leakage, and their capacities, in: IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014, IEEE, 2014, pp. 308–322. doi:10.1109/CSF.2014.29.

[28] S. Hamadou, C. Palamidessi, V. Sassone, E. ElSalamouny, Probable innocence in the presence of independent knowledge, in: P. Degano, J. D. Guttman (Eds.), Formal Aspects in Security and Trust, FAST

2009, Vol. 5983 of Lecture Notes in Computer Science, Springer, 2009, pp. 141–156.

[29] S. Hamadou, V. Sassone, C. Palamidessi, Reconciling belief and vulnerability in information flow, in: IEEE Symposium on Security and Privacy, IEEE Computer Society, 2010, pp. 79–92.

[30] A. P. Dempster, A generalization of bayesian inference, in: R. R. Yager, L. Liu (Eds.), Classic Works of the Dempster-Shafer Theory of Belief Functions, Vol. 219 of Studies in Fuzziness and Soft Computing, Springer, 2008, pp. 73–104.

[31] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.

[32] J. L. Massey, Guessing and entropy, in: In Proceedings of the 1994 IEEE International Symposium on Information Theory, 1994, p. 204.

[33] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, Anonymity protocols as noisy channels, Inf. Comput. 206 (2-4) (2008) 378–401.

[34] P. Malacaria, H. Chen, Lagrange multipliers and maximum information leakage in different observational models, in: Ú. Erlingsson, M. Pistoia (Eds.), PLAS, ACM, 2008, pp. 135–146.

[35] B. Köpf, D. A. Basin, An information-theoretic model for adaptive side-channel attacks, in: P. Ning, S. D. C. di Vimercati, P. F. Syverson (Eds.), ACM Conference on Computer and Communications Security, ACM, 2007, pp. 286–296.

[36] D. Clark, S. Hunt, P. Malacaria, A static analysis for quantifying information flow in a simple imperative language, Journal of Computer Security 15 (3) (2007) 321–371.

[37] D. Clark, S. Hunt, P. Malacaria, Quantitative analysis of the leakage of confidential data, Electr. Notes Theor. Comput. Sci. 59 (3).

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

[39] A. K. Hirsch, M. R. Clarkson, Belief semantics of authorization logic, in: A. Sadeghi, V. D. Gligor, M. Yung (Eds.), 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013, ACM, 2013, pp. 561–572. doi:10.1145/2508859.2516667.
URL http://doi.acm.org/10.1145/2508859.2516667

[40] M. R. Clarkson, F. B. Schneider, Quantification of integrity, Mathematical Structures in Computer Science 25 (2) (2015) 207–258. doi:10.1017/S0960129513000595.
URL http://dx.doi.org/10.1017/S0960129513000595

[41] S. H. Hussein, A precise information flow measure from imprecise probabilities, in: SERE, IEEE, 2012, pp. 128–137.

[42] G. J. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, Wiley-Interscience, 2005.

[43] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, On the bayes risk in information-hiding protocols, Journal of Computer Security 16 (5) (2008) 531–571. doi:10.3233/JCS-2008-0333.
URL http://hal.inria.fr/inria-00349224/en/

[44] C. Braun, K. Chatzikokolakis, C. Palamidessi, Quantitative notions of leakage for one-try attacks, in: Proceedings of the 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009), Vol. 249 of ENTCS, Elsevier B.V., 2009, pp. 75–91.

[45] M. Andrés, C. Palmidessi, P. van Rossum, G. Smith, Computing the amount of leakage in information-hiding systems, Tech. rep., LIX, Ecole Polytechnique (2009).

[46] R. Chadha, U. Mathur, S. Schwoon, Computing information flow using symbolic model-checking, in: V. Raman, S. P. Suresh (Eds.), 34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2014, December 15-17, 2014, New Delhi, India, Vol. 29 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2014, pp. 505–516. doi:10.4230/LIPIcs.FSTTCS.2014.505.

[47] K. Chatzikokolakis, T. Chothia, A. Guha, Statistical measurement of information leakage, in: J. Esparza, R. Majumdar (Eds.), Proceedings

of the 16th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS), Vol. 6015 of Lecture Notes in Computer Science, Springer, 2010, pp. 390–404.

[48] D. Dubois, H. Prade, Possibility theory, probability theory and multiple-valued logics: A clarification, Ann. Math. Artif. Intell. 32 (1-4) (2001) 35–66.

[49] L. A. Zadeh, Book review: A mathematical theory of evidence, AI Magazine 5 (3) (1984) 81–83.
URL http://www.aaai.org/ojs/index.php/aimagazine/article/view/452

[50] R. R. Yager, On the dempster-shafer framework and new combination rules, Inf. Sci 41 (2) (1987) 93–137.
URL http://dx.doi.org/10.1016/0020-0255(87)90007-7

[51] P. Smets, R. Kennes, The transferable belief model, in: R. R. Yager, L. Liu (Eds.), Classic Works of the Dempster-Shafer Theory of Belief Functions, Vol. 219 of Studies in Fuzziness and Soft Computing, Springer, 2008, pp. 693–736.
URL http://dx.doi.org/10.1007/978-3-540-44792-4_28

[52] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4 (1988) 244–264.

[53] P. Smets, Belief functions: The disjunctive rule of combination and the generalized bayesian theorem, Int. J. Approx. Reasoning 9 (1) (1993) 1–35.

[54] J. Dezert, F. Smarandache, A new probabilistic transformation of belief mass assignment, CoRR abs/0807.3669.

[55] P. Smets, Decision making in the tbm: the necessity of the pignistic transformation, Int. J. Approx. Reasoning 38 (2) (2005) 133–147.

[56] J.-Y. Jaffray, Application of linear utility theory to belief functions, in: B. Bouchon-Meunier, L. Saitta, R. R. Yager (Eds.), IPMU, Vol. 313 of Lecture Notes in Computer Science, Springer, 1988, pp. 1–8.

[57] C. K. Murphy, Combining belief functions when evidence conflicts, Decision Supp. Syst. 29 (1) (2000) 1–9.

[58] M. C. Florea, A.-L. Jousselme, É. Bossé, D. Grenier, Robust combination rules for evidence theory, Information Fusion 10 (2) (2009) 183–197.

[59] Z. Liu, J. Dezert, Q. Pan, G. Mercier, Combination of sources of evidence with different discounting factors based on a new dissimilarity measure, Decision Support Systems 52 (1) (2011) 133–141.

[60] A.-L. Jousselme, P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, Int. J. Approx. Reasoning 53 (2) (2012) 118–145.

Appendix A. Some notions of the Dempster-Shafer Theory

In this paper, we have considered the case where the adversary is combining information from an external (and potentially unreliable) source and the observables of the program/protocol in order to increase her chance of breaking the system. Many fusion theories such as *possibility theory* [48], *probability theory* [12], and *Dempster-Shafer theory* [30, 31] have been proposed for the combination of information from different sources. Both possibility and probability theories can model imprecise and uncertain data at the same time. But the Dempster-Shafer theory is more adapted because it generalizes the others and, in particular, it has higher ability to combine information from (partially or totally) disagreeing sources.

Appendix A.1. Belief functions

Let $\mathcal{X}$ be a *frame of discernment*, that is, a finite set of exhaustive and mutually exclusive hypotheses[6].

Definition 10 (Basic belief assignment (bba)). A basic belief assignment (aka *belief mass function*) is a mapping $m : 2^{\mathcal{X}} \longrightarrow [0, 1]$ that satisfies:

$$\begin{array}{ll} (1): & m(\emptyset) = 0; \\ (2): & \sum_{\Theta \subseteq \mathcal{X}} m(\Theta) = 1. \end{array}$$

[6]Non-exhaustive and infinite frames of discernment are not considered.

The mass $m(\Theta)$ expresses our confidence or belief that the actual state of the world belongs to the subset Θ. It represents only the proportion of all relevant and available evidence supporting Θ and makes no additional claims about any subset of Θ. Note that the condition $m(\emptyset) = 0$ specifies the closed-world assumption, that is, the exhaustiveness of the hypotheses in the frame of discernment.

The *focal elements* of a bba m are the subsets Θ of $\mathcal{X}$ such that $m(\Theta) > 0$. The set of the focal elements constitutes the *core* of the bba. A *categorical belief* is a bba m focusing on a single element, that is, there exists a subset Θ such that $m(\Theta) = 1$. A categorical belief focusing on the entire world $\mathcal{X}$ is called a *vacuous belief* and expresses a total ignorance about the actual state of the world. A bba focusing on singletons of $\mathcal{X}$ is a *Bayesian belief*. Finally, the set $\{(\Theta, m(\Theta) \,|\, \Theta \in 2^{\mathcal{X}} \text{ and } m(\Theta) > 0\}$ of focal elements together with their mass is called a *body of evidence* (BOE). By abuse of language we shall identify a BOE by its corresponding bba. We also write $m(x)$ for $m(\{x\})$.

Since our attacker may have collected her side information from many (potentially conflicting) sources, she needs to find a way of combining the different BOEs provided by her sources. Moreover, since some of these sources might be unreliable, she also needs a sound technique to discount beliefs from unreliable sources. Finally, since our adversary is a decision maker as she needs to make one single guess based on her belief, see needs a way of deriving a subjective probability distribution from her belief mass function. The following sections present such techniques.

Appendix A.2. Combination of information

Given two bodies of evidence m_1 and m_2, the more classical way of combining them is Dempster's rule of combination (DRC) defined as

$$m_1 \oplus m_2(\Theta) = \begin{cases} 0 & \text{if } \Theta = \emptyset \\ \frac{\sum_{\Omega_1 \cap \Omega_2 = \Theta} m_1(\Omega_1) m_2(\Omega_2)}{1 - \sum_{\Omega_1 \cap \Omega_2 = \emptyset} m_1(\Omega_1) m_2(\Omega_2)} & \text{otherwise.} \end{cases} \tag{A.1}$$

DRC is not the only rule to combine two bbas. Indeed, there is a 'jungle' of combination rules developed in the literature essentially to answer the following Zadeh's [49] famous counterexample. Consider $\mathcal{X} = \{x_0, x_1, x_2\}$ and two expert opinions given by $m_1(x_0) = 0.9$, $m_1(x_2) = 0.1$, and $m_2(x_1) = 0.9$, $m_2(x_2) = 0.1$. The DRC gives $m_1 \oplus m_2(x_2) = 1$. Hence, since the two experts are highly conflicting in the hypothesis they strongly support,

the alternative x_2, hardly supported by each of them, turn out to be fully supported after the combination.

Several alternatives solutions have been proposed that all reject the $m_1 \oplus m_2(x_2) = 1$ solution (e.g [50, 51, 52]). None has reached a universal acceptance and interestingly DRC remains the most widely accepted and used in many application. The main concern with all these rules is that they are all based on the assumption that the sources of information are fully reliable. A strong assumption hardly supported for real life applications. Indeed, in intelligence gathering as well as in many real world information fusion systems, one must rate separately the quality (reliability, trustworthiness) of both the source and the content of the report. If the source is judged unreliable then the contain must by discarded. Therefore, one must first determine the reliability of the sources of the bbas before combining them.

Appendix A.3. Belief reliability

If one assumes that the source of his belief is fully reliable then he must accept it as it is. On the contrary, if the source is unreliable, then the belief must be discarded. An intermediate case is where we accept that the source might be more or less unreliable. Thus, we are building a bba m_ρ over the space $\mathcal{R} = \{r, nr\}$ representing our beliefs about the fact that the source of our bba m is reliable (r) or not reliable (nr). Let $\alpha = m_\rho(r)$ be our belief about the fact that the source is reliable. If $\alpha = 1$ then we would accept m. If $\alpha = 0$ then we would discard m and our belief would become vacuous. More generally, said α is the confidence level on the reliability of the source of the bba m about Θ, the application of the General Bayesian Theorem [53] produces our discounted [31] belief about Θ,

$$m^\alpha(\Theta) = \begin{cases} \alpha m(\Theta) & \text{if } \Theta \neq \mathcal{X} \\ \\ \alpha m(\mathcal{X}) + (1 - \alpha) & \text{otherwise.} \end{cases} \tag{A.2}$$

Every mass in m is shrunk by a factor α and the mass lost is transferred to the universe (i.e. the frame of discernment). The factor α is called the *coefficient of the reliability* and $\delta = 1 - \alpha$ is the *discount rate*.

It is easy to see that discounting a discounted bba[7] is equivalent to discounting the bba by the product of the coefficients of the reliability.

[7]When for instance the evaluator of source of a bba need to be rated himself.

Lemma 3. $m^{\alpha\alpha'}(\Theta) = m^{\alpha\alpha'}(\Theta)$.

Appendix A.4. Pignistic transformation

The mass $m(\Theta)$ expresses our confidence that the actual state of the world belongs to the subset Θ and makes no additional claims about any subset of Θ. However, it is usually necessary to use the beliefs to make a decision, like the single-try attack scenario considered in this paper. In this case, a rational decision-maker should approximate the bba by a subjective probability measure, known as *pignistic probability* measure, based on the underlying frame of discernment. Many pignistic transformations (aka *pignistic probabilizations*) have been proposed (e.g [54]). Here, we only consider the well-known and widely used transformation defined as [55]

$$p_m(x) = \sum_{\Theta \subseteq \mathcal{X} \text{ s.t. } x \in \Theta} \frac{1}{|\Theta|} m(\Theta) \text{ for all } x \text{ in } \mathcal{X}, \tag{A.3}$$

where $|\Theta|$ is the cardinality of the set Θ.

The next lemma shows that the transformation of a categorical belief redistributes uniformly the mass of its unique focal element Θ to the singletons of Θ. In particular, when m is vacuous, its transformation is the uniform probability distribution over the frame of discernment $\mathcal{X}$. The transformation of a Bayesian belief m is equal to m. Finally, the discounted mass of a belief is uniformly redistributed to the singletons of the frame after the transformation.

Lemma 4. 1. *If m is categorical focusing on Θ, then*

$$p_m(x) = \begin{cases} \frac{1}{|\Theta|} & \text{if } x \in \Theta \\ 0 & \text{otherwise.} \end{cases}$$

2. *If m is vacuous then $p_m(x) = \frac{1}{|\mathcal{X}|}$ for all x in $\mathcal{X}$.*
3. *if m is a Bayesian belief then $p_m(x) = m(x)$ for all x in $\mathcal{X}$.*
4. $p_{m^\alpha}(x) = \alpha\, p_m(x) + \frac{1-\alpha}{|\mathcal{X}|}$.

In summary, the adversary constructs her belief as follow: she collects information from different sources represented by the BOEs, she discounts each

BOE according to the reliability of its source, she combines the discounted BOEs using the Dempster's rule of combination, then she finally derives the subjective probability distribution thanks to the pignistic transformation above. This subjective probability distribution models the adversary's belief in our framework.

Appendix B. Reliability and discounting

In Section 4.2, our belief updating based on the Bayes' rule requires the admissibility restriction. We introduce here a new approach which allow us to estimate the admissibility factor ϵ of an arbitrary belief given the channel matrix. More precisely, given an arbitrary belief p_β, we will compute its reliability factor α w.r.t. the evidence induced by the channel (viz its observables). The discounted belief p_β^α is ϵ-admissible, with $\epsilon = (1-\alpha)$. Indeed the following holds.

Proposition 13. *Let p_β be an arbitrary belief and α be the reliability factor of its source. Then the discounted belief p_β^α is $(1-\alpha)$-admissible.*

Proof. From Lemma 4 we have $p_\beta^\alpha(a) = \alpha p_\beta(a) + \frac{1-\alpha}{|\mathcal{A}|} \geq \frac{1-\alpha}{|\mathcal{A}|}$. Hence, p_β^α is $(1-\alpha)$-admissible when $\alpha \neq 1$. □

Appendix B.1. Belief reliability

Estimating the reliability of beliefs from different sources of evidence is a challenging and very popular topic of research. Indeed, the Theory of Evidence (viz Theory of Belief) is becoming increasingly popular and widely used for decision making [56, 57] in information fusion. Moreover, in many real applications, not all the sources of evidence might have the same reliability and there is no prior knowledge about it. Recently, many discounting frameworks (e.g. [58, 59]), mainly based on two different approaches, have been proposed to enhance the trustworthiness of information from unreliable sources of evidence. On the one hand, there are methods that propose distance metrics[8] to capture the level (viz the quantity) of the dissimilarity between different opinions, but cannot show whether or not they conflict in the hypothesis they strongly support. On the other hand, most of the metrics

[8]See [60] for more detail on distances between beliefs.

proposed in the literature measure the *conflict* between beliefs. These metrics capture well the quality of the dissimilarity of opinions, that is, whether or not they distribute most of their support to compatible (viz same) elements. Having noticed that both approaches fail to give satisfactory results in general and that each of them capture only one of two complementary aspects of the dissimilarity, [59] propose to combine both approaches.

The distance and conflict metrics of [59] translate in our setting as

$$\mathrm{Dist}(p,q) = \frac{1}{2}\Big(\sum_{a\in\mathcal{A}} |p(a) - q(a)|\Big)$$

and

$$\mathrm{Conf}(p,q) = \begin{cases} 0 & \text{if } \operatorname{argmax}_{a\in\mathcal{A}} p(a) \cap \operatorname{argmax}_{a\in\mathcal{A}} q(a) \neq \emptyset \\ p(a_0)q(a'_0) & \text{otherwise,} \end{cases}$$

where $a_0 \in \operatorname{argmax}_{a\in\mathcal{A}} p(a)$ and $a'_0 \in \operatorname{argmax}_{a\in\mathcal{A}} q(a)$.

An immediate consequence is that the distance between two beliefs is null if and only if the two opinions fully agree, i.e. they produces the same belief. Conversely, the distance is maximum if they fully contradict each other.

Lemma 5. *Dist*$(p,q) = 0$ *iff* $\forall a \in \mathcal{A},\ p(a) = q(a)$.

Lemma 6. *Dist*$(p,q) = 1$ *iff* $\forall a \in \mathcal{A},\ p(a) \neq 0 \Rightarrow q(a) = 0$.

Similarly, the conflict is maximum if and only if they fully support two distinct elements.

Lemma 7. *Conf*$(p,q) = 1$ *iff* $\exists a_0, a'_0 \in \mathcal{A}$ *s.t.* $p(a_0) = q(a'_0) = 1$ *and* $a_0 \neq a'_0$.

The disagreement between two beliefs is an indicator of the unreliability of at least one of them. If they totally disagree then at least one of them is unreliable, while if they totally agree then both are equi-reliable. In this framework we adopt the following *dissimilarity metric* as an estimation of the relative disagreement between two beliefs.

$$\mathrm{Dissim}(p,q) = \frac{\mathrm{Dist}(p,q) + \mathrm{Conf}(p,q)}{1 + \mathrm{Dist}(p,q)\cdot\mathrm{Conf}(p,q)}. \tag{B.1}$$

Appendix B.2. Discount rate

In order to determine the coefficient of the reliability of the belief, we first need to determine the belief induced by an evidence o. For example, considering the channel `PROG C5` (see Table 3), when the adversary sees o_2 then the belief mass m_{o_2} induced by her observation is

$$m_{o_2}(\Omega) = \begin{cases} 1 & \text{if } \Omega = \{a_2\} \\ 0 & \text{otherwise.} \end{cases}$$

But if she observes o_1 then she knows that the high input is either a_0 or a_3 but has no clear reason to prefer one over the other, even though o_1 is unlikely when the high input is a_0. In fact, if the actual a priori distribution of A is such that $p(a_3) = 0.01$, then when we observe o_1, the unlikely has happened either at the random generation of A or in the channel output. Therefore the only reasonable belief mass induced by o_1 is

$$m_{o_1}(\Omega) = \begin{cases} 1 & \text{if } \Omega = \{a_0, a_3\} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the belief mass induced by an observable o is the categorical belief mass

$$m_o(\Omega) \;=\; \begin{cases} 1 & \text{if } \Omega = \{a \in \mathcal{A} \,|\, p(o \,|\, a) > 0\} \\ 0 & \text{otherwise.} \end{cases}$$

Now let $\mathcal{A}_o = \{a \in \mathcal{A} \,|\, p(o \,|\, a) > 0\}$ be the focal set of the belief mass induced by o. Then by the pignistic probabilization, we obtain the following belief over $\mathcal{A}$ induced by o:

$$p_o(a) \;=\; \begin{cases} \frac{1}{|\mathcal{A}_o|} & \text{if } a \in \mathcal{A}_o \\ 0 & \text{otherwise.} \end{cases} \tag{B.2}$$

The dissimilarity metric defined in Equation B.1 allows us to estimate the relative disagreement between two beliefs. However in an uncertain environment where the reliability of the sources of both beliefs is unknown, we have no clear way to tell which of them is more reliable as the dissimilarity only measures their relative disagreement. Fortunately, in this paper, the belief p_o induced by a channel output is fully reliable. Hence the dissimilarity between p_o and p_β is a good indicator of the unreliability of the prebelief p_β.

Therefore we compute the discount rate of p_β induced by o as the relative dissimilarity between p_β and p_o given by

$$\delta_o = \frac{\mathrm{Dist}(p_\beta, p_o) + \mathrm{Conf}(p_\beta, p_o)}{1 + \mathrm{Dist}(p_\beta, p_o) \cdot \mathrm{Conf}(p_\beta, p_o)}. \tag{B.3}$$

It is easy to see that, on the one hand the discount rate is null if and only if the evidence o fully agrees with the initial belief. On the other hand, the initial belief is fully discounted if and only if o contradicts it.

Lemma 8. *(1)* $\delta_o = 0$ *iff* $\forall a \in \mathcal{A}_o,\ p_\beta(a) = \frac{1}{|\mathcal{A}_o|}$ *and (2)* $\delta_o = 1$ *iff* $\forall a \in \mathcal{A}_o,\ p_\beta(a) = 0$.

Proof. We start with Equivalence (1).

$$\begin{aligned}
\delta_o = 0 \quad &\iff \quad \mathrm{Dist}(p_\beta, p_o) = conf(p_\beta, p_o) = 0 \quad \text{(Lemma 5)} \\
&\iff \quad p_\beta = p_o \quad \text{(Eq B.2)} \\
&\iff \quad \forall a \in \mathcal{A}_o,\ p_\beta(a) = \frac{1}{|\mathcal{A}_o|}.
\end{aligned}$$

For Equivalence (2):

$$\begin{aligned}
\delta_o = 1 \quad &\iff \quad \Big(\mathrm{Dist}(p_\beta, p_o) = 1\Big) \vee \Big(\mathrm{Conf}(p_\beta, p_o) = 1\Big) \quad \text{(Lemmas 6 and 7)} \\
&\iff \quad \Big(\forall a \in \mathcal{A},\ p_o(a) \neq 0 \Rightarrow p_\beta(a) = 0\Big) \vee \\
&\qquad \Big(\exists a_0, a_0' \in \mathcal{A} \text{ s.t. } p(a_0) = q(a_0') = 1 \text{ and } a_0 \neq a_0'\Big) \\
&\iff \quad \forall a \in \mathcal{A}_o,\ p_\beta(a) = 0.
\end{aligned}$$

□

We finally define the discount rate of a belief w.r.t. a channel as the expectation of the discount rates induced by its observables. Note that our approach of deriving p_o as the pignistic probabilization of the categorical belief mass m_o is equivalent to Bayesian update of a vacuous belief (i.e. in the absence of any information) when observing o. Therefore, we take the expectation under the uniform distribution $p_u(a)$ over $\mathcal{A}$. Hence, the discount rate of a belief p_β w.r.t. a channel $p(o \,|\, a)$ is

$$\delta = \sum_{o \in \mathcal{O}} p_u(o)\delta_o \text{ with } p_u(o) = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} p(o|a). \tag{B.4}$$

Next we show that a belief cannot be fully discounted since it cannot totally conflict with all the observables.

Theorem 7. $0 \leq \delta < 1$.

Proof. We start with the proof of $0 \leq \delta$.

$$\forall o \in \mathcal{O},\ 0 \leq \delta_o \text{ and } 0 \leq p_u(o) \Longrightarrow 0 \leq \sum_{o \in \mathcal{O}} p_u(o)\delta_o \Longrightarrow \ \ 0 \leq \delta \text{ (Eq. B.4)}.$$

To prove $\delta < 1$, we show that $\delta = 1$ leads to a contradictory belief.

$$\begin{aligned}
\delta = 1 \quad &\Longleftrightarrow \quad \forall o \in \mathcal{O},\ \delta_o = 1 \\
&\Longleftrightarrow \quad \forall o \in \mathcal{O},\ \Big(\mathrm{Dist}(p_\beta, p_o) = 1\Big) \vee \Big(\mathrm{Conf}(p_\beta, p_o) = 1\Big) \\
&\Longleftrightarrow \quad \forall o \in \mathcal{O},\ \forall a \in \mathcal{A}, \Big(\ p_o(a) \neq 0 \Rightarrow p_\beta(a) = 0\Big) \vee \\
&\qquad\qquad \Big(\exists a_0, a_0' \in \mathcal{A} \text{ s.t. } p(a_0) = q(a_0') = 1 \text{ and } a_0 \neq a_0'\Big) \\
&\qquad\qquad \text{(From Lemmas 6 and 7)} \\
&\Longleftrightarrow \quad \forall o \in \mathcal{O},\ \forall a \in \mathcal{A}_o,\ p_\beta(a) = 0 \\
&\Longleftrightarrow \quad \forall a \in \mathcal{A},\ p_\beta(a) = 0 \quad (\text{ since } \cup_{o \in \mathcal{O}} \mathcal{A}_o = \mathcal{A}).
\end{aligned}$$

Hence p_β would be a contradictory belief, which is not possible under our closed-world assumption. □

From the first point of Lemma 8, from Theorem 7 and Proposition 13, we know that the discounted belief $p_\beta^{(1-\delta)}(a)$ is never in full contradiction with the evidence o. We therefore define the posteriori belief as the Bayesian update of the discounted belief, i.e.

$$\widetilde{p}_\beta\,(a \,|\, o) \ \ = \ \ \frac{p(o \,|\, a) p_\beta^\alpha(a)}{\sum_{a' \in \mathcal{A}} p(o \,|\, a') p_\beta^\alpha(a')}, \quad \text{whith } \alpha = 1 - \delta. \tag{B.5}$$