Theoretical Computer Science 777 (2019) 32–54

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

An axiomatization of information flow measures

Mário S. Alvim [a,*], Konstantinos Chatzikokolakis [b,c], Annabelle McIver [e], Carroll Morgan [f,g], Catuscia Palamidessi [d,c], Geoffrey Smith [h]

[a] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b] CNRS, Palaiseau, France
[c] LIX, École Polytechnique, Palaiseau, France
[d] Inria Saclay, Palaiseau, France
[e] Macquarie University, Sydney, Australia
[f] University of New South Wales, Sydney, Australia
[g] Data61, Sydney, Australia
[h] Florida International University, Miami, United States of America

ARTICLE INFO

Article history:
Received 30 June 2018
Received in revised form 10 October 2018
Accepted 16 October 2018
Available online 19 October 2018

Keywords:
Information flow
g-Vulnerability
Information theory
Confidentiality
Axioms

ABSTRACT

Quantitative information flow aims to assess and control the leakage of sensitive information by computer systems. A key insight in this area is that no single leakage measure is appropriate in all operational scenarios; as a result, many leakage measures have been proposed, with many different properties. To clarify this complex situation, this paper studies information leakage axiomatically, showing important dependencies among different axioms. It also establishes a completeness result about the *g*-leakage family, showing that any leakage measure satisfying certain intuitively-reasonable properties can be expressed as a *g*-leakage.

1. Introduction

The theory of *quantitative information flow* has seen rapid development over the past decade, motivated by the need for rigorous techniques to *assess* and *control* the leakage of sensitive information by computer systems. The starting point of this theory is the modeling of a *secret* as something whose value is known to the adversary only as a *prior probability distribution* π. This immediately suggests that the "amount" of secrecy might be quantified based on π, where intuitively a uniform π would mean "more" secrecy and a biased π would mean "less" secrecy. But how, precisely, should the quantification be done?

Early work in this area (e.g., [1]) adopted classic information-theoretic measures like *Shannon-entropy* [2] and *guessing-entropy* [3]. But these can be quite misleading in a security context, because they can be arbitrarily high even if π assigns a large probability to one of the secret's possible values, giving the adversary a large chance of guessing that secret correctly in just one try. This led to the introduction of *Bayes vulnerability* [4], which is simply the maximum probability that π assigns to any of the possible values of the secret. Bayes vulnerability indeed measures a basic security threat, but it implicitly assumes an operational scenario where the adversary must guess the secret exactly, in one try. There are of course many other possible scenarios, including those where the adversary benefits by guessing a *part* or a *property* of the secret

* Corresponding author.
E-mail address: msalvim@dcc.ufmg.br (M.S. Alvim).

https://doi.org/10.1016/j.tcs.2018.10.016
0304-3975/

or by guessing the secret within *three tries*, or where the adversary is *penalized* for making an incorrect guess. This led to the introduction of *g-vulnerability* [5], which uses *gain functions* g to model the operational scenario, enabling specific g-vulnerabilities to be tailored to each of the above scenarios, and many others as well.[1]

This situation may however strike us as a bit of a zoo. We have a multitude of exotic vulnerability measures, but perhaps no clear sense of what a vulnerability measure ought to be. Are all the g-vulnerabilities "reasonable"? Are there "reasonable" vulnerability measures that we are missing?

The situation becomes more complex when we turn our attention to systems. We model systems as information-theoretic *channels*, and the crucial insight, reviewed in Section 2.2 below, is that each possible output of a channel allows the adversary to update the prior distribution π to a *posterior distribution*, where the posterior distribution itself has a probability that depends on the probability of the output. Hence a channel is a mapping from prior distributions to *distributions on posterior distributions*, called *hyper-distributions* [6].

In assessing *posterior vulnerabilities*, by which we mean the vulnerability after the adversary sees the channel output, we have a number of choices. It is natural to consider the vulnerability of each of the posterior distributions, and take the *average*, weighted by the probabilities of the posterior distributions. Or (if we are pessimistic) we might take the *maximum*. Next we can define the *leakage* caused by the channel by comparing the posterior vulnerability and prior vulnerability, either multiplicatively or additively. These choices, together with the multitude of vulnerability measures, lead us to many different leakage measures, with many different properties. Is there a systematic way to understand them? Can we bring order to the zoo?

Such questions motivate the axiomatic study that we undertake in this paper. We consider a set of axioms that characterize intuitively-reasonable properties that vulnerability measures might satisfy, separately considering axioms for prior vulnerability (Section 4) and axioms for posterior vulnerability and for the *relationship* between prior and posterior vulnerability (Section 5). Addressing this relationship is an important novelty of our axiomatization, as compared with previous axiomatizations of entropy (such as [2,7,8]), which considered only prior entropy, or the axiomatization of *utility* by Kifer and Lin [9], which considers posterior utility without investigating its relation to prior utility. As a result, our axiomatization is able to consider properties of *leakage*, usually defined in terms of comparison between the posterior and prior vulnerabilities. We should however clarify that we do not view axiomatics as a matter of identifying "self-evident" truths. A variety of axioms may appear intuitively reasonable, so while it is sensible to consider intuitive justifications for them, such justifications should not be considered absolute. Rather we see the value of axiomatics as consisting more in understanding the logical dependencies among different properties, so that we might (for instance) identify a minimal set of axioms that is sufficient to imply all the properties that we care about.

The main contributions of this paper are of two kinds. One kind involves showing interesting *dependencies* among the various axioms. For instance, under axiom *averaging* for posterior vulnerability, we prove in Section 5 that three other axioms are equivalent: *convexity*, *monotonicity* (i.e., non-negativity of leakage), and the *data-processing inequality*. Convexity is the property that vulnerability is a *convex function* from distributions to reals; what is striking here is that it is a property that might not be intuitively considered "fundamental", yet our equivalence (assuming averaging) shows that it is. We also show an equivalence under the alternative axiom *maximum* for posterior vulnerability, which then involves *quasi-convexity*.

A second kind of contribution justifies the significance of g-vulnerability. Focusing on the axioms of *convexity* and *continuity* for prior vulnerability, we consider the class of *all* functions from distributions to reals that satisfy them, proving in Section 4 that this class *exactly coincides* with the class of g-vulnerabilities. This *soundness* and *completeness* result shows that if we accept averaging, continuity, and convexity (or monotonicity or the data-processing inequality) then prior vulnerabilities are exactly g-vulnerabilities.

Plan of the paper. The rest of the paper is structured as follows: Section 2 reviews the basic concepts of quantitative information flow, Section 3 sets up the framework of our axiomatization, and Sections 4 and 5 discuss axioms for prior and posterior vulnerabilities, respectively. Section 6 provides some discussion, Section 7 gives a more abstract perspective, Section 8 discusses related work, and Section 9 concludes.

A preliminary version of this paper appeared in [10]. Additional material presented here includes: (i) proofs; (ii) a thoroughly revised presentation of the soundness and completeness of g-vulnerabilities with respect to continuous, convex functions in Section 4.1, based on a deeper yet simplified characterization of the geometry of gain functions; and (iii) a more elaborate discussion of our results and their consequences.

2. Preliminaries

We now review some basic notions from quantitative information flow. A *secret* is something whose value is known to the adversary only as a *prior probability distribution* π: there are various ways for measuring what we will call its *vulnerability*. A *channel* models systems with observable behavior that changes the adversary's probabilistic knowledge, making the secret more vulnerable and hence causing information *leakage*.

[1] Note that *entropies* measure secrecy from the point of view of the *user* (i.e., more entropy means more secrecy), while *vulnerabilities* measure secrecy from the point of view of the *adversary* (i.e., more vulnerability means less secrecy). The two perspectives are complementary, but to avoid confusion this paper focuses almost always on the vulnerability perspective.

2.1. Secrets and vulnerability

The starting point of computer security is information that we wish to keep *secret*, such as a user's password, social security number, or current location. An adversary typically does not know the value of the secret, but still possesses some *probabilistic information* about it, captured by a probability distribution called the *prior*. We denote by $\mathcal{X}$ the finite set of possible secret values and by $\mathbb{D}\mathcal{X}$ the set of probability distributions over $\mathcal{X}$. A prior $\pi \in \mathbb{D}\mathcal{X}$ could either reflect a probabilistic procedure for choosing the secret—e.g., the probability of choosing a certain password—, or it could capture any knowledge the adversary possesses on the population the user comes from—e.g., a young person is likely to be located at a popular bar on Saturday night.

The prior π plays a central role at measuring how *vulnerable* a secret is. For instance, short passwords are not vulnerable because of their length (prefixing passwords with a thousand zeroes does not necessarily render them more secure), but because each password has a high probability of being chosen. To obtain a concrete vulnerability measure one needs to consider an *operational scenario* describing the adversary's capabilities and goals; vulnerability then measures the adversary's expected success in this scenario.

Bayes-vulnerability [4] considers an adversary trying to *guess* the secret in *one try* and measures the threat as the *probability* of the guess being correct. Knowing a prior π, a rational adversary will guess a secret to which it assigns the highest probability: hence Bayes-vulnerability is given by

$$V^b(\pi) \;=\; \max_{x\in\mathcal{X}} \pi_x \,,$$

where we write π_x for the probability π assigns to x. Note that Bayes-vulnerability is called simply "vulnerability" in [4], and is the basic notion behind *min-entropy*, defined as $H_\infty(\pi) = -\lg V^b(\pi)$. It is also the converse of the adversary's probability of error, also called *Bayes-risk* in the area of hypothesis testing [11].

Guessing-entropy [3] considers an adversary trying to guess the secret in an unlimited number of tries, and measures the adversary's uncertainty as the *number of guesses* needed on average. The best strategy is to try secrets in non-increasing order of probability: if x_i is an indexing of $\mathcal{X}$ in such an order, then guessing-entropy is given by

$$G(\pi) \;=\; \textstyle\sum_i i\,\pi_{x_i} \,.$$

Shannon-entropy [2] considers an adversary who tries to infer the secret using Boolean questions (i.e., of the form "does x belong to a certain subset $\mathcal{X}'$ of $\mathcal{X}$?") and measures the adversary's uncertainty as the *number of questions* needed on average. It can be shown that the best strategy is at each step to split the secret space in sets of equal probability (as far as possible). Under this strategy, a secret x will be guessed in $-\lg \pi_x$ steps, hence on average the number of questions needed is

$$H(\pi) \;=\; -\textstyle\sum_{x\in\mathcal{X}} \pi_x \lg \pi_x \,.$$

Note that Bayes-vulnerability measures the *threat* to the secret (the higher the better for the adversary). On the other hand, guessing- and Shannon-entropy measure the adversary's *uncertainty* about the secret (the lower the better for the adversary).

Although the operational scenarios described above capture realistic threats for the secret, one could envision a variety of alternative threats we might also be worried about. For instance, an adversary might be interested in guessing only *part* of the secret, an *approximate* value of the secret, a *property* of the secret, or guessing the secret in a fixed number of tries. It is for this reason that the more general *g-vulnerability* framework [5] was proposed: it allows one to adapt to many different adversarial models.

Its operational scenario is parametrized by a set $\mathcal{W}$ of *actions* (possibly infinite) that the adversary can make *about* the secret, and a *gain function* $g\colon \mathcal{W}\times\mathcal{X}\to\mathbb{R}$. The gain $g(w,x)$ expresses the adversary's benefit for having taken the action w when the actual secret is x. The g-vulnerability function measures the threat as the adversary's expected gain for an optimal choice of action w:

$$V_g(\pi) \;=\; \sup_{w\in\mathcal{W}} \textstyle\sum_{x\in\mathcal{X}} \pi_x g(w,x) \,. \tag{1}$$

Regarding the set $\mathcal{W}$ of allowable actions, one might assume that this should just be $\mathcal{X}$, the set of possible values of the secret. This is in fact too restrictive: the adversary's goal might be to guess a *piece* of the secret, or a value *close* to the secret, or some *property* of the secret. As a consequence we allow an arbitrary set of actions, possibly infinite, and make (almost) no restrictions on the choice of g. In particular, a negative value of $g(w,x)$ expresses situations when the adversary is *penalized* for making a particular action under a particular secret; such values are essential for obtaining the results of Section 4.1.3.

However, leaving g unrestricted has two side effects that are undesirable both *conceptually* and *technically*. First, V_g could potentially produce *negative* vulnerabilities. Conceptually, since we want to measure *how vulnerable* a secret is, it seems reasonable that the minimum possible vulnerability should be 0, meaning "not vulnerable at all". Technically, we will consider

multiplicative leakage, which measures leakage as the *ratio* of two vulnerabilities; but such a ratio seems mathematically meaningless if one vulnerability is positive and the other negative.

Second, V_g could potentially produce *infinite* vulnerabilities. Consider, for instance, the case when $\mathcal{W}$ is the set of all integers, $\mathcal{X} = \{x_1, x_2\}$, and g is given by $g(w, x_1) = w$ and $g(w, x_2) = -w$. For this example we find that V_g becomes not only infinite but also *discontinuous*: it is 0 on the uniform prior (for which the gain of one secret is exactly counterbalanced by the loss of the other), and ∞ for all other priors. Again, such behavior is both conceptually and technically problematic. Conceptually, it is unnatural for a system to be "infinitely vulnerable", and even more so for an adversary to be "infinitely risk-averse", perceiving an infinitesimal change in the prior as an infinite change in the vulnerability of the system. Technically, it is clear that discontinuous, infinite-valued functions are ill-behaved.

For these reasons, we always restrict to the class of gain functions $\mathbb{G}\mathcal{X}$, defined as

$$\mathbb{G}\mathcal{X} \;:=\; \{g \mid V_g : \mathbb{D}\mathcal{X} \to \mathbb{R}_{\geq 0}\}\,, \tag{2}$$

that is, those gain functions that produce non-negative and finite-valued vulnerabilities. In Section 4.1.2 we will see that this restriction also implies the continuity of V_g.

Note that, as its name suggests, V_g is a measure of vulnerability, i.e., of the threat to the secret. An equally expressive alternative is to define an "uncertainty" measure similarly, but using a *loss function* l instead of a gain function and assuming that the adversary wants to minimize loss. The uncertainty measure, parametrized by l, can be then defined dually as $U_l(\pi) = \inf_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} \pi_x l(w, x)$, and is often called *Bayes-risk* in the area of decision theory.

Due to the flexibility of gain functions, g-vulnerability is a very expressive framework, one that can capture a great variety of operational scenarios. This raises the natural question of which other vulnerability measures are expressible in this framework. Bayes-vulnerability is a straightforward example, captured by guessing the exact secret, i.e., taking $\mathcal{W} = \mathcal{X}$, and using the *identity* gain function defined as

$$g_{id}(w, x) = \begin{cases} 1, & \text{if } w = x, \\ 0, & \text{if } w \neq x. \end{cases}$$

Guessing-entropy can be also captured in this framework [12,13], this time using a loss function since it is an uncertainty measure. The adversary's action in this case is to guess a *permutation* of $\mathcal{X}$, i.e., the order in which secrets are chosen in the operational scenario of guessing-entropy. We can naturally define the loss $l(w, x)$ as the index of x in w, i.e. the number of guesses to find x, and using this loss function we get $U_l(\pi) = G(\pi)$.

Similarly, in the case of Shannon-entropy, the adversary tries to guess a strategy for constructing his questions. Strategies can be described as probability distributions: at each step questions split the search space into subsets of as even probability as possible. Hence, actions are $\mathcal{W} = \mathbb{D}\mathcal{X}$, and the loss can be defined as $l(w, x) = -\lg w_x$ (the number of steps needed to find x under the strategy w). Since the best strategy is to take $w = \pi$ itself, it can be shown [12] that under this loss function $U_l(\pi) = H(\pi)$.

In Section 4.1.3 we show that g-vulnerability exactly coincides with the generic class of continuous and convex vulnerability functions.

2.2. Channels, hypers and leakage

So far we have considered secrets for which a probabilistic prior is known, and have discussed different ways for measuring their vulnerability. We now turn our attention to *systems*, which are programs or protocols processing secret information and producing some *observable* behavior. Examples of such systems are password-checkers, implementations of cryptosystems, and anonymity protocols.

A system can be modeled as an *(information theoretic) channel*, a triple $(\mathcal{X}, \mathcal{Y}, \mathsf{C})$, where $\mathcal{X}, \mathcal{Y}$ are finite sets of (secret) input values and (observable) output values respectively and C is a $|\mathcal{X}| \times |\mathcal{Y}|$ channel matrix in which each entry $\mathsf{C}_{x,y}$ corresponds to the probability of the channel producing output y when the input is x. Hence each row of C is a probability distribution over $\mathcal{Y}$ (entries are non-negative and sum to 1). A channel is *deterministic* iff each row contains a single 1 identifying the only possible output for that input.

It is typically assumed that the adversary knows how the system works, i.e. knows the channel matrix C. Knowing also the prior distribution π, the adversary can compute the joint distribution $p(x, y) = \pi_x \mathsf{C}_{x,y}$ on $\mathcal{X} \times \mathcal{Y}$, producing joint random variables X, Y with marginal probabilities $p(x) = \sum_y p(x, y)$ and $p(y) = \sum_x p(x, y)$, and conditional probabilities $p(y|x) = {}^{p(x,y)}/_{p(x)}$ (if $p(x)$ is non-zero) and $p(x|y) = {}^{p(x,y)}/_{p(y)}$ (if $p(y)$ is non-zero). Note that p_{XY} is the unique joint distribution that recovers π and C, in that $p(x) = \pi_x$ and $p(y \mid x) = \mathsf{C}_{x,y}$ (if $p(x)$ is non-zero).[2]

For a given y (s.t. $p(y)$ is non-zero), the conditional probabilities $p(x|y)$ for each $x \in \mathcal{X}$ form the *posterior distribution* $p_{X|y}$, which represents the posterior knowledge the adversary has about input X after observing output y.

[2] When necessary to avoid ambiguity, we write distributions with subscripts, e.g. p_{XY} or p_Y.

Example 1. Given $\mathcal{X} = \{x_1, x_2, x_3\}$, $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and the channel matrix C below, (the uniform) prior $\pi = (1/3, 1/3, 1/3)$ combined with C leads to joint matrix J:

C	y_1	y_2	y_3	y_4
x_1	1	0	0	0
x_2	0	1/2	1/4	1/4
x_3	1/2	1/3	1/6	0

$\xrightarrow{\pi}$

J	y_1	y_2	y_3	y_4
x_1	1/3	0	0	0
x_2	0	1/6	1/12	1/12
x_3	1/6	1/9	1/18	0

Summing columns of J gives the marginal distributions $p_Y = (1/2, 5/18, 5/36, 1/12)$, and normalizing gives the posterior distributions $p_{X|y_1} = (2/3, 0, 1/3)$, $p_{X|y_2} = (0, 3/5, 2/5)$, $p_{X|y_3} = (0, 3/5, 2/5)$, and $p_{X|y_4} = (0, 1, 0)$.

The effect of a channel C is to update the adversary's knowledge from a prior π to a collection of posteriors $p_{X|y}$, each occurring with probability $p(y)$, called a *hyper-distribution*. A hyper (for short) on the input space $\mathcal{X}$ is of type $\mathbb{D}^2\mathcal{X}$, which stands for $\mathbb{D}(\mathbb{D}\mathcal{X})$, a distribution on distributions on $\mathcal{X}$. The support of a hyper is the set of possible posteriors that the application of channel C on prior π can produce: we call those posteriors *inners*. The probability assigned by the hyper to a particular inner is the marginal probability of the y that produced that inner. We call those probabilities the *outer* probabilities. We use Δ to denote a hyper, $\lceil\Delta\rceil$ for its *support* (the set of posteriors with non-zero probability), $[\pi]$ to denote the point-hyper assigning probability 1 to π, and $[\pi \triangleright \mathsf{C}]$ to denote the hyper obtained by the application of C on π. We say that $[\pi \triangleright \mathsf{C}]$ is the result of *pushing prior π through channel C*.

In Example 1, the hyper $[\pi \triangleright \mathsf{C}]$ assigns (outer) probabilities $(1/2, 15/36, 1/12)$ to the (inner) posteriors $(2/3, 0, 1/3)$, $(0, 3/5, 2/5)$, and $(0, 1, 0)$, respectively.[3]

Following [6,14], we can abstract from a *concrete channel* represented by its matrix C to an *abstract channel* C consisting just in the corresponding mapping from priors to hyper-distributions. Abstract channels ignore aspects of channel matrices that are irrelevant to leakage (e.g., labels and order of columns, and columns that are multiples of each other), and concentrate only on the essential information that affects leakage: the mapping from priors to hyper-distributions. We will denote concrete channels and channel matrices using a sans-serif font (C) and abstract channels using a math font (C).

Since the outcome of a channel is a hyper, it is natural to extend vulnerability measures from priors to hypers, obtaining a *posterior* vulnerability. For all measures described in Section 2.1 this has been done in a natural way by taking the vulnerability of each posterior and *averaging* them using the outer. Let

$$\mathcal{E}_\pi F \;:=\; \sum_x \pi_x F(x)$$

denote the *expected value* of some random variable $F : \mathcal{X} \to R$ (where R is usually the reals $\mathbb{R}$ but more generally can be a vector space) over a distribution $\pi : \mathbb{D}\mathcal{X}$. We can then define *posterior Bayes-vulnerability* $\widehat{V}^b : \mathbb{D}^2\mathcal{X} \to \mathbb{R}^+$ as

$$\widehat{V}^b\Delta \;=\; \mathcal{E}_\Delta V^b\,,$$

and similarly for Shannon-entropy, guessing-entropy and g-vulnerability. For hypers $[\pi \triangleright \mathsf{C}]$ produced by channels, from the above formula we can get an expression of each posterior vulnerability as a function of π and C, for instance,

$$\begin{aligned}
\widehat{V}^b[\pi \triangleright \mathsf{C}] &= \textstyle\sum_y \max_x \pi_x \mathsf{C}_{x,y}, && \text{for Bayes vulnerability, and}\\
\widehat{V}_g[\pi \triangleright \mathsf{C}] &= \textstyle\sum_y \sup_w \sum_x \pi_x \mathsf{C}_{x,y} g(w,x), && \text{for } g\text{-vulnerability.}
\end{aligned}$$

Note that, for point-hypers, we have by construction that $\widehat{V}^b[\pi] = V^b(\pi)$ and $\widehat{V}_g[\pi] = V_g(\pi)$, and similarly for the other measures.

Finally, the execution of a system is expected to disclose information about the secret to the adversary, and the *information leakage* of a channel C for a prior π is defined by comparing the vulnerability of the prior π—the adversary's prior knowledge—and that of $[\pi \triangleright \mathsf{C}]$—the adversary's posterior knowledge. The comparison is typically done either *additively* or *multiplicatively*, giving rise to two versions of leakage:

$$\text{additive:}\quad \mathcal{L}^{b,+}(\pi, \mathsf{C}) \;=\; \widehat{V}^b[\pi \triangleright \mathsf{C}] - V^b(\pi)\,, \qquad \text{and} \tag{3}$$

$$\text{multiplicative:}\quad \mathcal{L}^{b,\times}(\pi, \mathsf{C}) \;=\; \frac{\widehat{V}^b[\pi \triangleright \mathsf{C}]}{V^b(\pi)}\,. \tag{4}$$

Note that the logarithm of $\mathcal{L}^{b,\times}(\pi, \mathsf{C})$ is usually called *min-entropy leakage* [4]. Leakage can be similarly defined for all other measures.

[3] There might be fewer posteriors in the support of hyper $[\pi \triangleright \mathsf{C}]$ than there are columns in the joint distribution $p_{X,Y}$ from which it is derived, because if several columns of $p_{X,Y}$ normalize to the same posterior then the hyper will automatically coalesce them [14]. Columns y_2 and y_3 were coalesced in this case.

Table 1
Summary of axioms and their mnemonics for pairs of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$.

Axioms for prior vulnerabilities		
CNTY	$\forall \pi$:	$\mathbb{V}$ is a continuous function of π
CVX	$\forall \sum_i a_i \pi^i$:	$\mathbb{V}\left(\sum_i a_i \pi^i\right) \leq \sum_i a_i \mathbb{V}(\pi^i)$
Q-CVX	$\forall \sum_i a_i \pi^i$:	$\mathbb{V}\left(\sum_i a_i \pi^i\right) \leq \max_i \mathbb{V}(\pi^i)$
Axioms for posterior vulnerabilities		
NI	$\forall \pi$:	$\widehat{\mathbb{V}}[\pi] = \mathbb{V}(\pi)$
DPI	$\forall \pi, \mathsf{C}, \mathsf{R}$:	$\widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}] \geq \widehat{\mathbb{V}}[\pi \triangleright \mathsf{CR}]$
MONO	$\forall \pi, C$:	$\widehat{\mathbb{V}}[\pi \triangleright C] \geq \mathbb{V}(\pi)$
Possible definitions of posterior vulnerabilities		
AVG	$\forall \Delta$:	$\widehat{\mathbb{V}}\Delta = \mathcal{E}_\Delta \mathbb{V}$
MAX	$\forall \Delta$:	$\widehat{\mathbb{V}}\Delta = \max_{\lceil \Delta \rceil} \mathbb{V}$

3. An axiomatic view of vulnerabilities

In Section 2 we discussed vulnerability measures obtained by quantifying the threat to the secret in a specific operational scenario. Channels were then introduced, mapping prior distributions to hypers, and the vulnerability measures were naturally extended to posterior ones by averaging each posterior vulnerability over the hyper.

In this paper we shall follow a different approach: we axiomatize the study of vulnerabilities. We begin by considering generic vulnerability functions of type

$$\begin{aligned} \textit{prior vulnerability:} \quad & \mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}_{\geq 0}, \qquad \text{and} \\ \textit{posterior vulnerability:} \quad & \widehat{\mathbb{V}} : \mathbb{D}^2\mathcal{X} \to \mathbb{R}_{\geq 0}, \end{aligned}$$

and we consider a variety of properties that "reasonable" instantiations of these generic functions might be expected to have. We then formalize these properties as a set of *axioms* for vulnerability functions, and investigate their consequences.

We shall start, in the following section, by focusing on the prior case by giving axioms for prior vulnerabilities $\mathbb{V}$ alone. We then take convexity and continuity (made precise in Section 4 ahead) as our fundamental properties, and show that they lead to g-vulnerability exactly. After that, we turn our attention to axioms considering either both $\mathbb{V}$ and $\widehat{\mathbb{V}}$, or posterior $\widehat{\mathbb{V}}$ alone.

Moreover, we study two ways of constructing $\widehat{\mathbb{V}}$ from $\mathbb{V}$: by taking the average over the hyper, as we have been doing so far, and by considering the maximum-vulnerability over the hyper. It turns out that, in each case, several of the axioms become equivalent. An important observation is that the axioms purely affect the *relationship* between prior and posterior vulnerabilities, and are orthogonal to the way $\mathbb{V}$ and $\widehat{\mathbb{V}}$ are compared when used to measure leakage (i.e. multiplicatively or additively). Hence the results we obtain about the relationship among axioms are valid under both definitions of leakage.

It is important to have in mind that, although in this paper we consider axioms for generic vulnerability, dual axioms can be naturally stated for generic uncertainty measures.[4] In Table 1 we summarize the axioms we shall consider.

4. Axiomatization of prior vulnerabilities

We begin by introducing axioms that deal solely with prior vulnerabilities $\mathbb{V}$.

The first property we consider is that "small" changes on the prior π have a "small" effect on $\mathbb{V}$ applied to that prior. This intuition is formalized in the following axiom.

Definition 2 *(Axiom of continuity (CNTY)).* A vulnerability $\mathbb{V}$ is a continuous function of π (w.r.t. the standard topology[5] on $\mathbb{D}\mathcal{X}$).

Intuitively, the CNTY axiom captures adversaries who are not infinitely risk-averse. For instance, the non-continuous vulnerability function

$$\mathbb{V}^\lambda(\pi) \quad = \quad \begin{cases} 1, & \text{if } \max_x \pi_x \geq \lambda, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

[4] Recall that *vulnerabilities* measure secrecy from the point of view of the *adversary* (i.e. more vulnerability means less secrecy), whereas *entropies* measure *uncertainty*, that is, secrecy from the point of view of the *user* (i.e. more entropy means more secrecy). The two perspectives are complementary; in this paper we focus on the vulnerability perspective.

[5] The one induced by the Euclidean metric, or equivalently the total variation or 1/2-Manhattan metric used later in this section.

would correspond to an adversary who requires the probability of guessing correctly to be above a certain threshold λ in order to consider an attack effective at all. But this is an arguably unnatural behavior if we assume that the risk of changing the probability to $\lambda-\epsilon$, for an infinitesimal ϵ, should not be arbitrarily large. For instance, if $\lambda = 2/3$, distribution $\pi^1 = (2/3,\ 1/2)$ would be considered "insecure" by the function above, since $\mathbb{V}^{2/3}(\pi^1) = 1$, whereas distribution $\pi^2 = (2/3-\epsilon,\ 1/3+\epsilon)$ would be considered "secure", since $\mathbb{V}^{2/3}(\pi^2) = 0$.

The second property we consider is that V_g is a convex function of the prior.[6] More precisely, a *convex combination* of priors $\pi^1, \ldots, \pi^n$ is a sum $\sum_i a_i \pi^i$ where a_i's are non-negative reals adding up to 1. Since $\mathbb{D}\mathcal{X}$ is a convex set, a convex combination of priors is itself a prior. The property is then formalized as the following axiom.

Definition 3 *(Axiom of convexity (CVX)).* A vulnerability $\mathbb{V}$ is a convex function of π—that is, for all convex combinations $\sum_i a_i \pi^i$ we have

$$\mathbb{V}\left(\sum_i a_i \pi^i\right) \quad \leq \quad \sum_i a_i \mathbb{V}(\pi^i)\ .$$

The CVX axiom can be interpreted as follows. Consider a game in which a secret (say a password) is drawn from two possible distributions π^1 or π^2. The choice of distributions is itself random: we first select $i \in \{1, 2\}$ at random, with $i = 1$ having probability a_1 and $i = 2$ probability $a_2 = 1-a_1$, and then we use π^i to draw the secret.

Now consider the following two scenarios for this game: in the first scenario, the value of i is conveyed to the adversary, so that the actual prior the secret was drawn from is known. Using information in that π^i (whichever one it was) the adversary performs an attack, the expected success of which is measured by $\mathbb{V}(\pi^i)$. In this scenario the expected measure of success overall will be $\sum_i a_i \mathbb{V}(\pi^i)$. In the second scenario, the choice i is not disclosed to the adversary: she knows only that, on average, secrets are drawn from the prior $\sum_i a_i \pi^i$. Since to perform an attack the adversary can only use the information in $\sum_i a_i \pi^i$, the expected success of an attack in this case will be measured by $\mathbb{V}\left(\sum_i a_i \pi^i\right)$.

The CVX axiom corresponds to the intuition that, since in the first scenario the adversary has more information, the effectiveness of an attack can only be higher. Yet another way of seeing this axiom is to realize that an adversary should get no less information from a_1, π^1 and π^2, than from $\sum_i a_i \pi^i$, since the last value can be calculated if the first three are known.

Note that, in the definition of CVX, it is sufficient to use convex combinations of *two* priors, i.e., of the form $a\pi^1 + (1-a)\pi^2$; indeed we often use such combinations in proofs. Note also that CVX actually implies continuity everywhere except on the *boundary* of the domain, i.e., on priors having at least one element with probability exactly 0. The function $\mathbb{V}^\lambda$ from (5), for instance, for $\lambda = 1$ is convex but discontinuous. It captures an adversary that is only happy when knowing the secret with absolute certainty; such an adversary does satisfy CVX, yet continuity breaks for point priors (note that, for all $1/|\mathcal{X}| < \lambda < 1$, $\mathbb{V}^\lambda$ is neither continuous nor convex). The CNTY axiom, however, explicitly requires continuity everywhere.

Since the vulnerabilities $\mathbb{V}(\pi^i)$ in the definition of CVX are weighted by the probabilities a_i, we could have cases when the expected vulnerability $\sum_i a_i \mathbb{V}(\pi^i)$ is small although some individual $\mathbb{V}(\pi^i)$ is large. In such cases, one might argue that the bound imposed by CVX is too strict and could be loosened by requiring that $\mathbb{V}\left(\sum_i a_i \pi^i\right)$ is bounded only by the maximum of the individual vulnerabilities. This weaker requirement is formalized as the following axiom.

Definition 4 *(Axiom of quasi-convexity (Q-CVX)).* A vulnerability $\mathbb{V}$ is a quasi-convex function of π—i.e. for all convex combinations $\sum_i a_i \pi^i$ we have

$$\mathbb{V}\left(\sum_i a_i \pi^i\right) \quad \leq \quad \max_i \mathbb{V}(\pi^i)\ .$$

The justifications we have so far provided for the axioms of CVX and Q-CVX might not strike us as very intuitive at first, but it turns out that these axioms can in fact be justified as natural consequences of fundamental axioms relating prior and posterior vulnerabilities, and specific choices for constructing $\widehat{\mathbb{V}}$. We shall address these connections in detail in Section 5 ahead.

4.1. Soundness and completeness of V_g with respect to continuous, convex functions

It turns out that the vulnerability functions satisfying the axioms of CNTY and CVX are exactly those expressible as V_g for some gain function g. We will show each direction of this implication now.

4.1.1. A geometric view of gain functions

For our characterization of g-vulnerability, we exploit the geometry of gain functions together with some fundamental tools from convex analysis. In the following, $x \cdot x'$ denotes the dot product of vectors x, x'. A crucial observation is that an

[6] Note that, given the duality between vulnerability and uncertainty measures, for uncertainty measures a reasonable property would be *concavity* rather than *convexity*.

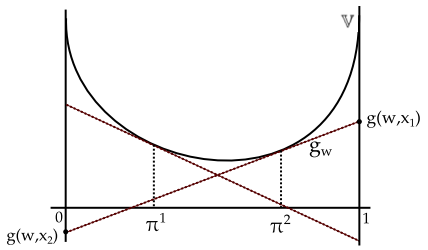

Fig. 1. Actions w constructed from subgradients on different priors.

action w can be thought of as a *vector* in $\mathbb{R}^n$ ($n = |\mathcal{X}|$), containing the gain for each secret, that is $w_x = g(w, x)$. Since a prior π is a vector itself, the expected gain of w can be written as

$$\mathcal{E}_\pi g(w, \cdot) \quad = \quad w \cdot \pi \ .$$

It is then clear that $w \cdot \pi$ is a linear function, as shown in Fig. 1.

A useful tool for the analysis of convex functions is that of *subgradients*, which generalize gradients for non-differentiable functions. A vector ϕ is a subgradient of $f : S \to \mathbb{R}$ at $x^* \in S$ iff

$$f(x) - f(x^*) \quad \geq \quad \phi \cdot (x - x^*) \qquad \text{for all } x \in S \ . \tag{6}$$

Note that the function $x \mapsto \phi \cdot (x - x^*) + f(x^*)$ is linear, always below f and touches f at x^*, as shown in Fig. 1. The set of all subgradients of f at x^* (called the *sub-differential*) is denoted by $\partial f(x^*)$.

A *norm* $\| \cdot \|$ is the standard tool for reasoning about the length of vectors. Any norm naturally induces a metric defined as $d_{\|\cdot\|}(x, x') := \|x - x'\|$. Moreover, any norm $\| \cdot \|$ has a *dual norm* $\| \cdot \|_*$ given by

$$\|x\|_* \quad := \quad \max_{z:\, \|z\| \leq 1} x \cdot z \ .$$

It is well known that the Euclidean norm is its own dual, while the dual of the Manhattan norm[7] $\| \cdot \|_1$ is the max norm $\| \cdot \|_\infty$ (in general the dual of a p-norm is the q-norm such that $1/p + 1/q = 1$). Directly from the definition of $\| \cdot \|_*$ we get that

$$|x \cdot y| \quad \leq \quad \|x\| \, \|y\|_* \qquad \text{for all } x, y \in \mathbb{R}^n \ , \tag{7}$$

which can be seen as a generalization of the Cauchy–Schwarz inequality.

The *interior* of a set $S \subseteq \mathbb{R}^n$ consists of points $x \in S$ such that some open ball centered at x is contained in S. It is common, however, that S is a lower dimensional object embedded in $\mathbb{R}^n$. A triangle, for instance, could be embedded in $\mathbb{R}^3$ although it is 2-dimensional. Such objects are "flat", so do not have a proper interior; however we can still talk about their *relative interior* w.r.t. the lower dimensional space they live in (i.e. their affine hull). For convex sets, the relative interior $\text{relint}(S)$ can be simply defined as the points $x \in S$ such that all line segments in S ending in x can be extended beyond x without leaving S [15, Thm. 6.1]. That is

$$\text{relint}(S) \quad := \quad \{x \in S \mid \forall z \in S. \exists \lambda > 1 : \lambda x + (1 - \lambda) z \in S\} \ .$$

Note that the probability simplex $\mathbb{D}\mathcal{X} \subset \mathbb{R}^n$ ($n = |\mathcal{X}|$) is an $(n-1)$-dimensional object that lies on the hyperplane $x \cdot \mathbf{1} = 1$ (probabilities sum up to 1). Hence its interior is empty, but $\text{relint}(\mathbb{D}\mathcal{X})$ is not: it consists exactly of all *full-support* distributions.

4.1.2. Every V_g satisfies continuity and convexity

We first show that any g-vulnerability, for $g : \mathbb{G}\mathcal{X}$, satisfies the axioms of CNTY and CVX. Let g be such a gain function with a possibly infinite set of actions $\mathcal{W}$. Recall that V_g can be expressed as the supremum of the family of functions:

$$V_g(\pi) \quad = \quad \sup_w w \cdot \pi \ .$$

Note that $w \cdot \pi$ is linear on π, hence both (trivially) convex and continuous.

The convexity of V_g then follows from the fact that the supremum of any family of convex functions is itself a convex function. On the other hand, showing continuity is more challenging, since the supremum of continuous functions is not necessarily continuous itself.[8]

[7] The *Manhattan norm* of a vector $(x_1, x_2, \ldots, x_n)$ is defined as $\|(x_1, x_2, \ldots, x_n)\|_1 = \sum_{i=1}^n |x_i|$.

[8] A counter-example would be $f(x) = \sup_{n>0} -(x-1)^n$, which is discontinuous at 0.

To show that V_g is continuous, we employ the concept of semi-continuity. Informally speaking, a function is lower (resp. upper) semi-continuous at x_0 if, for values close to x_0, the function is either close to $f(x_0)$ or greater than $f(x_0)$ (resp. smaller than $f(x_0)$).

Lower semi-continuity of V_g is straightforward from the following proposition:

Proposition 5. *If f is the supremum of a family of continuous functions then it is lower semi-continuous.*

Proof. Let $\mathcal{F}$ be a set of continuous functions and let $f(x) = \sup_{f' \in \mathcal{F}} f'(x)$. We show that f is lower semi-continuous.

Fix some $\alpha \in \mathbb{R}$. We need to show that $A = \{x \mid f(x) > \alpha\}$ is open. Let $x_0 \in A$; we are going to show that there exists a ball around x_0 contained in A. Since $\alpha < \sup_{f' \in \mathcal{F}} f'(x_0)$, there exists some $f' \in \mathcal{F}$ such that $f'(x_0) > \alpha$. Since f' is continuous, there exists some ball $B_\epsilon(x_0)$ such that $f'(x) > \alpha$ for all $x \in B_\epsilon(x_0)$. Hence $f(x) \geq f'(x) > \alpha$ for all $x \in B_\epsilon(x_0)$ which means that $B_\epsilon(x_0) \subseteq A$. □

On the other hand upper semi-continuity is much less straightforward. For this we appeal to the structure of the probability simplex and the Gale–Klee–Rockafellar theorem.

Theorem 6 *(Gale–Klee–Rockafellar). If $f : \mathcal{A} \to \mathbb{R}$ is convex and $\mathcal{A}$ is a polyhedron then f is upper semi-continuous.*

Note that the above theorem crucially requires f to be finite-valued, which the reason why the class $\mathbb{G}\mathcal{X}$ (2) enforces the finiteness of V_g. Hence, for $g : \mathbb{G}\mathcal{X}$, V_g is both lower semi-continuous (Proposition 5) and upper semi-continuous (Theorem 6), and it is trivial that any function satisfying both semi-continuities is (simply) continuous.

In fact, we can also show that V_g is Lipschitz-continuous, a property stronger than continuity. Given a metric d on $\mathbb{D}\mathcal{X}$, a function $F : \mathbb{D}\mathcal{X} \to \mathbb{R}$ is $k{\cdot}d$-Lipschitz iff

$$|F(\pi) - F(\pi)'| \quad \leq \quad k \cdot d(\pi, \pi') \qquad \text{for all } \pi, \pi' : \mathbb{D}\mathcal{X} .$$

Intuitively, this property limits the *steepness* of F. Changing π by at most ϵ (w.r.t. d) produces a change in $F(\pi)$ of at most $k \cdot \epsilon$.

Our metric of choice for $\mathbb{D}\mathcal{X}$ is the *total variation* distance, given by

$$\mathrm{tv}(\pi, \pi') \quad := \quad \sup_{X \subseteq \mathcal{X}} |\pi(X) - \pi'(X)| .$$

For discrete distributions, expressed as vectors, the total variation is equal to half of the Manhattan distance, that is $\mathrm{tv} = \frac{1}{2} d_{\|\cdot\|_1}$.

Finally, the *span* $\|g\|$ of a gain function g is defined as:

$$\|g\| \quad := \quad \sup_{w,x,x'} |g(w,x) - g(w,x')|$$

We are now ready to state our soundness result.

Theorem 7 *(Soundness). For any $g : \mathbb{G}\mathcal{X}$, V_g satisfies CNTY and CVX. Moreover it is $\|g\|{\cdot}\mathrm{tv}$-Lipschitz.*

Proof. Continuity is a direct corollary of Proposition 5 and Theorem 6, we here show the Lipschitz property. Viewing an action w as a vector, we first show that $\pi \mapsto w \cdot \pi$ is $\|g\|{\cdot}\mathrm{tv}$-Lipschitz. Let $\bot = \min_x w_x$ and $\Delta = \|g(w, \cdot)\|$. The elements of w lie in $[\bot, \bot + \Delta]$; if we shift it by $c = \bot + \Delta/2$ then we get the vector $w - c\mathbf{1}$ whose elements lie in $[-\Delta/2, \Delta/2]$, hence $\|w - c\mathbf{1}\|_\infty = \Delta/2$. We then reason

$$\begin{array}{lll}
 & |w \cdot \pi - w \cdot \pi'| & \\
= & |(w - c\mathbf{1}) \cdot (\pi - \pi')| & \text{“}c\mathbf{1} \cdot (\pi - \pi') = c - c = 0\text{”} \\
\leq & \|w - c\mathbf{1}\|_\infty \, \|\pi - \pi'\|_1 & \text{“(7), the dual of } \|\cdot\|_1 \text{ is } \|\cdot\|_\infty\text{”} \\
= & \|w - c\mathbf{1}\|_\infty \, 2\,\mathrm{tv}(\pi, \pi') & \text{“}\mathrm{tv}(\pi, \pi') = 1/2 \|\pi - \pi'\|_1\text{”} \\
\leq & \|g\|{\cdot}\mathrm{tv}(\pi, \pi') . & \text{“}2\|w - c\mathbf{1}\|_\infty = \Delta = \|g(w, \cdot)\| \leq \|g\|\text{”}
\end{array}$$

Finally, $V_g(\pi) = \sup_w w \cdot \pi$, and it remains to show that the supremum preserves the d-Lipschitz property.

So let F be a set of d-Lipschitz functions, we show that $F(a) = \sup_{f : F} f(a)$ is also d-Lipschitz. Fixing $a, a' : \mathcal{A}$, assume w.l.o.g. that $F(a) \geq F(a')$; we have

$$\begin{array}{lll}
 & |F(a) - F(a')| & \\
= & \sup_f f(a) - \sup_f f(a') & \text{“definition } F\text{; } F(a) \geq F(a')\text{”} \\
\leq & \sup_f (f(a) - f(a')) & \text{“}-\sup_{f'} f'(a') \leq -f(a')\text{”}
\end{array}$$

$$
\begin{array}{lll}
\leq & \sup_f |f(a) - f(a')| & \\
\leq & \sup_f d(a, a') & \text{“}f \text{ is } d\text{-Lipschitz”} \\
= & d(a, a') \ . \qquad \square &
\end{array}
$$

Note that, without the “finite-valued” restriction of $\mathbb{G}\mathcal{X}$, V_g might attain $+\infty$ and be discontinuous. As an example, take $\mathcal{X} = \{x_1, x_2\}$, $\mathcal{W} = \mathbb{N}$, $g(w, x_1) = w$ and $g(w, x_2) = -w$. For this g we compute that $V_g(\pi^u) = 0$ for the uniform prior, and $V_g(\pi) = +\infty$ for any other π.

4.1.3. Continuity and convexity exactly characterize V_g

Gain functions and g-vulnerability were introduced in order to capture a variety of operational scenarios. As discussed in Section 2, we can naturally retrieve a variety of well-known measures—Bayes-vulnerability, Shannon-entropy, guessing entropy, etc.—using properly constructed gain functions. This suggests the question of how expressive g-vulnerabilities are in general. Remarkably, it turns out that g-vulnerabilities are expressive enough to capture any vulnerability function $\mathbb{V}$ satisfying CNTY and CVX, although in the general case a countably infinite set $\mathcal{W}$ of guesses might be needed.

The geometric view of g, together with the following result from convex analysis, are fundamental for establishing the completeness of g-vulnerability.

Theorem 8. *Let $f : S \to \mathbb{R}$ be convex and let $x \in \mathit{relint}(S)$. Then $\partial f(x) \neq \emptyset$. Moreover, if f is $k \cdot d_{\|\cdot\|}$-Lipschitz then $\|\phi\|_* \leq k$ for all $\phi \in \partial f(x)$.*

Proof. The part $\partial f(x^*) \neq \emptyset$ comes from [15, Theorem 23.4]. The part $\|\phi\|_* \leq k$ comes from [16, Lemma 2.6]. $\square$

We are now ready to state our completeness result. The main idea is that, on each full support prior π^*, we will use a subgradient $\phi \in \partial\mathbb{V}(\pi^*)$ to form an action vector w, such that $w \cdot \pi$ is below $\mathbb{V}$ and touches it on π^*. The supremum of all such function $w \cdot \pi$ has to coincide with $\mathbb{V}$. Note that this result is a variant of the well-known fact that convex functions can be expressed as the sup of linear ones (which has, indeed, already been explored by Boreale and Pampaloni in the study of metrics for QIF [17,18]); our proof also establishes the Lipschitz property and the fact that countably many actions are sufficient.

Theorem 9 *(Completeness). Let $\mathbb{V} : \mathbb{D}\mathcal{X} \to \mathbb{R}_{\geq 0}$ satisfy CNTY and CVX. Then $\mathbb{V} = V_g$ for some $g : \mathbb{G}\mathcal{X}$ with a countable set of actions. Moreover, if $\mathbb{V}$ is $k \cdot$ tv-Lipschitz then $\|g\| \leq k$.*

Proof. Let A be the elements of $\mathit{relint}(\mathbb{D}\mathcal{X})$ (i.e. the full support priors) having *rational* coordinates. We will create one action vector w_π for each such $\pi \in A$. So fix some $\pi^* \in A$, since $\mathbb{V}$ is convex and $\pi^* \in \mathit{relint}(\mathbb{D}\mathcal{X})$, Theorem 8 guarantees that $\partial\mathbb{V}(\pi^*)$ is not empty. Fixing some subgradient $\phi \in \partial\mathbb{V}(\pi^*)$, define an action vector

$$w_{\pi^*} \ := \ \phi + \big(\mathbb{V}(\pi^*) - \phi \cdot \pi^*\big)\mathbf{1} \ ,$$

where $\mathbf{1}$ is the “all-ones” vector. Note that $\pi \cdot \mathbf{1} = 1$, hence we have that

$$w_{\pi^*} \cdot \pi \ = \ \phi \cdot (\pi - \pi^*) + \mathbb{V}(\pi^*) \ ,$$

the function that (from the definition of the subgradient (6)) is always below $\mathbb{V}$ and touches it on π^*. In other words, it holds that

$$
\begin{array}{lll}
w_{\pi^*} \cdot \pi^* & = & \mathbb{V}(\pi^*) \ , \qquad \text{and} \\
w_{\pi^*} \cdot \pi & \leq & \mathbb{V}(\pi) \qquad \text{for all } \pi : \mathbb{D}\mathcal{X} \ .
\end{array}
$$

Setting $\mathcal{W} = \{w_{\pi'} \mid \pi' : A\}$ (a countable set), we have that for all $\pi \in A$

$$V_g(\pi) \ = \ \sup_{\pi' : A} w_{\pi'} \cdot \pi \ = \ w_\pi \cdot \pi \ = \ \mathbb{V}(\pi) \ ,$$

that is, V_g and $\mathbb{V}$ coincide on A. Note that, although we have not yet established that $V_g = \mathbb{V}$ everywhere, we already know that $V_g(\pi) \leq \mathbb{V}(\pi)$ for all $\pi \in \mathbb{D}\mathcal{X}$, hence V_g is finite-valued, so from Proposition 5 and Theorem 6 we conclude that it is continuous.

As a consequence, since all irrationals are the limit of a sequence of rationals, from continuity we get that V_g and $\mathbb{V}$ coincide on the whole $\mathit{relint}(\mathbb{D}\mathcal{X})$. Similarly, boundary points are the limit of a sequence of interior points, hence we conclude that V_g and $\mathbb{V}$ coincide everywhere. This also clearly means that the constructed gain function belongs to $\mathbb{G}\mathcal{X}$.

Finally, assume that $\mathbb{V}$ is $k \cdot \mathrm{tv}$-Lipschitz. Since $\mathrm{tv}(\pi, \pi') = 1/2\|\pi - \pi'\|_1$, $\mathbb{V}$ is $k/2 \cdot d_{\|\cdot\|_1}$-Lipschitz. From Theorem 8, and the fact that the dual norm of $\|\cdot\|_1$ is $\|\cdot\|_\infty$, we get that for all subgradients ϕ used in the construction of the action vectors w_π, it holds that $\|\phi\|_\infty \leq k/2$. As a consequence

$$|g(w_\pi, x) - g(w_\pi, x')| \;=\; |\phi_x - \phi_{x'}| \;\leq\; |\phi_x| + |\phi_{x'}| \;\leq\; 2\|\phi\|_\infty \;\leq\; k\,,$$

from which we conclude that $\|g\| \leq k$. □

Note that a consequence of the construction in the above proof is that, for any full-support prior π with rational probabilities, the supremum in $V_g(\pi)$ is in fact attained by the constructed action w_π. If we are not interested in $\mathcal{W}$ being countable, we can extend the construction of w_π to the whole relint($\mathbb{D}\mathcal{X}$), hence the supremum will be attainable for all full-support priors. This property fails for non-full-support priors however, since subgradients are not guaranteed there; one would need to use gain functions with explicit $-\infty$ values to make the supremum attainable.

5. Axiomatization of posterior vulnerabilities

We will now consider axioms for posterior vulnerabilities, and axioms that relate posterior- and prior vulnerabilities. We consider three of them, and investigate how different definitions of posterior vulnerabilities shape the interrelation among those axioms.

The first property we consider states that an adversary who has learned with certainty, after observing the output of a channel, that the secret has distribution π will have the same amount of information $\mathbb{V}(\pi)$ she would have had from the prior distribution π itself.

This is formalized as the following axiom.

Definition 10 *(Axiom of non-interference (`NI`)).* The vulnerability of a point-hyper equals the vulnerability of the unique inner of this hyper:

$$\forall \pi\colon \qquad \widehat{\mathbb{V}}[\pi] \;=\; \mathbb{V}(\pi)\,.$$

As its name suggests, the `NI` axiom can be interpreted in terms of noninterference. A channel C is *noninterfering* if the result of pushing any prior π through C is the point-hyper $[\pi]$, meaning that the adversary's state of knowledge is not changed by the observation of the output of the channel; that is, C is the channel $\mathbb{1}$ that leaks nothing. It is well known that a channel matrix C is noninterfering iff all its rows are the same [19,20]. The `NI` axiom, then, is equivalent to stating that an adversary observing the output of a noninterfering channel does not gain or lose any information about the secret:

$$\forall \pi\colon \qquad \widehat{\mathbb{V}}[\pi \triangleright \mathbb{1}] \;=\; \mathbb{V}(\pi)\,.$$

The second axiom we consider is an analogue of the famous data-processing inequality for mutual information,[9] and is formalized as follows.

Definition 11 *(Axiom of data-processing inequality (`DPI`)).* Post-processing does not increase vulnerability:

$$\forall \pi, \mathsf{C}, \mathsf{R}\colon \qquad \widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}] \;\geq\; \widehat{\mathbb{V}}[\pi \triangleright \mathsf{CR}]\,,$$

where the number of columns in matrix C is the same as the number of rows in matrix R, and CR is the standard matrix multiplication, here called *cascading* of channels.

Note that, since the `DPI` axiom concerns the operation of cascading—which demands a matching between the outputs of the first channel matrix and the inputs of the second one—, it is formalized in terms of (concrete) channels. The `DPI` axiom can be interpreted as follows. Consider a secret that is fed into a (concrete) channel C, and then the produced output is post-processed by being fed into another (concrete) channel R (whose input set must be the same as the output set of C). Now consider two adversaries A and A' such that A can only observe the output of channel C, and A' can only observe output of the cascading $\mathsf{C}' = \mathsf{CR}$. For any given prior π on secret values, A's posterior knowledge about the secret is given by the hyper $[\pi \triangleright \mathsf{C}]$, whereas that of A''s is given by $[\pi \triangleright \mathsf{C}']$. Note, however, that from A's knowledge it is always possible to reconstruct A''s, but the converse is not necessarily true. To see that, note that A can use π and C to compute $[\pi \triangleright \mathsf{CR}']$ for any R', including the particular R used by A'. On the other hand, A' knows only π and C' and, in general, the decomposition of C' into a cascade of two channels is not unique (i.e. there may be several pairs C_i, R_i of matrices satisfying $\mathsf{C}' = \mathsf{C}_i\mathsf{R}_i$), so it is not always possible for A' to uniquely recover C from C' and compute $[\pi \triangleright \mathsf{C}]$. Given this asymmetry, `DPI` formalizes that a vulnerability $\widehat{\mathbb{V}}$ should not evaluate A's information as any less than A''s.

[9] The data processing-inequality for mutual information states that if $X \to Y \to Z$ forms a Markov chain, then $I(X;Y) \geq I(X;Z)$.

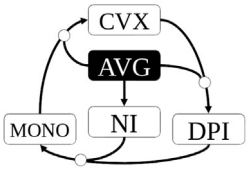

Fig. 2. Equivalence of axioms under AVG. The merging arrows indicate joint implication: for example, AVG ∧ CVX implies DPI.

The third property we consider is that by observing the output of a channel an adversary cannot lose information about the secret; in the worst case, the output can be ignored if it is not useful.[10] This property is formalized as the following axiom.

Definition 12 *(Axiom of monotonicity (MONO)).* Pushing a prior through a channel does not decrease vulnerability:

$$\forall \pi, C: \qquad \widehat{\mathbb{V}}[\pi \triangleright C] \geq \mathbb{V}(\pi)\,.$$

The MONO axiom has two direct consequences on the additive and multiplicative notions of leakage. Since posterior vulnerabilities are never smaller than the corresponding prior vulnerabilities, additive leakage is always non-negative, and multiplicative leakage is never smaller than 1.

5.1. Definitions of posterior vulnerabilities

Having introduced the axioms of NI, DPI and MONO, we now turn our attention to how posterior vulnerabilities can be defined so to respect them. In contrast with the case of prior vulnerabilities, in which the axioms considered (CVX and CNTY) were satisfied by, and only by, the family of prior g-vulnerabilities, in the case of posterior vulnerability the axioms considered so far are not satisfied by, and only by, the family of posterior g-vulnerabilities. For that reason, in the following we shall consider alternative definitions of posterior vulnerabilities, and discuss the interrelations of axioms each of them induces.

5.1.1. Posterior vulnerability as expectation

As we have seen, the posterior versions of Bayes vulnerability and g-vulnerability, as well as of Shannon entropy and guessing entropy, are all defined as the expectation of the corresponding prior measures over the (hyper-) distribution of posterior distribution, i.e. weighted by the probability of each posterior's being realized. The definition of posterior vulnerability as expectation is formalized as the following axiom.

Definition 13 *(Axiom of averaging (AVG)).* The vulnerability of a hyper is the expected value, w.r.t. the outer distribution, of the vulnerabilities of its inners:

$$\forall \Delta: \qquad \widehat{\mathbb{V}}\Delta \;=\; \mathcal{E}_{\Delta}\mathbb{V}\,,$$

where the hyper $\Delta : \mathbb{D}\mathcal{X}$ typically results from $\Delta = [\pi \triangleright C]$ for some π, C.

We will now consider the consequences of taking AVG as an axiom. As it turns out, by imposing AVG on a prior/posterior pair $(\mathbb{V}, \widehat{\mathbb{V}})$ of vulnerabilities, we can uncover a series of interesting relations among other axioms: if axiom of AVG holds, then so does the axiom of NI; and the axioms of CVX, DPI and MONO become equivalent to each other. Fig. 2 summarizes these relations, which we shall now demonstrate.

We begin by showing that AVG implies NI.

Proposition 14 *(AVG ⇒ NI). If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies AVG, then it also satisfies NI.*

Proof. If AVG is assumed then for any prior π we have $\widehat{\mathbb{V}}[\pi] = \mathcal{E}_{[\pi]}\mathbb{V} = \mathbb{V}(\pi)$, since $[\pi]$ is a point-hyper. □

Second, we show that the axioms of NI and DPI, taken together, imply MONO.

[10] Note that in this paper we adopt a *static* perspective of leakage, which considers the entire hyper-distribution $[\pi \triangleright C]$. Since this approach lets us consider *all* the possible posterior distributions that the adversary might learn, together with their probabilities, it is reasonable to assume that the adversary's information cannot decrease by observing the system. If, instead, we were to adopt a *dynamic* perspective of leakage, which considers the effect of the adversary's observing a *particular* channel output y, it might be reasonable to accept that some particular outputs can decrease the adversary's information about the secret.

Proposition 15 *(NI ∧ DPI ⇒ MONO). If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies NI and DPI, then it also satisfies MONO.*

Proof. For any π, C, let $\mathbb{1}$ denote the noninterfering channel with only one column and as many rows as the columns of C. Then

$$\begin{array}{lll} & \widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}] & \\ \geq & \widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}\mathbb{1}] & \text{“by DPI”} \\ = & \widehat{\mathbb{V}}[\pi \triangleright \mathbb{1}] & \text{“C}\mathbb{1} = \mathbb{1}\text{”} \\ = & \widehat{\mathbb{V}}[\pi] & \\ = & \mathbb{V}(\pi) & \text{“by NI”} \end{array}$$

□

Third, we show that the axioms AVG and MONO together imply CVX.

Proposition 16 *(AVG ∧ MONO ⇒ CVX). If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies AVG and MONO, then it also satisfies CVX.*

Proof. Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be a finite set, and let π^1 and π^2 be distributions over $\mathcal{X}$. Let $0<a<1$, so that also $\pi^3 = a\pi^1 + (1-a)\pi^2$ is a distribution on $\mathcal{X}$. (In case $a = 0$ or $a = 1$ we would have $\pi^3 = \pi^1$ or $\pi^3 = \pi^2$, respectively, and convexity would follow trivially.) Define C^* to be the two-column channel matrix

$$\mathsf{C}^* \quad = \quad \begin{bmatrix} a\pi^1_1/\pi^3_1 & (1-a)\pi^2_1/\pi^3_1 \\ \vdots & \vdots \\ a\pi^1_i/\pi^3_n & (1-a)\pi^2_i/\pi^3_n \\ \vdots & \vdots \\ a\pi^1_n/\pi^3_N & (1-a)\pi^2_n/\pi^3_N \end{bmatrix} \tag{8}$$

for every i such that $\pi^3_i \neq 0$. (Note that if $\pi^3_i = 0$ for some i, then x_i is not in the support neither of π^1 nor of π^2 (since $0<a<1$), and we can, without loss of generality, remove element x_i from both priors and from channel matrix C^* above.)

By pushing π^3 through C^* we obtain the hyper $[\pi^3 \triangleright \mathsf{C}^*]$ with outer distribution $(a, 1-a)$, and associated inners π^1 and π^2. Since AVG is assumed, we have

$$\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}^*] \quad = \quad a\mathbb{V}(\pi^1) + (1-a)\mathbb{V}(\pi^2)\,. \tag{9}$$

But note that by MONO, we also have

$$\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}^*] \;\geq\; \mathbb{V}(\pi^3) \;=\; \mathbb{V}(a\pi^1 + (1-a)\pi^2)\,. \tag{10}$$

Taking (9) and (10) together, we obtain CVX. □

Finally, we show that the axioms AVG and CVX together imply DPI. For that, we will need the following lemma.

Lemma 17. *Let $X \to Y \to Z$ form a Markov chain with triply joint distribution $p(x, y, z) = p(x)p(y|x)p(z|y)$ for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Then $\sum_y p(y|z)p(x|y) = p(x|z)$ for all x, y, z.*

Proof. First we note that the probability of z depends only on the probability of y, and not x, so $p(z|x, y) = p(z|y)$ for all x, y, z. Then we can use the fact that

$$p(y, z)p(x, y) \quad = \quad p(x, y, z)p(y) \tag{11}$$

to reason

$$\begin{array}{lll} & \sum_y p(y|z)p(x|y) & \\ = & \sum_y (p(y,z)/p(z))(p(x,y)/p(y)) & \text{“by definition of conditional”} \\ = & \sum_y p(x,y,z)p(y)/p(z)p(y) & \text{“by Equation (11)”} \\ = & \sum_y p(x, y \mid z) & \text{“by definition of conditional”} \\ = & p(x|z)\,. & \text{“by marginalization”} \end{array}$$

□

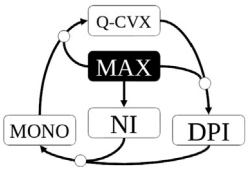

Fig. 3. Equivalence of axioms under MAX. The merging arrows indicate joint implication: for example, MAX ∧ Q-CVX implies DPI. Compare with Fig. 2.

Proposition 18 *(AVG ∧ CVX ⇒ DPI). If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies AVG and CVX, then it also satisfies DPI.*

Proof. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be sets of values. Let be π be a prior on $\mathcal{X}$, C be a (concrete) channel from $\mathcal{X}$ to $\mathcal{Y}$, and R be a (concrete) channel from $\mathcal{Y}$ to $\mathcal{Z}$. Note that the cascade CR of channels C and R is a channel from $\mathcal{X}$ to $\mathcal{Z}$.

Let $p(x, y, z)$ be the triply joint distribution defined $p(x, y, z) = \pi_x \mathsf{C}_{x,y} \mathsf{R}_{y,z}$ for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. By construction, this distribution has the property that the probability of z depends only on the probability of y, and not x, that is that $p(z \mid x, y) = p(z|y)$.

Note that, by pushing prior π through channel C, we obtain hyper $[\pi \triangleright \mathsf{C}]$, in which the outer distribution on y is $p(y)$, and the inners are $p_{X|y}$. Thus we can reason

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}] & \\
= & \sum_y p(y) \mathbb{V}(p_{X|y}) & \text{“by AVG”} \\
= & \sum_y \left(\sum_z p(z) p(y|z) \right) \mathbb{V}(p_{X|y}) & \text{“by marginalization”} \\
= & \sum_z p(z) \sum_y p(y|z) \mathbb{V}(p_{X|y}) & \text{“moving constants w.r.t. the sum”} \\
\geq & \sum_z p(z) \mathbb{V}\left(\sum_y p(y|z) p_{X|y} \right) & \text{“by CVX”} \\
= & \sum_z p(z) \mathbb{V}(p_{X|z}) & \text{“by Lemma 17”} \\
= & \widehat{\mathbb{V}}[\pi \triangleright \mathsf{CR}] \ . & \text{“by AVG”}
\end{array}
$$

□

5.1.2. Posterior vulnerability as maximum

An important consequence of AVG is that an observable's happening with very small probability will have a negligible effect on $\widehat{\mathbb{V}}$, even if it completely reveals the secret. If this is not acceptable, an alternative approach is to consider the maximum information that can be obtained from any single output of the channel—produced with non-zero probability—no matter how small that probability might be. This conservative approach represents a defender who worries about the worst possible amount of threat to the secret. The definition of posterior vulnerability in these terms is formalized as the following axiom.

Definition 19 *(Axiom of maximum (MAX)).* The vulnerability of a hyper is the maximum value among the vulnerabilities the inners in the support of its outer distribution:

$$\forall \Delta: \qquad \widehat{\mathbb{V}} \Delta \ = \ \max_{\lceil \Delta \rceil} \mathbb{V} \ ,$$

where the hyper $\Delta : \mathbb{D}\mathcal{X}$ typically results from $\Delta = [\pi \triangleright C]$ for some π, C.

Note that the definition above takes the support of the outer distribution because we ignore the vulnerability of inners that cannot happen (i.e. that have probability zero).

We shall now consider the consequences of taking MAX as an axiom. As it turns out, by imposing MAX on a prior/posterior pair $(\mathbb{V}, \widehat{\mathbb{V}})$ of vulnerabilities, we can derive relations among other axioms, just as we did for AVG. But they are different.

More precisely, if the axiom of MAX is satisfied, then again the axiom of NI too is implied; but this time it's the axioms of Q-CVX, DPI and MONO that become equivalent to each other. Fig. 3 summarizes these relations, which we shall now demonstrate.

We begin by showing that MAX implies NI.

Proposition 20 *(MAX ⇒ NI). If a pair of prior/posterior vulnerabilities* $(\mathbb{V}, \widehat{\mathbb{V}})$ *satisfies MAX, then it also satisfies NI.*

Proof. If the MAX axiom is assumed, for any prior π we will have $\widehat{\mathbb{V}}[\pi] = \max_{\lceil [\pi] \rceil} \mathbb{V} = \mathbb{V}(\pi)$, since $[\pi]$ is a point-hyper. □

However, in contrast to the case of AVG, the symmetry among CVX, MONO and DPI is broken under MAX: although the axioms of MONO and DPI are still equivalent (a result that we shall soon demonstrate), they are weaker than the axiom of CVX. Indeed, the following example shows a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfying the axioms of MAX, MONO and DPI, but not the axiom of CVX.

Example 21 *(MAX ∧ MONO ∧ DPI ⇏ CVX).* Consider the pair $(\mathbb{V}_1, \widehat{\mathbb{V}}_1)$ such that for every prior π and channel C:

$$\mathbb{V}_1(\pi) = 1 - \left(\min_x \pi_x\right)^2, \qquad \text{and}$$
$$\widehat{\mathbb{V}}_1[\pi \triangleright C] = \max_{\lceil[\pi \triangleright C]\rceil} \mathbb{V}_1.$$

(Note that $[\pi \triangleright C]$ denotes the *set* of inners in the hyper resulting from the application of C to π.)

We can see that $\mathbb{V}_1$ does not satisfy CVX by considering distributions $\pi^1 = (0, 1)$ and $\pi^2 = (1/2, 1/2)$, and their convex combination $\pi^3 = 1/2\,\pi^1 + 1/2\,\pi^2 = (1/4, 3/4)$. We then calculate $\mathbb{V}_1(\pi^1) = 1-0^2 = 1$, $\mathbb{V}_1(\pi^2) = 1-(1/2)^2 = 3/4$, $\mathbb{V}_1(\pi^3) = 1-(1/4)^2 = 15/16$, and $1/2\,\mathbb{V}_1(\pi^1) + 1/2\,\mathbb{V}(\pi^2) = 7/8$ to conclude that $\mathbb{V}_1(\pi^3) > 1/2\,\mathbb{V}_1(\pi^1) + 1/2\,\mathbb{V}_1(\pi^2)$ so that indeed CVX is not satisfied.

The pair $(\mathbb{V}_1, \widehat{\mathbb{V}}_1)$ satisfies MAX by construction. To show that $(\mathbb{V}_1, \widehat{\mathbb{V}}_1)$ satisfies MONO and DPI, we first notice that $\mathbb{V}_1$ is quasi-convex. Using results from Fig. 3 (more precisely, that MAX + Q-CVX ⇒ MONO + DPI, proved later in this section), we conclude that MONO and DPI are also satisfied.

The vulnerability function used in the counter-example above is quasi-convex. It turns out that this is not a coincidence: by replacing CVX with Q-CVX (a weaker property), the symmetry between the axioms can be restored. The remainder of this section establishes the equivalence of Q-CVX, MONO and DPI under MAX.

We first show that the axioms MAX and MONO together imply Q-CVX.

Proposition 22 *(MAX ∧ MONO ⇒ Q-CVX). If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies MAX and MONO, then it also satisfies Q-CVX.*

Proof. Assume for a contradiction that $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfy MAX and MONO, but do not satisfy Q-CVX.

Since Q-CVX is not satisfied, there must exist a value $0 \leq a \leq 1$ and three distributions π^1, π^2, π^3, such that $\pi^3 = a\pi^1 + (1-a)\pi^2$ and

$$\mathbb{V}(\pi^3) \;>\; \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right). \tag{12}$$

Now consider the (concrete) channel C^* defined as in (8). Then the hyper-distribution $[\pi^3 \triangleright \mathsf{C}^*]$ has outer distribution $(a, 1-a)$, and corresponding inner distributions π^1 and π^2. Since MAX is assumed, we have that

$$\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}^*] \;=\; \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right), \tag{13}$$

and because we assumed MONO, we also have that

$$\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}^*] \;\geq\; \mathbb{V}(\pi^3). \tag{14}$$

By replacing (13) in (14), we derive that $\mathbb{V}(\pi^3) \leq \max\left(\mathbb{V}(\pi^1), \mathbb{V}(\pi^2)\right)$, which contradicts our assumption in (12). □

We now show that the axioms MAX and Q-CVX together imply DPI.

Proposition 23 *(MAX ∧ Q-CVX ⇒ DPI). If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies MAX and Q-CVX, then it also satisfies DPI.*

Proof. Let π be a prior on $\mathcal{X}$, and C, R be channels from $\mathcal{X}$ to $\mathcal{Y}$ and from $\mathcal{Y}$ to $\mathcal{Z}$, respectively, with joint distribution $p(x, y, z)$ defined in the same way as in the proof of Proposition 18.

Note that, by pushing prior π through channel CR, we obtain hyper $[\pi \triangleright \mathsf{CR}]$ in which the outer distribution on z is $p(z)$, and the inners are $p_{X|z}$. Thus we can derive:

$$\begin{array}{lll}
& \widehat{\mathbb{V}}[\pi \triangleright \mathsf{CR}] & \\
= & \max_z \mathbb{V}(p_{X|z}) & \text{“by MAX”} \\
= & \max_z \mathbb{V}\left(\sum_y p(y|z)\, p_{X|y}\right) & \text{“by Lemma 17”} \\
\leq & \max_z \left(\max_y \mathbb{V}(p_{X|y})\right) & \text{“by Q-CVX”} \\
= & \max_y \mathbb{V}(p_{X|y}) & \text{“z not free”} \\
= & \widehat{\mathbb{V}}[\pi \triangleright \mathsf{C}]\,. & \text{“by MAX”}
\end{array}$$

□

Finally, note that, although Q-CVX is needed to recover the full equivalence of the axioms, CVX is strictly stronger than Q-CVX; hence, using a convex vulnerability measure (such as any V_g), MONO and DPI are still guaranteed under MAX.

Corollary 24 *(MAX ∧ CVX ⇒ MONO ∧ DPI). If a pair $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies MAX and CVX, then it also satisfies MONO and DPI.*

Proof. Follows from the results of Fig. 3 and the fact that CVX ⇒ Q-CVX. □

5.1.3. Other definitions of posterior vulnerabilities

Defining posterior vulnerabilities using the axioms of AVG or MAX is certainly reasonable in many scenarios, but we may wonder whether other definitions might also be meaningful in some context. In this section we discuss the consequences of defining posterior vulnerabilities with something more relaxed than AVG or MAX. We shall, however, consider that $\widehat{\mathbb{V}}$ should be related to $\mathbb{V}$ by the following condition: the vulnerability of a hyper-distribution should be bounded by the vulnerabilities of the inner distributions in its support. The next axiom formalizes this restriction.

Definition 25 *(Axiom of bounds (BNDS)).* The vulnerability of a hyper lies non-strictly between the vulnerabilities of the inners in its support:

$$\forall \Delta: \qquad \min_{\lceil \Delta \rceil} \mathbb{V} \;\leq\; \widehat{\mathbb{V}}\Delta \;\leq\; \max_{\lceil \Delta \rceil} \mathbb{V}\,,$$

where the hyper $\Delta : \mathbb{D}\mathcal{X}$ might result from $\Delta = [\pi \triangleright C]$ for some π, C.

Intuitively, the BNDS axiom states that the vulnerability of a hyper resulting from pushing a prior through a channel can only be as high as the vulnerability induced by the "most leaky" observable, and it can only be as low as the vulnerability induced by the "least leaky" observable.

We shall now consider the consequences of taking BNDS as an axiom. Whereas BNDS turns out to be strong enough to ensure NI, by replacing MAX with BNDS, the equivalence among Q-CVX, DPI and MONO no longer holds.

We begin by showing that the axiom of BNDS is sufficient to guarantee NI.

Proposition 26 *(BNDS ⇒ NI). If a pair of prior/posterior vulnerabilities $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies BNDS, then it also satisfies NI.*

Proof. If $(\mathbb{V}, \widehat{\mathbb{V}})$ satisfies BNDS, then $\min_{\lceil \Delta \rceil} \mathbb{V} \leq \widehat{\mathbb{V}}\Delta \leq \max_{\lceil \Delta \rceil} \mathbb{V}$ for every hyper Δ. Consider, then, the particular case when $\Delta = [\pi]$. Since $[\pi]$ is a point-hyper with inner π, we have that $\min_{\lceil [\pi] \rceil} \mathbb{V} = \max_{\lceil [\pi] \rceil} \mathbb{V} = \mathbb{V}(\pi)$. This in turn implies that $\mathbb{V}(\pi) \leq \widehat{\mathbb{V}}[\pi] \leq \mathbb{V}(\pi)$, which is NI. □

The next example shows that under BNDS, not even CVX—which is stronger than Q-CVX—is sufficient to ensure MONO or DPI.

Example 27 *(BNDS ∧ CVX ⇏ MONO and BNDS ∧ CVX ⇏ DPI).* Consider the pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ such that for every prior π and hyper Δ:

$$\begin{aligned} \mathbb{V}_2(\pi) &= \max_x \pi_x, && \text{and} \\ \widehat{\mathbb{V}}_2\Delta &= \frac{(\max_{\lceil \Delta \rceil} \mathbb{V}_2 + \min_{\lceil \Delta \rceil} \mathbb{V}_2)}{2}. \end{aligned}$$

The pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ satisfies the axiom of BNDS, since $\widehat{\mathbb{V}}_2$ is the simple arithmetic average of maximum and minimum vulnerabilities of the inners. The pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ also satisfies the axiom of CVX, since $\mathbb{V}_2(\pi)$ is just the Bayes vulnerability of π.

To see that the pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ does not satisfy the axiom of MONO, consider the prior $\pi^2 = (9/10, 1/10)$ and the (concrete) channel

$$\mathsf{C}_2 = \begin{array}{|c|c|}\hline 8/9 & 1/9 \\ \hline 0 & 1 \\ \hline \end{array}\,.$$

We can calculate that $\mathbb{V}_2(\pi^2) = 9/10$, and that $[\pi^2 \triangleright \mathsf{C}_2]$ has outer distribution $(4/5, 1/5)$, and inner distributions $(1, 0)$ and $(1/2, 1/2)$. Hence

$$\widehat{\mathbb{V}}_2[\pi^2 \triangleright \mathsf{C}_2] \;=\; (1+1/2)/2 \;=\; 3/4\,,$$

which violates MONO because $\widehat{\mathbb{V}}_2[\pi^2 \triangleright \mathsf{C}_2] < \mathbb{V}_2(\pi^2)$.

Now to see that the pair $(\mathbb{V}_2, \widehat{\mathbb{V}}_2)$ does not satisfy DPI, consider the prior $\pi^3 = (3/7, 4/7)$ and the (concrete) channels

$$\mathsf{C}_3 = \begin{array}{|c|c|}\hline 1/3 & 2/3 \\ \hline 1/4 & 3/4 \\ \hline \end{array}\,, \quad \text{and} \quad \mathsf{R}_3 = \begin{array}{|c|c|}\hline 1/4 & 3/4 \\ \hline 3/4 & 1/4 \\ \hline \end{array}\,.$$

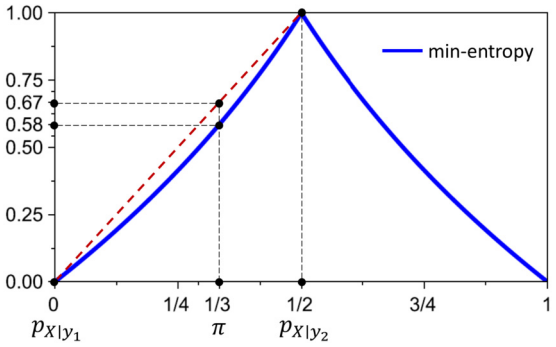

Fig. 4. A picture showing how posterior min-entropy can be greater than prior min-entropy. Pushing prior $\pi = (1/3, 2/3)$ through channel C gives hyper $[\pi \triangleright C]$ with outer $(1/3, 2/3)$ and inners $p_{X|y_1} = (0, 1)$ and $p_{X|y_2} = (1/2, 1/2)$. So $H_\infty(\pi) = -\log_2 2/3 \approx 0.58$ and $\widehat{H}_\infty[\pi \triangleright C] = 1/3 \cdot 0 + 2/3 \cdot 1 \approx 0.67$.

We can calculate that $[\pi^3 \triangleright \mathsf{C}_3]$ has outer distribution $(2/7, 5/7)$, and inners $(1/2, 1/2)$ and $(2/5, 3/5)$. Hence

$$\widehat{\mathbb{V}}_2[\pi^3 \triangleright \mathsf{C}_3] \;=\; {}^{(1/2+3/5)}/_2 \;=\; {}^{11}/_{20} \;=\; 0.55\,.$$

On the other hand, the cascade $\mathsf{C}_3\mathsf{R}_3$ yields the channel

$$\mathsf{C}_3\mathsf{R}_3 \;=\; \begin{array}{|c|c|}\hline {}^7/_{12} & {}^5/_{12} \\ \hline {}^5/_8 & {}^3/_8 \\ \hline \end{array}\,,$$

and we can calculate $[\pi^3 \triangleright \mathsf{C}_3\mathsf{R}_3]$ to have outer distribution $({}^{17}/_{28}, {}^{11}/_{28})$, and inners $({}^7/_{17}, {}^{10}/_{17})$ and $({}^5/_{11}, {}^6/_{11})$. Hence

$$\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}_3\mathsf{R}_3] \;=\; {}^{({}^{10}/_{17}+{}^6/_{11})}/_2 \;=\; {}^{106}/_{187} \;\approx\; 0.567\,,$$

which makes $\widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}_3\mathsf{R}_3] > \widehat{\mathbb{V}}[\pi^3 \triangleright \mathsf{C}_3]$ and violates the axiom of DPI.

6. Applications of axiomatization to understanding leakage measures

The relationships we have uncovered among axioms help us better understand the multitude of possible leakage measures one can adopt (e.g. what V_g to use, and what version of leakage—additive or multiplicative—to employ).

A first instance of an insight concerns the robustness of the *refinement relation* $\sqsubseteq$ discussed studied in [5,6,14]. Given channels C and D, both taking input X, C is refined by D, written $C \sqsubseteq D$, if $D = CR$ for some "refining" channel R. As proved in [5,14], refinement is *sound* and *complete* for the *strong g-leakage ordering*: we have $C \sqsubseteq D$ iff the g-leakage of D never exceeds that of C, regardless of the prior π or gain function g. Still, we might worry that refinement implies a leakage ordering *only* with respect to g-leakage, leaving open the possibility that the leakage ordering might conceivably fail for some yet-to-be-defined leakage measure. But Propositions 16 (AVG $\wedge$ MONO $\Rightarrow$ CVX) and 18 (AVG $\wedge$ CVX $\Rightarrow$ DPI) show that if the hypothetical new leakage measure is defined using AVG, and never gives negative leakage, then it also satisfies the data-processing inequality DPI. And hence refinement is also sound for the new leakage measure.

Another application concerns the possibility of negative leakage for some information measures. As an example, consider Rényi entropy, a family of entropy measures that has been used in the context of quantitative information flow. The family is defined by

$$H_\alpha(\pi) \;=\; \frac{1}{1-\alpha}\log_2\left(\textstyle\sum_{x\in\mathcal{X}} \pi_x^\alpha\right)$$

for $0 \leq \alpha \leq \infty$ (taking limits in the cases of $\alpha = 1$, which gives Shannon entropy, and $\alpha = \infty$, which gives min-entropy). It would be natural to use Rényi entropy to define a family of leakage measures by defining posterior Rényi entropy $\widehat{H}_\alpha$ using AVG and defining Rényi leakage by

$$\mathcal{L}^\alpha(\pi, C) \;=\; H_\alpha(\pi) - \widehat{H}_\alpha[\pi \triangleright C]\,.$$

However, it turns out that H_α is not concave for $\alpha > 2$. Therefore, by the dual version of Proposition 16 (AVG + MONO $\Rightarrow$ CVX), we find that Rényi leakage $\mathcal{L}^\alpha$ for $\alpha > 2$ would sometimes be *negative*. As an illustration, Fig. 4 shows how the nonconcavity of min-entropy H_∞ can cause posterior min-entropy to be greater than prior min-entropy, giving negative min-entropy leakage. This problem is avoided, however, if we do not define posterior min-entropy by using AVG, but instead by $\widehat{H}_\infty[\pi \triangleright C] = -\log_2 \widehat{V}_b[\pi \triangleright C]$.

7. A more abstract perspective

The axioms relating to Averaging are in fact instances of the general monad laws proved by Giry for probabilistic computation [21], and in this section we give details. The benefit of this generality is that it provides immediate access to well developed mathematical theories extending these results to infinite states and proper measures [22]. And its broader perspective gives a direct connection to higher-order reasoning tools that dramatically simplify proofs, thereby leading directly to practical frameworks for calculating leakage [23]. We now give details.

The operator $\mathbb{D}$ that takes a sample space $\mathcal{X}$ to (discrete) distributions $\mathbb{D}\mathcal{X}$ on that space is widely recognized as the "probability monad", that is in effect a type constructor that obeys a small collection of laws shared by other, similar constructors like the powerset operator $\mathbb{P}$ [21]. Each monad has two polymorphic functions η, for "unit", and μ, for "multiply", that interact with each other in elegant ways. For example in (the) $\mathbb{P}$ (monad), unit has type $\mathcal{X}\to\mathbb{P}\mathcal{X}$ and ηx is $\{x\}$, the singleton set containing just x (we sometimes omit function brackets to reduce clutter); correspondingly in $\mathbb{D}$ we have type $\mathcal{X}\to\mathbb{D}\mathcal{X}$ and ηx is $[x]$, the point-distribution on x. In $\mathbb{P}$, multiply μ is distributed union that takes a set of sets to the one set that is the union of them all, having thus the type $\mathbb{P}^2\mathcal{X}\to\mathbb{P}\mathcal{X}$; and in $\mathbb{D}$ we have $(\mu\Delta)_x = \sum_{\delta:\lceil\Delta\rceil}\Delta_\delta\delta_x$, with μ thus of type $\mathbb{D}^2\mathcal{X}\to\mathbb{D}\mathcal{X}$ and taking the outer-weighted average of all the inner distributions δ in the support of hyper Δ: it is the "weighted average" of the hyper, which we sometimes call "squash". That means, for example, that $\mu[\pi\triangleright C]=\pi$, i.e. that if you squash the hyper produced by prior π and channel C you get the prior back again.

Furthermore the monadic type-constructors are *functors*, whose significance for us is that they can be applied to functions as well as to objects: thus for f in $\mathcal{X}\to\mathcal{Y}$ the function $\mathbb{P}f$ of type $\mathbb{P}\mathcal{X}\to\mathbb{P}\mathcal{Y}$ is such that for X in $\mathbb{P}\mathcal{X}$ we have $f(X) = \{f(x) \mid x\in X\}$ in $\mathbb{P}\mathcal{Y}$. In $\mathbb{D}$ instead we get the *push forward* of f, so that for π in $\mathbb{D}\mathcal{X}$ we have $(\mathbb{D}f)(\pi)_y = \sum_{f(x)=y}\pi_x$.

With these tools, some of our axioms can be expressed in a very general way, for example

(1) AVG becomes $\widehat{\mathbb{V}}=\mu\circ\mathbb{D}\mathbb{V}$, the Kleisli lifting or indeed the `bind` of monadic functional programming.
(2) NI becomes $\widehat{\mathbb{V}}\circ\eta=\mathbb{V}$. Assuming (1), that follows from the general monad laws $\mu\circ\eta=1$ and $\mathbb{D}V\circ\eta=\eta\circ V$.
(3) CVX becomes $\mathbb{V}\circ\mu \le \mu\circ\mathbb{D}\mathbb{V}$.

A consequence of accepting averaging (1) is that $\widehat{\mathbb{V}}(\Delta^1{}_p{+}\Delta^2) = \widehat{\mathbb{V}}(\Delta^1)\,{}_p{+}\,\widehat{\mathbb{V}}(\Delta^2)$, i.e. we have linearity of $\widehat{\mathbb{V}}$, where ${}_p+$ takes the p-weighted sum of its operands: on the left we sum over hypers; on the right we sum over scalars. This is more generally $\widehat{\mathbb{V}}(\mu\underline{\Delta}) = \mu((\mathbb{D}\widehat{\mathbb{V}})\underline{\Delta})$ where $\underline{\Delta}$ is in $\mathbb{D}^3\mathcal{X}$, a distribution of hypers, another monad law when $\widehat{\mathbb{V}}=\mu\circ\mathbb{D}\mathbb{V}$.

The space $\mathbb{D}^3\mathcal{X}$ also gives a hyper-formulated definition of the refinement order $\sqsubseteq$ over hypers, i.e., that $\Delta^1\sqsubseteq\Delta^2$ just when $\widehat{\mathbb{V}}\Delta^1\ge\widehat{\mathbb{V}}\Delta^2$ for all $\widehat{\mathbb{V}}$ satisfying the axioms: it is that $\Delta^1\sqsubseteq\Delta^2$ just when there is a $\underline{\Delta}$ such that $\Delta^1=\mu\underline{\Delta}$ and $\Delta^2=(\mathbb{D}\mu)\underline{\Delta}$ [22,24]. This formulation allows soundness of $\sqsubseteq$, i.e. that it can only decrease g-vulnerability, to be shown even for infinite state-spaces $\mathcal{X}$ and general measures. (See Appendix A.)

Finally, the monadic structure coupled with the Kantorovich metric gives us continuity criteria not only for $\mathbb{V}$ but also for $\widehat{\mathbb{V}}$ [21,25]. If we give the underlying $\mathcal{X}$ the discrete metric, that $dist(x_1, x_2) = (0$ if $x_1{=}x_2$ else $1)$, then the Kantorovich-induced distance on $\mathbb{D}\mathcal{X}$ is equivalent to the total variation or $1/2$-Manhattan metric used in Section 4. But the great generality of these monads gives us more, for example that AVG, i.e., the axiom $\widehat{\mathbb{V}} = \mu\circ\mathbb{D}\mathbb{V}$, makes $\widehat{\mathbb{V}}$ continuous as well, this time with respect to the Kantorovich metric on hypers. That in turn allows higher-order calculations that limit information flow in a very robust way [12].

In Appendix A we give examples of these general constructions in action: how they enable more succinct definitions, shorter, more algebraic proofs (if somewhat denser), and how they improve our chances of further discoveries because of the wealth of similar constructions that already exist in this mathematical style.

8. Related work

Our axioms for posterior vulnerabilities can be seen generalizations of key properties of Shannon entropy [26]. As we mentioned, our axiom of data-processing inequality (DPI) is a straightforward generalization of the data-processing inequality for mutual information, which states that if $X\to Y\to Z$ forms a Markov chain, then $I(X;Y)\ge I(X;Z)$. Our axiom of monotonicity (MONO) is a generalization of Shannon entropy's *"information can't hurt"* property, which states that for every pair X, Y of random variables, $H(X|Y)\le H(X)$. Finally, our axiom of monotonicity (NI) can be seen as a generalization of the property that, for every pair of independent random variables X and Y, $H(X\mid Y) = H(X)$.

The conservative approach represented by the use of the axiom of maximum (MAX) is employed, for instance, in the original definition of *differential privacy* [27].

8.1. Relation with other axiomatizations of entropy measures

Csiszár has surveyed [28] the most commonly used postulates for a function f of the uncertainty contained in a finite probability distribution $(p_1,\ldots,p_N)$ for $N{>}0$. They are: (P1) *positivity*: $f(p_1,\ldots,p_N)\ge 0$; (P2) *expansibility*: $f(p_1,\ldots,p_N,0) = f(p_1,\ldots,p_N)$; (P3) *symmetry*: $f(p_1,\ldots,p_N)$ is invariant under permutations of $(p_1,p_2,\ldots,p_N)$; (P4) *continuity*: $f(p_1,\ldots,p_N)$ is a continuous function of $(p_1,\ldots,p_N)$, for fixed n; (P5) *additivity*: $f(P\times Q) = f(P) + f(Q)$, where $P\times Q$ is the product-distribution of P and Q (i.e. the distribution in which events have probability p_iq_j for

each $p_i \in P$ and $q_j \in Q$); (P6) *subadditivity*: $f(A, B) \leq f(A) + f(B)$, where A and B are discrete random variables; (P7) *strong additivity*: $f(A, B) = f(A) + f(B|A)$; (P8) *recursivity*: $f(p_1, p_2, \ldots, p_N) = f(p_1+p_2, p_3, \ldots, p_N)+(p_1+p_2) f(p_1/(p_1+p_2), p_2/(p_1+p_2))$; and (P9) *sum-property*: $f(p_1, \ldots, p_N) = \sum_{n=1}^{N} g(p_n)$ for some function g.

Shannon entropy is the only uncertainty measure to satisfy all axioms (P1–9) listed by Csiszár; but in fact various proper subsets of these axioms are sufficient to characterize Shannon entropy fully. In particular, Shannon himself showed that continuity, strong additivity, and the property that the uncertainty of a uniform distribution should not decrease as the number of elements in the distribution increases, are sufficient to determine entropy up to a constant factor [2]. Khinchin proved a similar result using strong additivity, expansibility, and the property that the maximum uncertainty should be realized in a uniform distribution [7].

Rényi explored ways to relax the axiomatization of Shannon entropy to derive more general uncertainty measures [8]. He showed that Shannon entropy could be characterized by five postulates: (R1) symmetry; (R2) continuity; (R3) $f(1/2, 1/2) = 1$; (R4) additivity; and (R5) the entropy of the union of two subdistributions is the arithmetic weighted average of each individual subdistribution. By replacing the weighted average in postulate (R5) with the (more relaxed) exponential mean, Rényi uniquely determined the family of Rényi entropies for full probability distributions $H_\alpha(p_1, p_2, \ldots, p_n) = 1/(1-\alpha) \log_2(\sum_{k=1}^{n} p_k^\alpha)$, where $0<\alpha<\infty$, with $\alpha\neq1$, is a parameter. In the limit of α tending to 1, H_α coincides with Shannon entropy, and in the limit of α tending to infinity, H_α is min-entropy (i.e. the negated log of Bayes vulnerability).

Following Denning's seminal work [29], Shannon entropy has been widely used in the field of quantitative information flow for the leakage of confidential information [1] [30] [31] [32] [33] [34] [35]. But as the field of quantitative information flow continued to evolve, new measures of uncertainty and of information were proposed. Contrary to Rényi's motivation, however, most measures were not derived from mathematical principles, but instead were motivated by specific operational scenarios. That was the case for guessing entropy, Bayes vulnerability, and g-vulnerability, for instance. Although many "healthiness properties" have been proved after the fact for these measures (e.g. non-negativity, non-decrease of uncertainty by post-processing, etc.), there has not always been a derivation of such measures from basic principles, or attempts to verify whether they can be unified in a more general framework.

Naturally, since measures other than Shannon entropy cannot satisfy all postulates (P1–9), the axioms for vulnerability considered in this paper differ from those listed by Csiszár. Some differences are unimportant: they are just adaptations of axioms of uncertainty to axioms of vulnerability (e.g. conditioning of random variables reduces uncertainty, but increases vulnerability, so some inequalities must be reversed).

Other differences are more fundamental, however, as they reflect our departure from Shannon's indifference to the *meaning* of different secret values. The axiom of symmetry (P3), for instance, assumes that all secret values are equally informative—and in many scenarios that is false. For instance, not everyone's bank account is as worth breaking in to as everyone else's, and so evidently a permutation on the probabilities of every particular account being broken into does not amount to the same vulnerability. The axioms of additivity (P5), subadditivity (P6) and strong additivity (P7) assume that the uncertainty of a pair of joint random variables is a function only of the correlation of the random variables, which is also not a valid assumption in many security scenarios: the information of the combination of two secrets may exceed the information content of each secret separately: for instance, the benefit of knowing someone's PIN *and* bank-account number at the same time greatly surpasses the sum of the benefits of knowing each one on its own. Recursivity (P8) and the sum-property (P9) assume that the probability of each secret value contributes on equal terms to the overall uncertainty of the probability distribution, which also is a false assumption for many relevant measures. Bayes vulnerability, for instance, satisfies neither recursivity nor the sum-property, as the information of a probability distribution is a function of the maximum probability only.

8.2. Relation with Kifer and Lin's work

Kifer and Lin's work is the one most closely related to ours. In a series of papers [36–38,9], these authors proposed an axiomatic characterization of "good" properties that sanitization mechanisms should provide, focusing in particular on *privacy* and *utility measures*. They considered utility as *information preservation*, which captures how "faithful" the output of the mechanism is to its input,[11] and as such is closely related to our notion of vulnerability. This notion derives from the more general concept of utility used in decision theory. Kifer and Lin argued that utility has not been studied systematically in the context of privacy, and that some proposals have led to inconsistencies and paradoxes.

In the following we summarize the connection between our paper and their work. We start by briefly recalling their basic concepts and notation. A sanitization mechanism $\mathcal{M}$ is a randomized algorithm from inputs to outputs,[12] whose behavior is described by conditional probabilities $P_{\mathcal{M}}(o|i)$ of observing output o when input is i. Such privacy mechanisms correspond exactly to our channels. Given two mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$ and $p \in [0, 1]$, $\mathcal{M}_1 \oplus_p \mathcal{M}_2$ denotes the mechanism that, on input D, returns $\mathcal{M}_1(D)$ with probability p and $\mathcal{M}_2(D)$ with probability $1-p$, and also reveals whether the output was created using $\mathcal{M}_1$ or $\mathcal{M}_2$.

[11] This is in contrast with utility as *usability*, which expresses how easily the output can be used. An example of the difference is provided by an encryption mechanism, which perfectly preserves information, but whose output is not usable except by users who know the decryption key.

[12] In Kifer and Lin's work, the inputs of a mechanism are assumed to be datasets, and denoted by D. However, the discussion of this section applies to inputs and outputs of any kind.

A measure of information preservation is a function μ mapping a mechanism $\mathcal{M}$ to a real value. Lin and Kifer [9] describe five axioms that such measures should satisfy:

(1) *Sufficiency*: $\mu(\mathcal{M}) \geq \mu(\mathcal{A} \circ \mathcal{M})$ for any randomized algorithm $\mathcal{A}$. Here $\circ$ represents functional composition.
(2) *Continuity*: μ is continuous in the components of $\mathcal{M}$ (viewed as a matrix).
(3) *Branching*: Given a mechanism $\mathcal{M}$ with output space $\{o_1, \ldots o_n\}$ there is a function G such that $\mu(\mathcal{M}) = G(P_{\mathcal{M}}(o_1|\cdot), P_{\mathcal{M}}(o_2|\cdot)) + \mu(\mathcal{M}')$, where $\mathcal{M}'$ is obtained from $\mathcal{M}$ by adding together the columns $P_{\mathcal{M}}(o_1|\cdot), P_{\mathcal{M}}(o_2|\cdot)$ and leaving the others unchanged.
(4) *Quasi-convexity*: $\mu(\mathcal{M}_1 \oplus_p \mathcal{M}_2) \leq \max(\mu(\mathcal{M}_1), \mu(\mathcal{M}_2))$.
(5) *Quasi-concavity*: $\mu(\mathcal{M}_1 \oplus_p \mathcal{M}_2) \geq \min(\mu(\mathcal{M}_1), \mu(\mathcal{M}_2))$.

Lin and Kifer analyzed in [9] many popular measures of utility from the literature of privacy, and showed that almost all of them fail to satisfy the above axioms. One exception is the notion of g-vulnerability, as we will see in a moment.

By observing that our notion of vulnerability is essentially the utility of the adversary, we can make several connections between Kifer and Lin's principles and our own. First, their sufficiency axiom is clearly related to our data-processing inequality (DPI), since $\mathcal{A} \circ \mathcal{M}$ represents the post-processing of $\mathcal{M}$ by $\mathcal{A}$. Furthermore, they showed in [9] that Axioms (1)–(3) characterize a measure based on posterior g-vulnerability. More formally[13]:

Theorem 28 *(Lin and Kifer [9], Theorem 6.2). It is the case that:*

(a) $\forall g \, \forall \pi \, \exists \mu$ *satisfying (1)–(5):* $\forall \mathcal{M} \, \widehat{V}_g[\pi \triangleright \mathcal{M}] = \mu(\mathcal{M})$, *and*
(b) $\forall \mu$ *satisfying (1)–(3)* $\exists \pi \, \exists g$: $\forall \mathcal{M} \, \widehat{V}_g[\pi \triangleright \mathcal{M}] = \mu(\mathcal{M})$.

From previous sections, we know that any function satisfying continuity (CNTY),[14] convexity (CVX), and averaging (AVG) corresponds to a posterior g-vulnerability for some g. Together with the above result, this suggests a strong relation between information preservation and the notion of average-based posterior vulnerability explored in this paper.

However, there are important differences. First of all, the type of μ and that of posterior vulnerability are different: posterior vulnerability applies to a hyper-distribution, typically derived from a channel $\mathcal{M}$ and a prior π. On the other hand, μ applies only to a channel $\mathcal{M}$. This means that the prior π is *implicitly encoded* into μ, and that the utility $\mu(\mathcal{M})$ is the utility of $\mathcal{M}$ under the fixed prior π. A second (related) difference is that, while we can express the prior vulnerability as a particular case of posterior vulnerability, this is not the case for μ. In fact, we can express the utility of the distribution π associated to μ as $\mu(\bar{0})$, but we cannot express the utility of a generic distribution via the same μ. Indeed, because of Axiom (1), for any $\mathcal{M}$, $\mu(\mathcal{M})$ has an utility greater than or equal to that of $\mu(\bar{0})$, thus it cannot represent the utility of any π' that has less utility than π. As a consequence, it seems that the relation between prior and posterior measures, which is a major contribution of our paper, cannot be expressed in Kifer and Lin's framework. At least, not by using μ alone: one would need to introduce and axiomatize a new function. In particular, the averaging axiom (AVG) cannot be formulated by using μ alone. Similarly, the maximum (MAX) and the bounds (BNDS) axioms cannot be formulated, despite the resemblance of the latter with the axioms (4) and (5) above.

In summary, a main novelty with respect to the work of Kifer and Lin is that we investigate the relation between prior and posterior vulnerabilities. Another novel contribution is the study of the relationships between alternative sets of axioms. In general, indeed, our focus is different from that of Kifer and Lin: they focused on finding a collection of axioms for analyzing utility specifically, and used them to review the current practices in the field of privacy. In contrast, our main motivation is to establish the scientific principles which can help in the development or adaptation of new measures in response to novel situations. Thus, we explored different sets of possible axioms, thereby clarifying the implications between the principles themselves.

8.3. Relation with Boreale and Pampaloni's work

In [17,18], Boreale and Pampaloni have conducted one of the first studies of adaptive adversaries in the context of quantitative information flow. They did not consider explicitly an axiomatic framework, but, in order for their results to be as general as possible, they adopted a generic notion of entropy, specified by a few properties which turn out to be our axioms of concavity, continuity, and averaging. Furthermore, in [18] they pointed out a known theorem in decision theory, which states that a function $H : \mathbb{D}\mathcal{X} \to \mathbb{R}^+$, satisfies concavity and continuity iff it is of the form $H(\pi) = \sum_x \pi_x S(x, \pi)$, where $S : \mathcal{X} \times \mathbb{D}\mathcal{X} \to \mathbb{R}^+$, is any function which satisfies the condition that $\sum_x \pi_x S(x, \pi')$ is minimal when $\pi' = \pi$. Such function S, called *Proper Scoring Rule* in decision theory, is similar to the (converse of) *gain functions* used in g-vulnerability,

[13] Theorem 28 was actually formulated for the converse functions: the *information loss* and the *expected error of a Bayesian decision maker*, which are converse of the information preservation and of the posterior g-vulnerability, respectively.
[14] Note that (CNTY) and (2) refer to different type of arguments.

and therefore the above definition is related to that of prior g-entropy. Thus this result is similar to that of the completeness of g-vulnerability with respect to our axiomatization of the prior vulnerability (Theorem 9).

9. Conclusion

We have presented axioms that might be satisfied by intuitively reasonable measures of the prior- and posterior vulnerability of a secret as it is being processed by a system; this allowed us to derive properties of leakage. Our first main contribution was (1) the equivalence of the axioms of convexity, monotonicity (i.e. non-negativity of leakage), and data-processing inequality (DPI) when posterior vulnerability is defined as the average vulnerability of the posteriors, and (2) the equivalence of quasi-convexity, monotonicity and DPI when posterior vulnerability is defined as the worst-case vulnerability of posterior distributions. A deep implication of these results is that convexity (and quasi-convexity) of information measures do not need to be taken as fundamental properties, but are derivable from more intuitive principles, such as averaging (or worst-case analysis) and DPI.

The second main contribution was the demonstration of the soundness and completeness of g-vulnerabilities with respect to the axioms of convexity and continuity. Moreover, because of the equivalences we established, it follows that g-vulnerability exactly captures all average-based information measures that respect DPI or monotonicity.

We now want to further investigate the full family of vulnerabilities under quasi-convexity and continuity, characterizing all worst-case based vulnerabilities that respect DPI or monotonicity.

Acknowledgements

Mário S. Alvim was partially supported by the National Council for Scientific and Technological Development – CNPq, by CAPES, and by FAPEMIG. Konstantinos Chatzikokolakis and Catuscia Palamidessi were supported by MAGIC, and by the Equipe Associée LOGIS. Annabelle McIver and Carroll Morgan were supported by the Australian Research Council Grant DP140101119. Geoffrey Smith was partially supported by the National Science Foundation under grants CNS-1116318 and CNS-1749014. Also, the authors are grateful for support from Digiteo and the INRIA Équipe Associée Princess.

Appendix A. Elementary examples supporting Section 7

Recall that we take our base space to be $\mathcal{X}$, distributions on that to be $\mathbb{D}\mathcal{X}$, and hypers to be $\mathbb{D}^2\mathcal{X}$. Typical elements of $\mathbb{D}\mathcal{X}$ are lower-case Greek letters, possibly superscripted. Thus π_x is the probability π assigns to x and π^1_x is the probability π^1 assigns to x and $\Delta^2_{\delta^3}$ is the probability that hyper $\Delta^2{:}\mathbb{D}^2\mathcal{X}$ assigns to distribution $\delta^3{:}\mathbb{D}\mathcal{X}$. In Δ_δ usually δ will be in the support $\lceil\Delta\rceil$ of Δ; if not, then of course the assigned probability is zero.

We (continue to) write the point, or "singleton" distribution on x as $[x]$, so that $[x]_{x'} = 1$ IF $x = x'$ ELSE 0—it is the same as ηx if we are using a monad. A "doubleton" distribution say $\delta{=}x_1{}_p{\oplus}x_2$ is such that $\delta_{x_1}{=}p$ and $\delta_{x_2}{=}1{-}p$. The p-weighted sum of two values is defined $x_1{}_p{+}x_2 = px_1{+}(1{-}p)x_2$, thus not the same thing as $x_1{}_p{\oplus}x_2$: for example if $\mathcal{X}$ were the reals $\mathbb{R}$, then $x_1{}_p{+}x_2$ would also be a real, but $x_1{}_p{\oplus}x_2$ would be a (doubleton) distribution in $\mathbb{D}\mathbb{R}$. Indeed we have $x_1{}_p{\oplus}x_2 = [x_1]{}_p{+}[x_2]$. In both cases the p-factor applies on the left.

In this section we use μ more generally than multiply of a monad, as introduced in Section 7 above: here μ will as well simply average *any* distribution taken over a vector space. Thus in particular we have $\mu(\delta^1{}_p{\oplus}\delta^2) = \delta^1{}_p{+}\delta^2$ because $\delta^1{}_p{\oplus}\delta^2$, a hyper with just two inners, is in $\mathbb{D}^2\mathcal{X} = \mathbb{D}(\mathbb{D}\mathcal{X})$ and $\mathbb{D}\mathcal{X}$ is a vector space.

We return first return to the higher-order formulation $\widehat{\mathbb{V}}{=}\mu{\circ}\mathbb{D}\mathbb{V}$ of AVG. With a doubleton hyper for illustration, say $\Delta = \pi^1{}_p{\oplus}\pi^2$, that gives

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\Delta = (\mu{\circ}\mathbb{D}\mathbb{V})\Delta & \text{“apply AVG to } \Delta\text{”} \\
\Rightarrow & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = (\mu{\circ}\mathbb{D}\mathbb{V})(\pi^1{}_p{\oplus}\pi^2) & \text{“}\Delta = \pi^1{}_p{\oplus}\pi^2\text{”} \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mu(\mathbb{D}\mathbb{V}(\pi^1{}_p{\oplus}\pi^2)) & \text{“composition”} \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mu(\mathbb{V}\pi^1{}_p{\oplus}\mathbb{V}\pi^2) & \text{“definition functor } \mathbb{D}\text{”} \\
\text{iff} & \widehat{\mathbb{V}}(\pi^1{}_p{\oplus}\pi^2) = \mathbb{V}\pi^1{}_p{+}\mathbb{V}\pi^2\ , & \text{“property of } \mu\text{”}
\end{array}
$$

showing that $\widehat{\mathbb{V}}$ indeed takes the weighted sum of $\mathbb{V}$ applied to the (two, in this case) posteriors in Δ. As an illustration of more general reasoning, we give the proof promised in Section 7, i.e., that applies even when the weighted average is not necessarily over just two hypers. Note that the resulting proof is *less* cluttered than the one immediately above, and also less so than a conventional presentation would be with summations and subscripts. We have

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}} \circ \mu & \\
= & \mu \circ \mathbb{D}\mathbb{V} \circ \mu & \text{“assumption AVG”} \\
= & \mu \circ \mu \circ \mathbb{D}^2\mathbb{V} & \text{“}\mu\text{ is natural transformation } \mathbb{D}^2{\to}\mathbb{D}\text{”} \\
= & \mu \circ \mathbb{D}\mu \circ \mathbb{D}^2\mathbb{V} & \text{“monad coherence condition on } \mu\text{”} \\
= & \mu \circ \mathbb{D}(\mu \circ \mathbb{D}\mathbb{V}) & \text{“}\mathbb{D}\text{ functor”} \\
= & \mu \circ \mathbb{D}\widehat{\mathbb{V}}\ , & \text{“assumption AVG”}
\end{array}
$$

which overall equality says intuitively that applying $\widehat{\mathbb{V}}$ to the weighted sum of some hypers, i.e., $\widehat{\mathbb{V}}(\mu\underline{\Delta})$, is the same as applying $\widehat{\mathbb{V}}$ to the hypers separately and then taking the weighted sum of the results, i.e., $\mu(\mathbb{D}\widehat{\mathbb{V}}\underline{\Delta})$. (†)

The higher-order NI captures its traditional definition via $\widehat{\mathbb{V}}[\pi]{=}\widehat{\mathbb{V}}(\eta\pi){=}(\widehat{\mathbb{V}}\circ\eta)\pi{=}\mathbb{V}\pi$. Here is how the higher-order version of NI follows from AVG and the monad laws, as we claimed in Section 7:

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\circ\eta & \\
= & \mu\circ\mathbb{D}\mathbb{V}\circ\eta & \text{“assumption AVG”} \\
= & \mu\circ\eta\circ\mathbb{V} & \text{“}\eta\text{ is natural transformation }1\to\mathbb{D}\text{”} \\
= & \mathbb{V}\ . & \text{“monad coherence condition }\mu\circ\eta=1\text{”}
\end{array}
$$

Applying that to arbitrary π gives $\widehat{\mathbb{V}}(\eta\pi)=\mathbb{V}\pi$.

For CVX again we use a doubleton hyper $\Delta=\pi^1{}_p{\oplus}\pi^2$ as an example, so that the traditional formulation of CVX is found in the middle of the following string of equalities:

$$
\begin{array}{lll}
 & (\mathbb{V}\circ\mu)\Delta & \\
= & (\mathbb{V}\circ\mu)(\pi^1{}_p{\oplus}\pi^2) & \text{“definition }\Delta\text{”} \\
= & \mathbb{V}(\mu(\pi^1{}_p{\oplus}\pi^2)) & \text{“composition”} \\
= & \mathbb{V}(\pi^1{}_p{+}\pi^2) & \text{“property of }\mu\text{”} \\
\le & \mathbb{V}\pi^1{}_p{+}\,\mathbb{V}\pi^2 & \text{“traditional formulation of CVX”} \\
= & \mu(\mathbb{V}\pi^1{}_p{\oplus}\,\mathbb{V}\pi^2) & \text{“property of }\mu\text{”} \\
= & \mu(\mathbb{D}\mathbb{V}(\pi^1{}_p{\oplus}\,\pi^2)) & \text{“definition functor }\mathbb{D}\text{”} \\
= & (\mu\circ\mathbb{D}\mathbb{V})(\pi^1{}_p{\oplus}\,\pi^2) & \text{“composition”} \\
= & (\mu\circ\mathbb{D}\mathbb{V})\Delta\ , & \text{“definition }\Delta\text{ again”}
\end{array}
$$

showing how CVX for this particular Δ agrees with the higher-order formulation $\mathbb{V}\circ\mu\le\mu\circ\mathbb{D}\mathbb{V}$ at (3) in Section 7.

Now we return to the formulation of the partial order $\sqsubseteq$ on hypers in terms of the surprising “hyper–hyper” $\underline{\Delta}$ in $\mathbb{D}^3\mathcal{X}$. As we did above (at †), we will assist the intuition by taking a simple case $\underline{\Delta}=\Delta^1{}_p{\oplus}\Delta^2$, thus a doubleton hyper-hyper over two hypers Δ^1 with probability p and Δ^2 with probability $1{-}p$. We show that the higher-order definition of $\sqsubseteq$ implies the $\mathbb{V}$-based definition, in this case, provided we assume CVX. (The reverse direction is harder, related to the *Coriaceous Conjecture* described in [5] and proved in [6,12].)

We start by setting $\Delta^+{=}\mu\underline{\Delta}$ and $\Delta^-{=}(\mathbb{D}\mu)\underline{\Delta}$, as in the higher-order formulation of $\sqsubseteq$ from which we would expect to be able to prove that $\widehat{\mathbb{V}}\Delta^+\ge\widehat{\mathbb{V}}\Delta^-$. Then we have

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\Delta^+ & \\
= & \widehat{\mathbb{V}}(\mu\underline{\Delta}) & \text{“definition }\Delta^+\text{”} \\
= & \widehat{\mathbb{V}}(\mu(\Delta^1{}_p{\oplus}\Delta^2)) & \text{“definition }\underline{\Delta}\text{”} \\
= & \widehat{\mathbb{V}}(\Delta^1{}_p{+}\Delta^2) & \text{“property of }\mu\text{”} \\
= & \widehat{\mathbb{V}}\Delta^1{}_p{+}\,\widehat{\mathbb{V}}\Delta^2 & \text{“linearity of }\widehat{\mathbb{V}}\text{, implied by AVG”} \\
= & (\mu\circ\mathbb{D}\mathbb{V})\Delta^1{}_p{+}\,(\mu\circ\mathbb{D}\mathbb{V})\Delta^2 & \text{“assume AVG”} \\
\ge & (\mathbb{V}\circ\mu)\Delta^1{}_p{+}\,(\mathbb{V}\circ\mu)\Delta^2 & \text{“assume CVX”} \\
= & \mu((\mathbb{V}\circ\mu)\Delta^1{}_p{\oplus}\,(\mathbb{V}\circ\mu)\Delta^2) & \text{“property }\mu\text{”} \\
= & \mu(\mathbb{D}(\mathbb{V}\circ\mu)(\Delta^1{}_p{\oplus}\,\Delta^2)) & \text{“functor }\mathbb{D}\text{”} \\
= & (\mu\circ\mathbb{D}\mathbb{V}\circ\mathbb{D}\mu)\underline{\Delta} & \text{“composition; functor }\mathbb{D}\text{; definition }\underline{\Delta}\text{”} \\
= & \widehat{\mathbb{V}}(\mathbb{D}\mu\underline{\Delta}) & \text{“assume AVG; composition”} \\
= & \widehat{\mathbb{V}}\Delta^-\ . & \text{“definition }\Delta^-\text{”}
\end{array}
$$

As in the earlier examples, once one is familiar with the general monadic operators, one gains access to a presentation of essentially the same proof but in a shorter, less cluttered (with no super- or subscripts), more algebraic style. More significantly, however, is that these operators have general monadic formulations that accommodate infinite state-spaces and proper measures. The algebraic proof applies as-is, unchanged. Thus for general $\underline{\Delta}$, is in effect a soundness proof for $\sqsubseteq$, that it can only decrease vulnerability (given CVX and AVG); and because of the great generality of the monad framework [21] it applies even for infinite state spaces $\mathcal{X}$ and measures. It is

$$
\begin{array}{lll}
 & \widehat{\mathbb{V}}\circ\mu & \\
= & \mu\circ\mathbb{D}\widehat{\mathbb{V}} & \text{“linearity of }\widehat{\mathbb{V}}\text{, proved earlier at †”} \\
\ge & \mu\circ\mathbb{D}(\mathbb{V}\circ\mu) & \text{“AVG and CVX; see ‡ below”} \\
= & \mu\circ\mathbb{D}\mathbb{V}\circ\mathbb{D}\mu & \text{“}\mathbb{D}\text{ functor”} \\
= & \widehat{\mathbb{V}}\circ\mathbb{D}\mu\ . & \text{“AVG with “the other }\mu\text{””}
\end{array}
$$

The longer proof just above is recovered by applying each line to $\underline{\Delta}=\Delta^1{}_p{\oplus}\Delta^2$.

The "see below" appeals to the elementary general fact that if two functions $f, f' : \mathcal{S} \to \mathbb{R}$ satisfy $f(s) \geq f'(s)$ for all $s : \mathcal{S}$, then also $(\mu \circ \mathbb{D} f)(\delta) \geq (\mu \circ \mathbb{D} f')(\delta)$ for all δ in $\mathbb{D}\mathcal{S}$, in words that if two random variables over the same distribution satisfy $\geq$ everywhere, then so do their expected values. Above we used $f = \widehat{\mathbb{V}}$ and $f' = \mathbb{V} \circ \mu$ and $\mathcal{S} = \mathbb{D}^2\mathcal{X}$, appealing to AVG and CVX for the inequality. (‡)

References

[1] D. Clark, S. Hunt, P. Malacaria, Quantitative information flow, relations and polymorphic types, J. Logic Comput. 18 (2) (2005) 181–199.
[2] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 625–656.
[3] J.L. Massey, Guessing and entropy, in: Proceedings of the IEEE Int. Symposium on Information Theory, IEEE, 1994, p. 204.
[4] G. Smith, On the foundations of quantitative information flow, in: Proc. of FOSSACS, in: Lecture Notes in Computer Science, vol. 5504, Springer, 2009, pp. 288–302.
[5] M.S. Alvim, K. Chatzikokolakis, C. Palamidessi, G. Smith, Measuring information leakage using generalized gain functions, in: Proc. of CSF, 2012, pp. 265–279.
[6] A. McIver, L. Meinicke, C. Morgan, Compositional closure for Bayes risk in probabilistic noninterference, in: Proc. of ICALP, in: Lecture Notes in Computer Science, vol. 6199, Springer, 2010, pp. 223–235.
[7] A. Khinchin, Mathematical Foundations of Information Theory, Dover Books on Mathematics, Dover Publications, 1957.
[8] A. Rényi, On measures of entropy and information, in: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability, 1961, pp. 547–561.
[9] B. Lin, D. Kifer, Information measures in statistical privacy and data processing applications, ACM Trans. Knowl. Discov. Data 9 (4) (2015) 28.
[10] M.S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, G. Smith, Axioms for information leakage, in: Proc. of CSF, 2016, pp. 77–92.
[11] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, On the Bayes risk in information-hiding protocols, J. Comp. Secur. 16 (5) (2008) 531–571.
[12] M.S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, G. Smith, Additive and multiplicative notions of leakage, and their capacities, in: Proc. of CSF, IEEE, 2014, pp. 308–322.
[13] M.S. Alvim, A. Scedrov, F.B. Schneider, When not all bits are equal: worth-based information flow, in: Proc. 3rd Conference on Principles of Security and Trust (POST 2014), 2014, pp. 120–139.
[14] A. McIver, C. Morgan, G. Smith, B. Espinoza, L. Meinicke, Abstract channels and their robust information-leakage ordering, in: Proc. of POST, in: Lecture Notes in Computer Science, vol. 8414, Springer, 2014, pp. 83–102.
[15] R.T. Rockafellar, Convex Analysis, Princeton Mathematical Series, Princeton University Press, Princeton, NJ, 1970.
[16] S. Shalev-Shwartz, Online learning and online convex optimization, Found. Trends Mach. Learn. 4 (2) (2012) 107–194.
[17] M. Boreale, F. Pampaloni, Quantitative information flow under generic leakage functions and adaptive adversaries, in: Proc. of FORTE, in: Lecture Notes in Computer Science, vol. 8461, Springer, 2014, pp. 166–181.
[18] M. Boreale, F. Pampaloni, Quantitative information flow under generic leakage functions and adaptive adversaries, Log. Methods Comput. Sci. 11 (4) (2015).
[19] J. McLean, Security models and information flow, in: Proc. of S&P, IEEE, 1990, pp. 180–189.
[20] G. Smith, D. Volpano, Secure information flow in a multi-threaded imperative language, in: Proc. of POPL, POPL '98, ACM, New York, NY, USA, 1998, pp. 355–364.
[21] M. Giry, A categorical approach to probability theory, in: Categorical Aspects of Topology and Analysis, in: Lecture Notes in Mathematics, vol. 915, Springer-Verlag, 1981, pp. 68–85.
[22] A. McIver, L. Meinicke, C. Morgan, A Kantorovich-monadic powerdomain for information hiding, with probability and nondeterminism, in: Proc. LiCS 2012, 2012.
[23] T. Schrijvers, A monadic model for computations that leak secrets, http://www.cse.unsw.edu.au/~carrollm/LiCS15-TS.pdf, 2015.
[24] A. McIver, L. Meinicke, C. Morgan, Hidden-Markov program algebra with iteration, Math. Structures Comput. Sci. 25 (2) (2015) 320–360.
[25] F. van Breugel, The metric monad for probabilistic nondeterminism, draft available at http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf, 2005.
[26] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd edition, J. Wiley & Sons, Inc., 2006.
[27] C. Dwork, Differential privacy, in: Proc. of ICALP, in: Lecture Notes in Computer Science, vol. 4052, Springer, 2006, pp. 1–12.
[28] I. Csiszár, Axiomatic characterizations of information measures, Entropy 10 (3) (2008) 261.
[29] D. Denning, Cryptography and Data Security, Addison–Wesley, 1983.
[30] P. Malacaria, Assessing security threats of looping constructs, in: Proc. of POPL, ACM, 2007, pp. 225–235.
[31] K. Chatzikokolakis, C. Palamidessi, P. Panangaden, Anonymity protocols as noisy channels, Inform. and Comput. 206 (2–4) (2008) 378–401.
[32] P. Malacaria, H. Chen, Lagrange multipliers and maximum information leakage in different observational models, in: Proc. of PLAS, ACM, 2008, pp. 135–146.
[33] I.S. Moskowitz, R.E. Newman, P.F. Syverson, Quasi-anonymous channels, in: Proc. of CNIS 2003, IASTED, 2003, pp. 126–131.
[34] I.S. Moskowitz, R.E. Newman, D.P. Crepeau, A.R. Miller, Covert channels and anonymizing networks, in: Proc. of WPES, ACM, 2003, pp. 79–88.
[35] M.S. Alvim, M.E. Andres, C. Palamidessi, Information flow in interactive systems, J. Comp. Secur. 1 (20) (2012) 3–50.
[36] D. Kifer, B.-R. Lin, Towards an axiomatization of statistical privacy and utility, in: Proc. of PODS, ACM, 2010, pp. 147–158.
[37] B. Lin, D. Kifer, Reasoning about privacy using axioms, in: M.B. Matthews (Ed.), Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers, ACSCC 2012, Pacific Grove, CA, USA, November 4–7, 2012, IEEE, 2012, pp. 975–979.
[38] D. Kifer, B.-R. Lin, An axiomatic view of statistical privacy and utility, J. Priv. Confidential. 4 (1) (2012) 5–49.