2014 IEEE 27th Computer Security Foundations Symposium

Additive and multiplicative notions of leakage, and their capacities

Mário S. Alvim* Konstantinos Chatzikokolakis† Annabelle McIver‡
Carroll Morgan§ Catuscia Palamidessi† Geoffrey Smith¶

* Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte

† INRIA, CNRS and LIX
École Polytechnique
Paris

‡ Department of Computing
Macquarie University
Sydney

§ School of Computer Science & Engineering
University of New South Wales and NICTA
Sydney

¶ School of Computing & Information Sciences
Florida International University
Miami

***Abstract*—Protecting sensitive information from improper disclosure is a fundamental security goal. It is complicated, and difficult to achieve, often because of unavoidable or even unpredictable operating conditions that can lead to breaches in planned security defences. An attractive approach is to frame the goal as a quantitative problem, and then to design methods that measure system vulnerabilities in terms of the amount of information they leak. A consequence is that the precise operating conditions, and assumptions about prior knowledge, can play a crucial role in assessing the severity of any measured vunerability.**

We develop this theme by concentrating on vulnerability measures that are *robust* in the sense of allowing general leakage bounds to be placed on a program, bounds that apply whatever its operating conditions and whatever the prior knowledge might be. In particular we propose a theory of channel capacity, generalising the Shannon capacity of information theory, that can apply both to additive- and to multiplicative forms of a recently-proposed measure known as g-leakage. Further, we explore the computational aspects of calculating these (new) capacities: one of these scenarios can be solved efficiently by expressing it as a Kantorovich distance, but another turns out to be NP-complete.

We also find capacity bounds for arbitrary correlations with data not directly accessed by the channel, as in the scenario of Dalenius's Desideratum.

Keywords*—*Quantitative information flow, channel capacity, confidentiality, Dalenius Desideratum.

I. Introduction

Protecting sensitive information from improper disclosure is a fundamental security goal, but one that is clearly not being achieved well in today's cyber infrastructure. The issue is complicated by the fact that some leakage of sensitive information is often unavoidable in practice, due either to system functionality (e.g. a statistical database must by design reveal information derived from the database entries, even if those entries themselves should be confidential) or due to side-channels (e.g. it is difficult to prevent running time or power consumption from depending on secrets). For this reason, it is attractive to approach the problem quantitatively, and indeed there has been good progress in recent years in the area of *quantitative information flow*, including works with a foundational focus [1]–[11], works aimed at the development of verification techniques [12]–[19], and works analyzing the leakage in real-system vulnerabilities [20]–[22].

One important quantitative theme has been the development of leakage measures with strong *operational significance* for confidentiality, so that the amount of information leaked is associated with the constraints and rewards of real attacks. In this respect, notable measures include *min-entropy leakage* [6] and *g-leakage* [11]. Assuming that an adversary's prior knowledge about the secret is represented as a probability distribution π, min-entropy leakage is based on the ratio of the secret's posterior- and prior *vulnerability* to being guessed correctly in one try by the adversary. One-try vulnerability is clearly a basic and important measure of confidentiality, but it should be recognized that it is not appropriate in all situations. For this reason, g-leakage generalizes vulnerability with a *gain function* g that models the operational scenario, allowing us to consider situations in which the adversary gains e.g. by guessing the secret partially, approximately, or within k tries. It can also model scenarios where there is a *penalty* for making incorrect guesses.

A second important theme concerns *robustness*. The leakage caused by a channel C depends on both the prior distribution π and the particular leakage measure used (e.g. Shannon entropy, guessing entropy, min-entropy, g-leakage). But what if we do not know π, or don't know the leakage measure, or don't know either? Then we might worry that our conclusions about information leakage lack robustness. Precisely for this reason, there has been considerable interest in achieving results that are less sensitive to the particular assumptions made of prior or entropy-measure. *Capacity* is the maximum leakage over all prior distributions—and its utility is that when a channel has small capacity, its leakage is small, no matter what the prior may be. It is also interesting to establish capacity relationships between different leakage measures. For instance, the *Miracle Theorem* of [11] shows that *min-capacity* (the maximum min-entropy leakage over all priors) is an upper bound on g-leakage for *every* gain function g—this means that a channel with small min-capacity has small g-leakage, no matter what prior or gain function is used. A second approach to robustness concerns *comparisons* between channels. It has recently been shown in the *Coriaceous Theorem* of [23] that a channel A *never* leaks

DOI 10.1109/CSF.2014.29

more than a channel B, on any prior π and gain function g, if and only if A *secures* B in the sense that that A is equivalent to the combination of B with some "post-processing" R.

This paper builds on these successes of the g-leakage model, achieving advances in both operational significance and robustness.

With respect to operational significance, recall that (as mentioned above) standard g-leakage is based on the *ratio* of the posterior and prior g-vulnerabilities. But there are situations where it seems more significant to consider instead the *difference* of the vulnerabilities. For instance, a system whose prior- and posterior vulnerabilities are 2^{-1000} and 2^{-700} respectively has a huge multiplicative increase of 2^{300}, but its additive increase is less than 2^{-700}, which might well be considered negligible. Moreover, g-vulnerabilities can be used to model the economic value to the adversary of the prior- and posterior situations; here too it seems more significant to consider their difference rather than their ratio. Finally, we discover that both guessing-entropy and Shannon-entropy leakage can be expressed as additive g-leakages. For these reasons, in this paper we develop the theory of *additive g-leakage*, complementing the existing multiplicative theory.

With respect to robustness, this paper presents a systematic study of *g-capacity*. Since the g-leakage of a channel C depends on both the prior π and the gain function g, we consider several versions of capacity: we can fix g and universally quantify π, or we can fix π and universally quantify g, or we can universally quantify both π and g. By considering both multiplicative- and additive leakage, we get a total of *six* capacity scenarios. We also consider a seventh capacity scenario related to the famous *ad-omnia* privacy desideratum proposed by Dalenius [24] for statistical databases, which states that nothing about an individual should be learnable with the database that could not equally well have been learned without it. As shown by Dwork [25], this desideratum is too strong to be achievable: for any useful database there is always some prior knowledge the adversary can use to induce a leakage of information. Analogously, we show here that g-leakage can quantify the leakage about a secret X caused by an apparently unrelated channel C from Y to Z when the adversary knows some *correlations* between X and Y.

Beyond our systematic study, we make novel and significant contributions in four of those seven capacity scenarios:

- Given a prior π, we show that a gain function g_π maximizing the multiplicative leakage can be computed efficiently. (§VI-B)
- For the specific gain function g_{id}, we show that the maximum additive leakage over all priors is *NP-complete*. (§VII-A)
- Given a prior π, we show that the maximum additive leakage over all gain functions can be computed in *linear* time, because this leakage can be expressed as a *Kantorovich distance*. (§VII-B)
- We show that the "Dalenius" leakage of X caused by a channel C from Y to Z is *bounded* by the capacity of C, regardless of the correlations that may exist between X and Y. (§IX)

The rest of this paper is organized as follows. Section II gives preliminaries, Section III motivates additive g-leakage, and Section IV presents useful algebraic properties for gain functions. Section V then defines the seven capacity scenarios that we consider. Sections VI and VII present our results on the multiplicative- and additive scenarios respectively, with Section VIII giving some examples. Section IX presents our "Dalenius" bounds, proven using new results about the additive capacity of channel cascades. Finally, Sections X and XI discuss related work and conclude.

II. Preliminaries

A. Basic notations

We begin by recalling the basic definitions of information-theoretic channels [26]. A *channel* is a triple $(\mathcal{X}, \mathcal{Y}, C)$ where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets (typically of secret input values and observable output values resp.) and C is an $|\mathcal{X}|\times|\mathcal{Y}|$ *channel matrix* whose entries are between 0 and 1 and whose rows each sum to 1. Typically we use upper-case Roman letters (like C) for channels, and calligraphic letters (like $\mathcal{X}, \mathcal{Y}$) for the sets over which they operate. We write $C_{x,y}$ for the element of C at row x in $\mathcal{X}$ and column y in $\mathcal{Y}$. For the (entire) row x we write $C_{x,-}$, and for column y we write $C_{-,y}$.

The value $C_{x,y}$ is the conditional probability of output y's being produced by C from input x. A channel is *deterministic* just when all its elements are either 0 or 1, implying that each input row contains a single 1 which identifies that input's unique corresponding output.

Typically we assume knowledge of a *prior* distribution π on $\mathcal{X}$, using lower-case Greek letters for simple distributions like that; by analogy with $C_{x,y}$ indexing a matrix, as above, we write π_x for π at x, thus the probability that π gives to x. For more complex distributions we use upper-case Greek letters;[1] for example the *joint distribution* on $\mathcal{X}\times\mathcal{Y}$ determined by input distribution π and channel C is typically written $\Pi_{x,y}=\pi_x C_{x,y}$; when Π is understood, we can use a Π-implicit convention to write that as just $p(x, y)$. The jointly distributed random variables X, Y have *marginal* probabilities that (again Π-implicitly) we write $p(x)=\sum_y p(x, y)$ (which of course is just π_x again) and $p(y)=\sum_x p(x, y)$; occasionally we also write X, Y informally for "the (unnamed) distribution on $\mathcal{X}, \mathcal{Y}$" resp. The *conditional probabilities* are then given by $p(y|x)=p(x,y)/p(x)$ (if $p(x)$ is non-zero) and $p(x|y)=p(x,y)/p(y)$ (if $p(y)$ is non-zero). Note that Π is the *unique* joint distribution that recovers π and C, in that $p(x)=\pi_x$ and $p(y|x)=C_{x,y}$ (if $p(x)$ is non-zero). When necessary to avoid ambiguity, we write these Π-implicit distributions with subscripts, e.g. p_X or p_{XY} or p_Y.

For example given $|\mathcal{X}|=3$ and $|\mathcal{Y}|=4$, we could consider a channel C (at left) which, when applied to (the uniform) prior $\pi = (1/3, 1/3, 1/3)$ produces the joint matrix Π (at right):

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

$\xrightarrow{\pi}$

Π	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

(1)

By summing Π's columns we get the $\mathcal{Y}$-marginal distribution $p_Y = (1/2, 5/18, 5/36, 1/12)$, and by normalizing those columns

[1] Where the upper-case Greek- and Roman letters are the same, we will write the Greek in sans-serif: thus upper-case τ would be T.

we obtain the four 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 $p_{X|y_4} = (0, 1, 0)$.

We use $\mathbb{D}()$ to construct the set of distributions on a set, so that $\mathbb{D}\mathcal{X}$ is typically the set of distributions on the input and $\mathbb{D}\mathcal{Y}$ is the set of distributions on the output. For a point- or equivalently Dirac distribution assigning probability 1 to some x in $\mathcal{X}$ (and thus 0 to all other elements of $\mathcal{X}$) we write $[x]$. We write $\lceil\pi\rceil$ for the support of distribution π, that is the (sub-)set of elements where it is not zero.[2] For the sum of a (sub-)distribution π we write just $\Sigma\pi$; using that we could write more succinctly $p(x) = \sum \Pi_{x,-}$ and $p(y) = \sum \Pi_{-,y}$.

For the *expected value* over a distribution $\pi: \mathbb{D}\mathcal{X}$ 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, we write $\mathcal{E}_\pi F$; thus in the discrete case we have $\mathcal{E}_\pi F = \sum_x \pi_x F(x)$. If $\mathcal{X}$ *itself* is a vector space, then we abbreviate $\mathcal{E}_\pi(\mathrm{id})$ by just $\mathcal{E}\pi$, the "average" of the distribution π on $\mathcal{X}$.

Leakage measures can be defined based on various entropy-like functions of the prior distribution π and the posterior distributions $p_{X|y}$, together with their probabilities $p(y)$. Here are some examples.

Shannon leakage is based on the Shannon entropy $H[\pi] = -\sum_x \pi_x \lg \pi_x$ [27] of the prior distribution [3] and the expected Shannon entropy of the posterior distributions $H[\pi, C] = \sum_y p(y) H[p_{X|y}]$. It is their difference $H[\pi] - H[\pi, C]$ that equals the mutual information $I(\pi, C)$.[4]

Guessing-entropy leakage is based on the guessing entropy [28] of the prior distribution, that is $G[\pi] = \sum_i i\, \pi_{x_i}$ where the i-indexing of $\mathcal{X}$ is in non-increasing probability order, and on the expected guessing entropy of the posterior distributions $G[\pi, C] = \sum_y p(y) G[p_{X|y}]$. The guessing-entropy leakage then is the difference $G[\pi] - G[\pi, C]$.

The operational significance of both Shannon- and guessing entropy can be stated in terms of the expected number of brute-force guesses an adversary would need to find the secret. [5]

Min-entropy leakage [6] is based on the prior *vulnerability* $V[\pi] = \max_x \pi_x$, the probability of the secret's being guessed in one try, and the expected vulnerability of the posterior distributions $V[\pi, C] = \sum_y p(y) V[p_{X|y}]$.

The min-entropy leakage $\mathcal{L}(\pi, C)$ is the logarithm of the ratio of the posterior- and prior vulnerabilities:

$$\mathcal{L}(\pi, C) \quad = \quad \lg(V[\pi, C] / V[\pi]) \ . \tag{2}$$

Note that vulnerability implicitly assumes an operational scenario in which the adversary gains only by guessing the secret *exactly*, and in *one try*.

For this reason, *g-leakage* [11] generalizes vulnerability to incorporate a *gain function* g, the choice of which allows the modelling of differing operational scenarios. In each scenario, there will be some set $\mathcal{W}$ of *guesses* that the adversary could make about the secret, and for any guess w and secret x, there will be some *gain* $g(w, x)$ that the adversary gets by having chosen w when the secret's actual value was x; gains range from 0, when choosing w has no value at all, to 1, when w is ideal. (In §III-A, and Def. 3 in §V this is generalised, however.)

Formally $g: \mathcal{W}\times\mathcal{X} \to [0, 1]$, where $\mathcal{W}$ is a finite set. Given a gain function g, the prior *g-vulnerability* is defined as the maximum expected gain over all possible guesses:

$$V_g[\pi] \quad = \quad \max_w \sum_x \pi_x g(w, x) \ . \tag{3}$$

The posterior g-vulnerability and the g-leakage are then defined as for min-entropy leakage above, so that we have

$$V_g[\pi, C] \quad = \quad \sum_y p(y) V_g[p_{X|y}] \ , \tag{4}$$

and then $\mathcal{L}_g(\pi, C) = \lg(V_g[\pi, C] / V_g[\pi])$.

Note that the *identity gain function* g_{id}, which gives gain 1 for guessing the secret correctly and 0 otherwise, makes g-leakage coincide with min-entropy leakage, which latter is thus a special case. But gain functions can do much more. As explained in [11], they can model a wide variety of practical operational scenarios, including those where the adversary benefits from guessing a value *close* to the secret, guessing a *part* of the secret, guessing a *property* of the secret or guessing the secret within some bounded number of tries. They can also model scenarios where there is a *penalty* for incorrect guesses.

We now recall some previous results about *capacity*. *Shannon capacity* and *min-capacity* are, respectively, the maximum Shannon leakage and maximum min-entropy leakage, over all priors. The min-capacity of C, denoted $\mathcal{ML}(C)$, is always realized on the uniform prior and it is equal to the logarithm of the sum of the column maximums of C [20], [29]. Min-capacity is an upper bound on Shannon capacity [11], and the two coincide on deterministic channels [30]. Finally, the *Miracle Theorem* [11] shows that min-capacity is an upper bound on g-leakage, for *every* prior π and gain function g.

B. Hyper-distributions and their notations

A more abstract view of the action of a channel on a prior (e.g. more abstract than producing a joint distribution) is that it produces a distribution of posteriors. Following [7], [23], we call those *hyper-distributions*, for brevity writing just "hyper".

A hyper on the input space $\mathcal{X}$ is of type $\mathbb{D}^2\mathcal{X}$. Consider a prior π on $\mathcal{X}$ and a channel $C: \mathcal{X}\to\mathcal{Y}$ which, as usual, determines a joint distribution $\Pi: \mathbb{D}(\mathcal{X}\times\mathcal{Y})$. As above, with Π understood we have a distribution p_Y on the outputs and, for each y, a corresponding posterior distribution $p_{X|y}$ on the inputs. If instead of considering p_Y to be a distribution on $\mathcal{Y}$ we consider it to be a distribution on the corresponding posteriors (i.e. on the normalised columns of Π themselves, rather than on their $\mathcal{Y}$-labels), then we have a hyper which we write $[\pi, C]$. For the example at (1) above, that hyper would assign probabilities $(1/2, 15/36, 1/12)$ to the posteriors $(2/3, 0, 1/3)$, $(0, 3/5, 2/5)$, and $(0, 1, 0)$, respectively.[6]

[2] Note that the characteristic function of the support set is indeed the ceiling of the distribution as a function.

[3] We write lg for $\log_2$ (Knuth).

[4] The more usual notation for these quantities is $H(X)$, $H(X|Y)$, and $I(X; Y)$. We explain at the end of §II why we use brackets $[\cdot]$.

[5] For Shannon entropy, this follows from a result by Massey [28].

[6] There might be fewer posteriors in the support of hyper $[\pi, C]$ than there are columns in the joint distribution Π from which it derives, because if several columns of Π normalise to the same posterior then the hyper will automatically coalesce them [23]: columns y_2 and y_3 were coalesced in this case.

Because hyper $[\pi, C]$ is of type $\mathbb{D}^2\mathcal{X}$, that is $\mathbb{D}\mathcal{V}$ where $\mathcal{V}$=$\mathbb{D}\mathcal{X}$ is a vector space, we can just average that hyper itself, in effect "collapsing" it; and that gives the original prior again. That is we have the convenient identity $\mathcal{E}[\pi, C] = \pi$.

A special case of a hyper is the *point* hyper on a particular (simple) distribution, say π: we write that as $[\pi]$, as stated above for point distributions generally and of which this is just a special case. In this way we can regard g-vulnerability V_g as acting on hypers in all cases. In particular $V_g[\pi]$ viewed this way, i.e. as a function V_g of type $\mathbb{D}^2\mathcal{X}\to\mathbb{R}$ applied to a (point) hyper, remains equal to $\max_w \sum_x \pi_x g(w, x)$ as we defined it above.

III. Additive Leakage

Additive leakage is the *difference* between the posterior and prior vulnerabilities, and we write it

$$\mathcal{L}^+_g(\pi, C) \quad = \quad V_g[\pi, C] - V_g[\pi] \ .$$

Note that the additive version of min-entropy leakage (i.e. with respect to g_{id}) was considered previously in [29] and [15]. (Standard g-leakage is defined *multiplicatively*, based on the *ratio* of the posterior and prior g-vulnerabilities.[7])

We observe that calculating leakage additively rather than multiplicatively will not have any impact on channels' leakage *ordering*, since the additive leakage ordering $\mathcal{L}^+_g(\pi, A) \le \mathcal{L}^+_g(\pi, B)$ and its multiplicative version $\mathcal{L}_g(\pi, A) \le \mathcal{L}_g(\pi, B)$ are both equivalent to $V_g[\pi, A] \le V_g[\pi, B]$ (provided that $V_g[\pi]>0$ in the multiplicative case). But we believe that additive g-leakage is interesting and important in several respects, explained below.

A. Operational significance

First, there are a number of situations where additive leakage may have greater operational significance than does multiplicative leakage. To illustrate, consider a secret array X of uniformly distributed 10-bit passwords for 1,000 users. The following probabilistic channel (it is (Ex1) from [11]) reveals *some* randomly-chosen user's password:

$$u \overset{?}{\leftarrow} \{0..999\};$$
$$Y = (u, X[u]);$$

When analyzed using min-entropy leakage, it leaks 10 bits, since it increases the vulnerability from 2^{-10000} to 2^{-9990}, a factor of 2^{10}. But this analysis is inadequate if we are concerned about the loss of *any* user's password, and not just about the loss of the *entire* database. Hence [11] defines a gain function g where

$$\mathcal{W} \quad = \quad \{(u, x) \mid 0 \le u \le 999 \text{ and } 0 \le x \le 1023\}$$

and

$$g((u, x), X) \quad = \quad \begin{cases} 1, & \text{if } X[u] = x \\ 0, & \text{otherwise.} \end{cases}$$

Note that this g gives the adversary a gain of 1 for correctly guessing the password of *any* user. When analyzed using g, the channel now increases the g-vulnerability from 2^{-10} to 1. But this is again a factor of 2^{10}, which means that the g-leakage is 10 bits, exactly the same as was the min-entropy leakage!

The point is that if we consider only the *ratios* of the vulnerabilities, an increase from 2^{-10000} to 2^{-9990} is exactly equivalent to an increase from 2^{-10} to 1. But, additively, the former increase is less than 2^{-9990}, a completely negligible quantity, while the latter is about 0.999, a huge difference.[8]

Another important motivation for considering additive leakage is the possibility of viewing vulnerabilities in *economic terms*. That is, we could define a gain function g so that the prior and posterior vulnerabilities would model the *economic value* to the adversary of its prior- and posterior knowledge about the secret. In this case, the additive leakage would represent the expected *monetary gain* from carrying out the attack represented by the channel C. Since making the attack might involve costs to the adversary, costs that he can predict, the additive leakage would then be crucial in his decision of whether the attack makes sense economically.

A technical issue to be considered concerns the *range* of gain functions. In [11], [23], gain functions are required to return values in the range $[0, 1]$. Note that in the multiplicative context, negative vulnerabilities would cause mathematical problems, and merely allowing a larger range $[0, a]$ would make no difference. In the additive context, however, it would make good sense to allow a wider range of gain values—large values could represent large economic gains, and negative values could represent *losses* (as in the gain function g_{tiger} of [11]).[9] We will say more about this issue in Section V below.

B. Expressiveness

Additive leakage is also interesting because some important *existing* leakages are special cases of additive g-leakages.

First, we find that we can express *guessing entropy* as a negative g-vulnerability. If we let $\mathcal{W}$ be the set of all *permutations* f of $\mathcal{X}$, we can then define the gain function

$$g(f, x) \quad = \quad -\, i$$

where i is the unique *index* of x within permutation f, assumed to range from 1 to n. (Note that this requires extending the allowed range of gain functions, since g has range in $[-n, -1]$.) With this definition, we find that $V_g[\pi] = -G[\pi]$ and $V_g[\pi, C] = -G[\pi, C]$, which means that the additive g-leakage is exactly the guessing-entropy leakage:

$$\mathcal{L}^+_g(\pi, C) \ = \ V_g[\pi, C] - V_g[\pi] \ = \ G[\pi] - G[\pi, C] \ .$$

We might of course wonder whether guessing-entropy leakage could also be expressed as a *multiplicative* g-leakage, but we suspect that in general this is not possible.

Finally, we can even express *Shannon leakage* (mutual information) as an additive g-leakage, although this does require further extension.

[7] The g-leakage is the *logarithm* of this ratio, which just changes the scale.

[8] On the other hand, increasing vulnerability from 0.000001 to 0.300001 is far more significant than increasing it from 0.4 to 0.7, even though both give additive leakage of 0.3. Perhaps we could say, roughly speaking, that leakage is 'significant' only if *both* the multiplicative- and additive leakage are 'large'.

[9] Theorem 4.1 of [11] says that we always have $V_g(\pi, C) \ge V_g(\pi)$. It remains true even if gain values are allowed to be negative, which means that additive leakage will always be non-negative.

To do this, we let $\mathcal{W}$ be the (uncountably infinite) set of all *probability distributions* ω on $\mathcal{X}$, and we define g by

$$g(\omega, x) \quad = \quad \lg \omega_x \ .$$

Since ω_x is in $[0, 1]$, this gain function has range $[-\infty, 0]$. The g-vulnerability associated with this gain function is

$$V_g[\pi] \quad = \quad \sup_{\omega} \sum_x \pi_x \lg \omega_x \ .$$

It turns out that V_g is realized when $\omega{=}\pi$, thanks to Gibbs' inequality [31], which implies that $V_g[\pi] = -H[\pi]$. This implies that the additive g-leakage for this g is exactly the Shannon leakage.

Using continuity arguments, we could restrict to a *countably* infinite $\mathcal{W}$ and a gain function with range $(-\infty, 0]$. Still, $\mathcal{W}$ needs to be infinite and g unbounded in order to capture Shannon leakage. Although such gain functions are interesting from the point of view of expressiveness, in this paper we restrict to finite sets of guesses $\mathcal{W}$.

IV. Gain-function algebra

Using gain functions to analyse leakage properties of channels can be simplified by establishing some general algebraic properties. Fix a channel C, prior π and gain function g. We define the following special g-functions, for $k{:}\,\mathbb{R}$:

$$\begin{array}{lcl} g_{\times k}(w, x) & = & g(w, x) \times k \\ g_{+k}(w, x) & = & g(w, x) + k \\ g_{+k@x'}(w, x) & = & g(w, x) + (k \text{ if } x{=}x' \text{ else } 0) \ . \end{array} \tag{5}$$

(Notice that in this section we are allowing completely unrestricted gain functions $\mathcal{W}{\times}\mathcal{X} \to \mathbb{R}$, for some finite $\mathcal{W}$.)

These new gain functions satisfy the following properties, which we prove in App. A:

$$\begin{array}{lcl} x\text{-Shift} & V_{g_{+k@x'}}[\pi, C] & = & V_g[\pi, C] + k\pi_{x'} \\ \text{Global Shift} & V_{g_{+k}}[\pi, C] & = & V_g[\pi, C] + k \\ \text{Scale} & V_{g_{\times k}}[\pi, C] & = & k \times V_g[\pi, C] \ , \\ & & & \text{for } k{\geq}0 \end{array} \tag{6}$$

In consequence, we have the following *invariance results* for leakages, simple corollaries of the above equalities:

Lemma 1: Scaling invariance For fixed gain function g and channel C, multiplicative leakage is invariant under positive Scale.

Lemma 2: Shifting invariance For fixed gain function g and channel C, additive leakage is invariant under Shift (of either kind).

V. Channel capacity

Shannon's original definition of channel capacity uses mutual information as the measure of information flow and maximises over all priors. The classic significance of the value obtained is that it provides a tight bound on the amount of information that can be transmitted reliably over the channel. In the security context, we look instead for bounds that tell us something about the "level of secrecy" of information, formulated for us in terms of relative (multiplicative) or comparative (additive) increases in g-vulnerability.

A useful notion of g-capacity would provide a robust measure of leakage that allowed meaningful predictions about how an attacker could use the channel for his gain. For robustness, we can consider universally quantifying over the prior π, over the gain function g, or over both. In quantifying over the prior π we are acknowledging that, in many situations, the prior is unknown and the assumption that it is uniform is not reasonable. (For example, large-scale studies have shown that password selection is not at all uniform [32].) In quantifying over the gain function g we are acknowledging that we might not know the value to the adversary of different sorts of partial information about the secret, neither now nor even in the future. (In medical research, for instance, new correlations between diet and susceptibility to disease are constantly discovered: although knowing someone's favorite dish might be relatively harmless today, it might not be so in the future.)

Since we can universally quantify over the prior π, over the gain function g, or over both, and since we are interested in both multiplicative- and additive g-leakage, we arrive at a total of six scenarios, which we write as follows:

$$\mathcal{L}_g(\forall, C), \mathcal{L}_\forall(\pi, C), \mathcal{L}_\forall(\forall, C), \mathcal{L}^+_g(\forall, C), \mathcal{L}^+_\forall(\pi, C), \mathcal{L}^+_\forall(\forall, C).$$

Each of them is of independent interest, and we define them precisely below.

For those scenarios that quantify over the gain function g, a first issue to be resolved is the set of gain functions that we allow. In previous work [11], [23], a gain function is required to have type $\mathcal{W} \times \mathcal{X} \to [0, 1]$ for some finite set $\mathcal{W}$. But, as motivated in Section III, with additive leakage we may wish to allow gain functions with a range different from $[0, 1]$. Can we find a suitable set of gain functions to use in both multiplicative and additive g-capacity?

Note first that *negative gain values* are mathematically problematic for multiplicative leakage. But they are inessential for additive leakage, since negative gain values can be eliminated by shifting (g is bounded since $\mathcal{W}$ is assumed finite), with no effect on additive leakage (Lemma 2). We therefore restrict our attention to non-negative gain functions.

Next consider an *unbounded range* $[0, \infty)$. Note that although any individual g is bounded (since $\mathcal{W}$ is finite), still an unbounded *range* allows to choose arbitrarily large gain functions when quantifying over them. For the multiplicative case this makes no difference, since any g (since it's bounded) can be scaled down to $[0, 1]$ without affecting the leakage (Lemma 2). On the other hand, scaling does affect the additive leakage, and the unbounded range of g can cause the additive capacities $\mathcal{L}^+_\forall(\pi, C)$, $\mathcal{L}^+_\forall(\forall, C)$ to be *ill defined*. Instead of bounding the range of g, we adopt the more permissive and mathematically attractive approach (for reasons explained in §VII) to restrict to *1-spanning gain functions*, whose gain values, for a given guess w, span a range of size at most 1:

Definition 3: 1-spanning A gain function g is *1-spanning* if $|g(w, x) - g(w, x')| \leq 1$ for all $x, x'{:}\,\mathcal{X}, w{:}\,\mathcal{W}$. □

We write $\mathbb{G}^1\mathcal{X}$ to denote the set of non-negative, 1-spanning gain functions on $\mathcal{X}$. When we formally define the different g-capacities below, this is the set of gain functions that we allow. Note that bounded range $[0, k]$ gain functions are a special case: they are scalar multiples of 1-spanning gain functions.

A. Multiplicative capacities defined

Recall that $\mathbb{D}\mathcal{X}$ is the set of distributions over a finite (and fixed) space of secrets $\mathcal{X}$. Both priors π and posterior distributions are elements of $\mathbb{D}\mathcal{X}$.

1) Fixed g, universally quantified π — $\mathcal{L}_g(\forall, C)$:

Definition 4: Multiplicative g-capacity For channel C and gain function g, the multiplicative g-capacity is given by

$$\mathcal{L}_g(\forall, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} \mathcal{L}_g(\pi, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} \lg(V_g[\pi, C]/V_g[\pi])\ .$$

□

2) Universally quantified g, fixed π — $\mathcal{L}_\forall(\pi, C)$:

Definition 5: Multiplicative π-capacity For channel C and prior π, the multiplicative π-capacity is given by

$$\mathcal{L}_\forall(\pi, C) \;=\; \sup_{g:\,\mathbb{G}^1\mathcal{X}} \mathcal{L}_g(\pi, C) \;=\; \sup_{g:\,\mathbb{G}^1\mathcal{X}} \lg(V_g[\pi, C]/V_g[\pi])\ .$$

□

It is worth pointing out that quantifying over gain functions in $\mathbb{G}^1\mathcal{X}$ is exactly equivalent here (in the multiplicative case) to quantifying over gain functions with range $[0, 1]$ (as used in previous work). First, gain functions with range $[0, 1]$ are in $\mathbb{G}^1\mathcal{X}$. Second, any $g\colon \mathbb{G}^1\mathcal{X}$ can be scaled to have range $[0, 1]$, with no effect on the multiplicative leakage.

3) Universally quantified g and π — $\mathcal{L}_\forall(\forall, C)$:

Definition 6: Multiplicative capacity For channel C the general multiplicative capacity is given by

$$\mathcal{L}_\forall(\forall, C) \;=\; \sup_{g:\,\mathbb{G}^1\mathcal{X}} \mathcal{L}_g(\forall, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} \mathcal{L}_\forall(\pi, C)\ .$$

□

B. Additive capacities defined

1) Fixed g, universally quantified π — $\mathcal{L}^+_g(\forall, C)$:

Definition 7: Additive g-capacity For channel C and gain function g, the additive g-capacity is given by

$$\mathcal{L}^+_g(\forall, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} \mathcal{L}^+_g(\pi, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} V_g[\pi, C] - V_g[\pi]\ .$$

□

This generalises Shannon capacity, but with Shannon entropy replaced by some g-vulnerability.

2) Universally quantified g, fixed π — $\mathcal{L}^+_\forall(\pi, C)$:

Definition 8: Additive π-capacity For channel C and prior π, the additive π-capacity is given by

$$\mathcal{L}^+_\forall(\pi, C) \;=\; \sup_{g:\,\mathbb{G}^1\mathcal{X}} \mathcal{L}^+_g(\pi, C) \;=\; \sup_{g:\,\mathbb{G}^1\mathcal{X}} V_g[\pi, C] - V_g[\pi]\ .$$

□

As mentioned above, restricting gain functions to $\mathbb{G}^1\mathcal{X}$ does make a difference in additive leakage. But note that a gain function with arbitrary range can always be expressed as a gain function in $\mathbb{G}^1\mathcal{X}$ followed by some shifting (which has no effect on the additive leakage) and scaling (which multiplies the additive leakage by a constant k). Hence we can, if desired, simply scale $\mathcal{L}^+_\forall(\pi, C)$ to account for a larger class of gain functions.

3) Universally quantified g and π — $\mathcal{L}^+_\forall(\forall, C)$:

Definition 9: Additive capacity For channel C the general additive capacity is given by

$$\mathcal{L}^+_\forall(\forall, C) \;=\; \sup_{g:\,\mathbb{G}^1 X} \mathcal{L}^+_g(\forall, C) \;=\; \sup_{\pi:\,\mathbb{D}\mathcal{X}} \mathcal{L}^+_\forall(\pi, C)\ .$$

□

C. A capacity scenario related to Dalenius's Desideratum

In addition to the six capacity scenarios considered above, we can also identify a seventh scenario that is considerably more "open ended".

In the context of *differential privacy* [25], [33], a major motivation is the possibility that the adversary knows interesting *correlations* among secrets. Recall Dwork's argument for the impossibility of *Dalenius's Desideratum*—she imagines the prior knowledge "Alan Turing is two inches taller than the average Lithuanian woman." With such knowledge, revealing information about Lithuanian women has the surprising effect of revealing information about Alan Turing.

And indeed we can consider such scenarios in the context of g-leakage. Suppose we have a secret X with prior π, and an adversary interested *only* in learning about X, as measured by a gain function $g : \mathcal{W}\times\mathcal{X} \to \mathbb{R}$. Now imagine a channel C from Y to Z, apparently having nothing at all to do with X. But still the adversary might know some interesting correlation between X and Y, expressed as a joint distribution Π on $\mathcal{X}\times\mathcal{Y}$ (which must induce marginal distribution π on $\mathcal{X}$). In this case, we can quantify how much C (surprisingly) teaches the adversary about X. For we can extend C to a channel C^* from (X, Y) to Z, defined by $C^*_{(x,y),z} = C_{y,z}$. (Note that C^* ignores the value of X.) And we can extend gain function g to a gain function g^* for (X, Y), defined by $g^*(w, (x, y)) = g(w, x)$; this reflects the fact that the adversary cares only about X. With this done, we find that $\mathcal{L}_{g^*}(\Pi, C^*)$ tells how much channel C leaks about X, given knowledge of correlations Π.

Note here that $V_{g^*}[\Pi] = V_g[\pi]$, but $V_{g^*}[\Pi, C^*]$ could be much larger. This means that the adversary's knowledge of the correlations Π between X and Y does not immediately teach him anything new about X. It is only observing the output of C^* that gives him new information about X—and this in spite of the fact that channel C^* completely ignores the value of X.

Hence it would be especially significant to establish upper bounds on the capacity (whether additive or multiplicative) on C^* in terms of the capacity of C—such bounds would give us a way to limit the harm that channel C could cause to X, no matter what correlations might subsequently be discovered to exist between X and Y. We give such bounds in Corollaries 22 and 23 in §IX below.

VI. Calculating multiplicative capacities

We now consider the efficiency of calculating capacities in our three multiplicative scenarios.

A. Fixed g, universally quantified π — $\mathcal{L}_g(\forall, C)$

This scenario appears quite challenging. For the identity gain function g_{id} (i.e. min-entropy leakage), we know from [20], [29] that a uniform prior always realizes the capacity. But this is not true in general. In [11, §V-D] are a channel and gain function whose multiplicative leakage is not maximized by a uniform prior but instead by $\pi = (1/2, 1/2, 0)$. It is unclear whether there is an efficient algorithm in this case.

B. Universally quantified g, fixed π — $\mathcal{L}_\forall(\pi, C)$

Here we have discovered a complete solution:

Theorem 10: For any full-support π, there exists a gain function g_π such that for any C we have $\mathcal{L}_{g_\pi}(\pi, C) = \mathcal{ML}(C)$. (As a consequence, $\mathcal{L}_{g_\pi}(\pi, C) = \mathcal{L}_\forall(\pi, C)$ since, by the Miracle Theorem of [11], we have $\mathcal{ML}(C) = \mathcal{L}_\forall(\forall, C)$.)

Proof: Let $g_\pi: \mathcal{X}\times\mathcal{X} \to [0, 1]$ be given by

$$g_\pi(x, x') \quad = \quad \begin{cases} a/\pi_x & \text{if } x{=}x' \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $a = \min_x \pi_x$. Note that g_π is independent of C; it can be written as a diagonal matrix whose entries are the *reciprocals* of the entries of π, normalized to be in $[0, 1]$. Intuitively, g_π cancels out the "hints" coming from a skewed prior, making *every* element of $\mathcal{X}$ equally good *a priori*.

Now straightforward calculations show that $V_{g_\pi}[\pi] = a$ and $V_{g_\pi}[\pi, C] = a \sum_y \max_x C_{x,y}$, giving $\mathcal{L}_{g_\pi}(\pi, C) = \lg \sum_y \max_x C_{x,y}$. The latter is the logarithm of the sum of the column maximums of C, which is $\mathcal{ML}(C)$ by the theorem of [20], [29]. ■

As an example, consider the imperfect cancer test from [30], with channel matrix

C	*positive*	*negative*
cancer	0.90	0.10
no cancer	0.07	0.93

For prior $\pi = (1/125, 124/125)$ the min-entropy leakage is 0, since the best guess is "no cancer", regardless of the test result. But with gain function

g_π	*cancer*	*no cancer*
cancer	1	0
no cancer	0	$1/124$

.

we find that the g_π-leakage is $\log 1.83 = \mathcal{ML}(C)$.

Finally, note that if π is not full support, we can just delete the rows of C corresponding to values of $\mathcal{X}$ not in the support.

C. Universally quantified g and π — $\mathcal{L}_\forall(\forall, C)$

Here we also have a complete solution. By the Miracle Theorem, multiplicative leakage is maximized by g_{id} and the uniform π, which give the min-capacity $\mathcal{L}_\forall(\forall, C)$. In fact, by §VI-B we can let π be any full-support prior and let g be the "reciprocals of π" gain function g_π.

VII. Calculating additive capacities

We now consider the efficiency of calculating capacities in our three additive scenarios.

A. Fixed g, universally quantified π — $\mathcal{L}^+_g(\forall, C)$

Here we know a number of interesting things. For g_{id} (which gives additive min-entropy leakage), it is shown in [29] that the additive leakage is maximized by some "corner point" distribution, i.e. one that is uniform on some *subset* of $\mathcal{X}$. If $|\mathcal{X}|{=}n$, there are $2^n{-}1$ corner-point distributions, making the maximum additive leakage computable. But, because the number of corner-point distributions is exponential in the size of C, this does not give an efficient algorithm. And the following theorem shows that an efficient algorithm probably does not exist: it is NP-complete to decide whether C's additive g_{id}-capacity exceeds a given threshold.[10]

Theorem 11: Given a channel matrix C and a threshold t, it is NP-complete to decide whether $\mathcal{L}^+_{g_{id}}(\forall, C) \geq t$.

Proof: Before beginning the proof, we establish some basic properties and notation for *corner-point* distributions. Given a non-empty subset $\mathcal{S}$ of $\mathcal{X}$, the corner-point distribution $\pi^\mathcal{S}$ gives probability $\frac{1}{|\mathcal{S}|}$ to the values in $\mathcal{S}$ and (therefore) gives probability 0 to the values in $\mathcal{X}{-}\mathcal{S}$. Now observe that

$$\begin{array}{ll} & \mathcal{L}^+_{g_{id}}(\pi^\mathcal{S}, C) \\ = & V[\pi^\mathcal{S}, C] - V[\pi^\mathcal{S}] \qquad \text{"}g_{id}\text{ gives ordinary vulnerability"} \\ = & \sum_{y:\,\mathcal{Y}} \max_{x:\,\mathcal{X}} \ \pi^\mathcal{S}_x C_{x,y} - \max_{x':\,\mathcal{X}} \pi^\mathcal{S}_{x'} \\ = & \sum_{y:\,\mathcal{Y}} \max_{x:\,\mathcal{S}} \ \frac{1}{|\mathcal{S}|} C_{x,y} - \frac{1}{|\mathcal{S}|} \\ = & \frac{1}{|\mathcal{S}|}(\sum_{y:\,\mathcal{Y}} \max_{x:\,\mathcal{S}} C_{x,y} - 1) \ . \end{array}$$

Notice that $\sum_{y:\,\mathcal{Y}} \max_{x:\,\mathcal{S}} C_{x,y}$ is simply the *sum of the column maximums of C, restricted to the rows in $\mathcal{S}$*. To simplify notation we let $\sigma_\mathcal{S}$ abbreviate this sum, allowing us to write

$$\mathcal{L}^+_{g_{id}}(\pi^\mathcal{S}, C) \quad = \quad \frac{\sigma_\mathcal{S} - 1}{|\mathcal{S}|} \ . \tag{8}$$

Now we proceed with the proof. By the result in [29], it is easy to see that the decision problem "$\mathcal{L}^+_{g_{id}}(\forall, C) \geq t$?" is in NP. For we can simply guess a non-empty subset $\mathcal{S}$ and verify that $\mathcal{L}^+_{g_{id}}(\pi^\mathcal{S}, C) \geq t$.

We complete the NP-completeness proof by showing a reduction from the *Set-Packing* problem [SP3] of [34], which is the problem of deciding whether a collection of sets $\mathcal{U} = \{U_1, \ldots, U_n\}$ contains k pairwise disjoint sets.

Suppose that the elements of the U_i are drawn from universe $\{a_1, \ldots, a_m\}$. We begin our reduction by constructing an $n \times m$ channel matrix[11] $C^\mathcal{U}$ where

$$C^\mathcal{U}_{i,j} \quad = \quad \begin{cases} \frac{1}{|U_i|} & \text{if } a_j \in U_i \\ 0 & \text{otherwise.} \end{cases}$$

Now notice that if $\mathcal{U}$ contains k pairwise disjoint sets, $k \geq 1$, then $C^\mathcal{U}$ contains k pairwise non-overlapping rows. This means that $C^\mathcal{U}$ is a *perfect* channel on the corresponding k inputs, since any possible output from those inputs will uniquely determine the input. Hence the additive leakage on the corner-point

[10] It is worth noting that this theorem is fundamentally different from the intractability results of [18]—those results assume that the channel is represented as a *program*, which inevitably leads to intractability of analysis. In contrast, our theorem is based on the *channel matrix*, which gives the channel's behaviour explicitly.

[11] Note that if $\emptyset \in \mathcal{U}$, we do not actually get a channel matrix, since $\emptyset$ leads to an all-zero row. But in this case notice that $\mathcal{U}$ contains k pairwise disjoint sets iff $\mathcal{U}{-}\{\emptyset\}$ contains $k{-}1$ pairwise disjoint sets.

distribution on those k inputs is $\frac{k-1}{k}$, since the vulnerability is increased from $\frac{1}{k}$ to 1.

We can see this more formally from equation (8) above. For if we let $\mathcal{S}$ be a set consisting of the indices of k pairwise disjoint sets in $\mathcal{U}$, then we have

$$\mathcal{L}^+_{g_{id}}(\pi^S, C^{\mathcal{U}}) \quad = \quad \frac{\sigma_{\mathcal{S}} - 1}{k}$$

where $\sigma_{\mathcal{S}}$ is the sum of the column maximums of $C^{\mathcal{U}}$, restricted to the rows in $\mathcal{S}$. But, since the rows in $\mathcal{S}$ are non-overlapping, we know that each nonzero entry in each of these rows is the *only* nonzero entry in its column (restricted to the rows in $\mathcal{S}$), and hence is included in the sum $\sigma_{\mathcal{S}}$. So, since each row sums to 1, we get that $\sigma_{\mathcal{S}} = k$, which gives

$$\mathcal{L}^+_{g_{id}}(\pi^S, C^{\mathcal{U}}) \quad = \quad \frac{k-1}{k} \ .$$

So far we have shown that if $\mathcal{U}$ contains k pairwise disjoint sets, then we have $\mathcal{L}^+_{g_{id}}(\forall, C^{\mathcal{U}}) \geq \frac{k-1}{k}$. But does the converse hold? Not necessarily, as shown by the following counterexample:

$$C^{\mathcal{U}} \quad = \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} .$$

In this case, $\mathcal{U}$ does *not* contain 2 disjoint sets, but nevertheless $\mathcal{L}^+_{g_{id}}(\forall, C^{\mathcal{U}}) \geq \frac{2-1}{2} = \frac{1}{2}$, since the corner-point distribution on all 3 rows gives additive leakage of $\frac{1}{3}(\frac{5}{2} - 1) = \frac{1}{2}$.

To correct this difficulty, we need to convert $\mathcal{U}$ into a channel matrix in a slightly more complicated way, one which boosts the "penalty" for including an overlapping row in a corner-point distribution.

To this end, let $p \geq 1$ be the maximum cardinality of any of the sets in $\mathcal{U}$.[12] Let $C^{\mathcal{U}}$ be embedded into a block matrix $\overline{C^{\mathcal{U}}}$ with p new rows and p new columns:

$$\overline{C^{\mathcal{U}}} \quad = \quad \begin{array}{|c|c|} \hline I_{p\times p} & 0_{p\times m} \\ \hline 0_{n\times p} & C^{\mathcal{U}} \\ \hline \end{array}$$

(Here I denotes the identity matrix, and 0 a zero matrix.)

Extending our previous calculation, note that if $\mathcal{U}$ contains k pairwise disjoint sets, then we have

$$\mathcal{L}^+_{g_{id}}(\forall, \overline{C^{\mathcal{U}}}) \quad \geq \quad \frac{p+k-1}{p+k} \ ,$$

since the p new rows are all non-overlapping.

But we can now argue that the converse also holds. For if $\mathcal{L}^+_{g_{id}}(\forall, \overline{C^{\mathcal{U}}}) \geq \frac{p+k-1}{p+k}$, then there exists a set $\mathcal{S}$ of indices such that $\mathcal{L}^+_{g_{id}}(\pi^S, \overline{C^{\mathcal{U}}}) \geq \frac{p+k-1}{p+k}$. Assume further that we choose $\mathcal{S}$ of *minimum size* that can achieve this additive-leakage threshold. Then we can argue by contradiction that the rows in $\mathcal{S}$ must be nonoverlapping, because if $\mathcal{S}$ contains a row that overlaps some other row of $\mathcal{S}$, then we can *delete* it without decreasing the additive leakage.

[12] In fact, the Set-Packing problem remains NP-complete even if all sets in $\mathcal{U}$ have cardinality at most 3 [34, p. 221].

For suppose that $|\mathcal{S}| = s$ for $s \geq 2$, and that $\mathcal{S}'$ is formed by deleting a row with some overlap with the other rows of $\mathcal{S}$. Recall that

$$\mathcal{L}^+_{g_{id}}(\pi^S, \overline{C^{\mathcal{U}}}) \quad = \quad \frac{\sigma_{\mathcal{S}} - 1}{s} \ ,$$

where $\sigma_{\mathcal{S}}$ is the sum of the column maximums of $\overline{C^{\mathcal{U}}}$, restricted to the rows in $\mathcal{S}$. Now observe that when we pass from $\mathcal{S}$ to $\mathcal{S}'$, we lose the contribution to $\sigma_{\mathcal{S}}$ from the deleted row; this loss is at most $\frac{p-1}{p}$, since the deleted row must have overlap of at least $\frac{1}{p}$ with the other rows of $\mathcal{S}$. Hence we have

$$\begin{array}{lll} & \mathcal{L}^+_{g_{id}}(\pi^{\mathcal{S}'}, \overline{C^{\mathcal{U}}}) & \\ = & (\sigma_{\mathcal{S}'} - 1)/(s-1) & \\ \geq & (\sigma_{\mathcal{S}} - \frac{p-1}{p} - 1)/(s-1) & \\ \geq & (\sigma_{\mathcal{S}} - \frac{p+k-1}{p+k} - 1)/(s-1) & \text{“}k \geq 0\text{”} \\ \geq & (\sigma_{\mathcal{S}} - \frac{\sigma_{\mathcal{S}}-1}{s} - 1)/(s-1) & \text{“}\mathcal{L}^+_{g_{id}}(\pi^S, \overline{C^{\mathcal{U}}}) \geq \frac{p+k-1}{p+k}\text{”} \\ = & (\frac{(\sigma_{\mathcal{S}}-1)(s-1)}{s})/(s-1) & \\ = & (\sigma_{\mathcal{S}} - 1)/s & \\ = & \mathcal{L}^+_{g_{id}}(\pi^S, \overline{C^{\mathcal{U}}}) \ . & \end{array}$$

Having proved that the rows in $\mathcal{S}$ are nonoverlapping, we now note that $\mathcal{S}$ must contain at least $p+k$ rows, since a nonoverlapping set of size u gives additive leakage $\frac{u-1}{u}$, which is less than $\frac{p+k-1}{p+k}$ unless $u \geq p+k$. And, finally, at least k of the nonoverlapping rows must come from $C^{\mathcal{U}}$, which implies that $\mathcal{U}$ contains k pairwise disjoint sets.

We have thus shown that $\mathcal{U}$ contains k pairwise disjoint sets iff $\mathcal{L}^+_{g_{id}}(\forall, \overline{C^{\mathcal{U}}}) \geq \frac{p+k-1}{p+k}$. Hence indeed Set Packing polynomial-time reduces to Maximum Additive Leakage. ■

Given the result for g_{id} specifically, we cannot of course expect the problem to be efficiently solvable for g in general. But we observe that we can use the linear-programming-based algorithm for Problem 6.1 of [11] to find a maximizing π: given channels C_1 and C_2, that algorithm finds a π that maximizes $V_g[\pi, C_1] - V_g[\pi, C_2]$. (It is aimed at testing whether C_1's g-leakage can ever exceed C_2's.) Here we can just let C_1 be C and let C_2 be $\mathbb{O}$, where $\mathbb{O}$ is a channel with just one column,[13] so that it satisfies *noninterference*, giving $V_g[\pi, \mathbb{O}]=V_g[\pi]$. Hence we get $V_g[\pi, C_1]-V_g[\pi, C_2] = V_g[\pi, C]-V_g[\pi] = \mathcal{L}^+_g(\pi, C)$.

To use linear programming, the algorithm in [11] applies a *strategy*-based formulation of posterior g-vulnerability. A strategy S *reifies* the choice of best guess as a channel from $\mathcal{Y}$ (the set of outputs of C) to $\mathcal{W}$ (the set of allowable guesses). We see then that we can formulate posterior vulnerability as

$$V_g[\pi, C] \quad = \quad \max_S \sum_{x,y,w} \pi_x C_{x,y} S_{y,w} g(w, x) \ .$$

But notice that if we then try to compute the maximization

$$\max_\pi (V_g[\pi, C_1] - V_g[\pi, C_2]) \ ,$$

we have *nested maxes* and a *quadratic* objective function, since the entries of both π and S are variables.

However, we can assume without loss of generality that the strategy S is *deterministic*, which allows the algorithm

[13] We write $\mathbb{O}$ because it is a right-zero of channel cascading.

to try explicitly *all* strategies S_1 for C_1 and S_2 for C_2, adding linear constraints to express that π is a prior for which S_1 and S_2 are optimal. In this way, it is able to find a π that maximizes additive g-leakage by solving $|\mathcal{W}|^{|\mathcal{Y}|+1}$ linear-programming problems. Of course, solving exponentially many linear-programming problems is not going to be feasible, except on very small channels.

But it appears that we can reformulate our problem as a single *Quadratic-Programming* problem, where the objective function is quadratic but the constraints are still linear. Quadratic programming cannot be solved efficiently in general, but there exist many mature quadratic-programming tools, and it appears likely that these could be applied fruitfully to the computation of $\mathcal{L}^+_g(\forall, C)$.

B. Universally quantified g, fixed π — $\mathcal{L}^+_\forall(\pi, C)$

We next study the complexity of calculating $\mathcal{L}^+_\forall(\pi, C)$. A first observation is that we can use the linear-programming-based algorithm for Problem 6.2 of [11] to find a maximizing g. This algorithm, very similar to the one described just above in §VII-A, requires the solution of exponentially many linear-programming problems.

But we find that we can do far better by exploiting some abstract mathematics. We show here that the quantification over $\mathbb{G}^1\mathcal{X}$ can be eliminated, because $\mathcal{L}^+_\forall(\pi, C)$ reduces to the well-known *Kantorovich distance* [35] between $[\pi]$ and $[\pi, C]$. Remarkably, this insight enables us to compute $\mathcal{L}^+_\forall(\pi, C)$ in time *linear* in the size of C.

We begin by recalling the Kantorovich distance between discrete probability distributions, and then we show how it applies in this context. Given a discrete probability distribution $\alpha : \mathbb{D}\mathcal{A}$ and random variable F recall that we write $\mathcal{E}_\alpha F$ for the expected value of $F : \mathcal{A} \to \mathbb{R}$ over α. Observe that when α is a point distribution centred at a that $\mathcal{E}_\alpha F = F(a)$.

Definition 12: Kantorovich Construction Let $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ be a metric on $\mathcal{A}$. We write $(\mathbb{D}\mathcal{A}, d)$ for the probability space on $\mathcal{A}$ determined by the sigma algebra generated by the open sets on $\mathcal{A}$. Given two discrete distributions $\alpha, \alpha' : \mathbb{D}\mathcal{A}$, the Kantorovich distance between them is

$$\mathbb{K}_{\mathbb{D}\mathcal{A}}(\alpha, \alpha') \quad = \quad \sup_{F\colon \mathcal{L}^1\mathcal{A}} |\mathcal{E}_\alpha F - \mathcal{E}_{\alpha'} F| \ ,$$

where $\mathcal{L}^1\mathcal{A}$ is the set of 1-Lipschitz functions on $\mathcal{A}$ [35]. □

The distance $\mathbb{K}_{\mathbb{D}\mathcal{A}}(\alpha, \alpha')$ on distributions is again a metric, and can be used via Def. 12 to construct $(\mathbb{D}^2\mathcal{A}, \mathbb{K}_{\mathbb{D}\mathcal{A}}(\alpha, \alpha'))$ [36]. In our context Def. 12 gives equivalent Kantorovich presentations for $\mathbb{D}\mathcal{X}$ and $\mathbb{D}^2\mathcal{X}$, allowing us to access very general calculational tools for evaluating channel capacities.

In our setting we start with the discrete metric on $\mathcal{X}$ to distinguish secrets because it simply treats all secrets as distinct. With this view we can use Def. 12 to derive the space of probability distributions over $\mathbb{D}\mathcal{X}$ together with an associated metric on $\mathbb{D}\mathcal{X}$. The next lemma recalls however that this derived metric in fact turns out to be (half of) the well-known Manhattan metric [37]. (See App. B.)

Lemma 13: Discrete Kantorovich metric Let $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ on $\mathcal{X}$ be the discrete metric, i.e. such that $d(x_1, x_2)$ is 0 or 1 according to $x_1 = x_2$. Then $\mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi, \pi')$ is half the Manhattan metric on $\mathbb{D}\mathcal{X}$: that is

$$\mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi, \pi') \quad = \quad \frac{1}{2} \sum_x |\pi'_x - \pi_x|$$

We next use the Manhattan metric to generate the probability space for $\mathbb{D}^2\mathcal{X}$, together with its associated Kantorovich distance between hypers. Our first observation is that by regarding V_g as a 1-Lipschitz function on $\mathbb{D}\mathcal{X}$, we immediately obtain that $\mathcal{L}^+_\forall(\pi, C)$ is well-defined. In particular the connection between leakage calculations and expected values is given by the equality $V_g[\pi, C] = \mathcal{E}_{[\pi,C]} V_g$, a technical result that we prove in App. A as Lem. 24.

Lemma 14: Well defined capacity $\mathcal{L}^+_\forall(\pi, C)$ is well-defined, and satisfies $\mathcal{L}^+_\forall(\pi, C) \leq \mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C])$.

Proof: For any (1-spanning) $g : \mathbb{G}^1\mathcal{X}$, the gain-function test V_g regarded as a function from $\mathbb{D}\mathcal{X} \to \mathbb{R}$, which takes a distribution $\pi : \mathbb{D}\mathcal{X}$ and maps it to $V_g[\pi]$, is a 1-Lipschitz function (Lem. 26 of App. A). This implies from Def. 12 that $\mathcal{L}^+_\forall(\pi, C)$ can be no more than the Kantorovich distance between $[\pi]$ and $[\pi, C]$. ■

An important property of the Kantorovich distance between probability distributions is that it has a dual formulation as the Wasserstein metric, commonly known as the "earth-moving metric" [35]; this in particular gives access to efficient algorithms for computing it, and strengthening the upper bound in Lem. 14. Earth moving formalises the idea of measuring the cost of transforming one distribution into another; it uses the underlying Kantorovich distance to define the cost function, which is then averaged over a strategy for re-distributing the probability mass in the resulting transformation. The earth-moving distance is defined by the minimum such cost.

We specialise the Wasserstein metric to the case of a discrete distribution on a set $\mathcal{D}$, with some metric $d : \mathcal{D}^2 \to \mathbb{R}$.

Definition 15: The earth-moving metric on $\mathbb{D}\mathcal{D}$ Consider two distributions $\sigma, \tau : \mathbb{D}\mathcal{D}$ with σ the "source" distribution and τ the "target" distribution. Then an earth-moving *strategy* on $\mathbb{D}\mathcal{D}$ is a (joint) distribution in $\mathbb{D}(\mathcal{D}^2)$ whose two marginals are σ and τ. Write $\mathcal{S}_{\sigma,\tau}$ for the set of such strategies.

The earth-moving distance between σ and τ is then the infimum over all such strategies $S : \mathcal{S}_{\sigma,\tau}$ of $\mathcal{E}_S d$. Informally it is the least sum of products "amount of earth moved" times "distance moved". More precisely we define

$$\mathbb{W}_{\mathbb{D}\mathcal{D}}(\sigma, \tau) \quad = \quad \inf_{S\colon \mathcal{S}_{\sigma,\tau}} \mathcal{E}_S d \ .$$ □

Now we specialise the Wasserstein distance to the same domain as our definition Def. 15 of the Kantorovich distance, that is taking $\mathcal{D}$ to be $\mathbb{D}\mathcal{X}$, so that $\mathbb{D}\mathcal{D}$ becomes $\mathbb{D}^2\mathcal{X}$, and the metric d on $\mathbb{D}\mathcal{X}$ is half the Manhattan distance. Thus $\sigma, \tau : \mathbb{D}\mathcal{D}$ become hypers in $\mathbb{D}^2\mathcal{X}$ and, furthermore, those hypers have finite support (since we are only considering finite channels). That means that the strategies S considered in Def. 15 can also be limited to discrete distributions for our purposes.

That done, we conclude that the two metrics are the same:

Theorem 16: Kantorovich-Rubinstein [35] The Kantorovich distance and the earth-moving distance between hypers are the same.

Thm. 16 applied here will allow us to show that channel capacities are in fact the same as Kantorovich distances between the prior, considered as a point hyper in $\mathbb{D}^2\mathcal{X}$, and the hyper produced by the channel acting on that prior.

Theorem 17: Additive π-capacity as Kantorovich distance Let $C: \mathcal{X} \to \mathcal{Y}$ be a channel and π a prior on $\mathcal{X}$. Then the channel capacity of C at π is equal to the Kantorovich distance between the hypers $[\pi]$ and $[\pi, C]$. That is, we have

$$\mathcal{L}^+_\forall(\pi, C) \quad = \quad \mathbb{W}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C]) \quad = \quad \mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C]) \ .$$

(Note that the second equality is immediate from Thm. 16.)

Proof: Adopt the notation of Def. 15 so that our source distribution is the point hyper $\Sigma{=}[\pi]$ and our target is $\mathsf{T}{=}[\pi, C]$, where we are using capital letters here (instead of σ, τ as earlier) to remind us that the source and target are hypers.

Because Σ is a point distribution (centred on the prior π), there is only one strategy in $\mathcal{S}_{\Sigma,\mathsf{T}}$ that has the appropriate marginals $\Sigma{=}[\pi]$ and $\mathsf{T}{=}[\pi, C]$. Given any two marginals Σ, T with Σ being a point distribution, the set $\mathcal{S}_{\Sigma,\mathsf{T}}$ of all joint distributions with the given marginals Σ, T contains only one member: *any* discrete joint distribution whose left (say) marginal is 1 at some x and zero elsewhere is expressible as a matrix where *only* the x-row is non-zero — and that row is the right marginal exactly. Put informally, if there is only one source "pile of earth," then any earth in any target pile can have come only from that single source.

For general $\sigma, \tau: \mathbb{D}\mathcal{X}$ the optimal strategy is given by

$$S_{\sigma,\tau} \quad = \quad \mathsf{T}_\tau \text{ if } \sigma{=}\pi \text{ else } 0 \ , \tag{9}$$

so that *all* earth transfers from $\lceil\Sigma\rceil$ start from π; and the proportion of earth "at π" that is moved to any posterior τ in $\lceil\mathsf{T}\rceil$ is the probability $[\pi, C]_\tau$ $(= \mathsf{T}_\tau)$ that the hyper $[\pi, C]$ assigns to that τ.

Thus we calculate

$$\begin{array}{lll}
& \mathbb{W}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C]) & \\
= & \min_{S : \mathcal{S}_{\Sigma,\mathsf{T}}} \mathcal{E}_S d & \text{“}d \text{ is underlying metric; Def. 15”} \\
= & \mathcal{E}_S d & \text{“take unique } S \text{ in } \mathcal{S} \text{ as at (9)”} \\
= & \mathcal{E}_{[\pi,C]} d_\pi & \text{“define } d_\sigma(\tau) = d(\sigma, \tau) \text{ to allow Currying;} \\
 & & \quad S \text{ at (9) has point left-marginal } [\pi]\text{”} \\
= & V_g[\pi, C] - 1 & \text{“define } g \text{ so that } V_g[\tau] = d_\pi(\tau){+}1\text{; see below”} \\
= & V_g[\pi, C] - V_g[\pi] & \text{“}V_g[\pi] = d_\pi(\pi){+}1 = 1\text{”} \\
\leq & \mathcal{L}^+_\forall(\pi, C) & \text{“Def. 8”} \\
\leq & \mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C]) \ , & \text{“Lem. 14”}
\end{array}$$

whence the result follows because we have sandwiched $\mathcal{L}^+_\forall(\pi, C)$ between $\mathbb{W}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C])$ and $\mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C])$ — which from Thm. 16 are equal.

For the "see below" we exhibit a $\mathcal{W}$ together with a non-negative and 1-spanning gain function $g: \mathcal{W}{\times}\mathcal{X} \to \mathbb{R}$ such that $V_g[\tau] = d(\pi, \tau){+}1$. To do this we observe that

$$\begin{array}{lll}
& d(\pi, \tau) + 1 & \\
= & \mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi, \tau) + 1 & \text{“definition } d\text{”} \\
= & \sum_x |\tau_x{-}\pi_x|/2 + 1 & \text{“Lem. 13”} \\
= & \max_w \sum_x (\tau_x{-}\pi_x) w(x)/2 + 1 & \text{“define } \mathcal{W} = \mathcal{X}{\to}\{-1, 1\}\text{”} \\
= & \max_w \sum_x \tau_x g(w, x) & \text{“define } g(w, x) = \\
 & & \quad 1 + (w(x) + \sum_{x'} \pi_{x'} w(x'))/2\text{; see App. C”} \\
= & V_g[\tau] \ . & \text{“definition } V_g \text{ for this } \mathcal{W}\text{”}
\end{array}$$

We now have

$$\mathcal{E}_{[\pi,C]} d_\pi \ = \ \mathcal{E}_{[\pi,C]}(V_g{-}1) \ = \ (\mathcal{E}_{[\pi,C]} V_g) - 1 \ = \ V_g[\pi, C] - 1 \ ,$$

where the last equality follows from Lem. 24 of App. A. We note that g is at least zero, since $[w(x) + \sum_{x'} \pi_{x'} w(x')]/2$ has minimum value at least -1, and finally g is 1-spanning since

$$|g(w, x) - g(w, x')| \quad = \quad |w(x) - w(x')|/2 \quad \leq \quad 1 \ .$$

■

We can now compute $\mathcal{L}^+_\forall(\pi, C)$ efficiently.

Corollary 18: Efficient calculation of additive capacity Given channel $C: \mathcal{X} \to \mathcal{Y}$ and prior $\pi: \mathbb{D}\mathcal{X}$, the additive capacity $\mathcal{L}^+_\forall(\pi, C)$ has complexity $O(|\mathcal{X}||\mathcal{Y}|)$.

Proof: From Thm. 17 we have the equality $\mathcal{L}^+_\forall(\pi, C) = \mathbb{W}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C])$, whose rhs. is the optimal cost of earth-moving $[\pi]$ to $[\pi, C]$. In the proof of that theorem we see that the cost is the expected value of the Kantorovich distance between prior π and the posteriors $p_{X|y}$ in $[\pi, C]$ (i.e. $\mathcal{E}_{[\pi,C]} d_\pi$).

That is, the optimal earth-move is calculated from the $\mathcal{Y}$-marginal p_Y of the joint distribution Π generated from π and C, using the Manhattan distance (Lem. 13) between π and each posterior $p_{X|y}$. Thus the computational cost is given by the cost of computing the $\mathcal{Y}$-marginal plus the cost of computing the average Manhattan distance, if we know the marginal.

Computing the $\mathcal{Y}$-marginal takes $|\mathcal{X}||\mathcal{Y}|$ time and, once done, computing each $p_{X|y}$ takes $|\mathcal{X}|$ time. Similarly, computing the Manhattan distance between each π and $p_{X|y}$ takes $|\mathcal{X}|$ time. Finally the cost of computing the expected optimal earth move takes $|\mathcal{X}||\mathcal{Y}|$ time, giving a total overall complexity of $O(|\mathcal{X}||\mathcal{Y}|)$. ■

C. *Universally quantified g and π — $\mathcal{L}^+_\forall(\forall, C)$*

For additive capacity (unlike multiplicative), it is not the case that the prior that optimises the capacity is known (or that it can be easily calculated). Nevertheless, we have made some progress: thanks to Thm. 17 we can remove one level of quantification over the gain functions, still leaving an optimisation problem for computing the capacity but it is less complex than before. Let $C: \mathcal{X} \to \mathcal{Y}$ be a channel. We can reformulate Def. 9 as follows:

$$\begin{array}{lll}
& \sup_\pi \mathcal{L}^+_\forall(\pi, C) & \\
= & \sup_\pi \mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi], [\pi, C]) & \text{“Thm. 17”} \\
= & \sup_\pi \sum_{y,x} \pi_x |C_{xy} - \sum_{x'} C_{x'y} \pi_{x'}|/2 \ . & \text{“Def. 15; see App. B”}
\end{array}$$

This has the form of an optimisation over variables π_x constrained by $\pi_x \geq 0$ and $\sum_x \pi_x = 1$. The $O(|\mathcal{X}||\mathcal{Y}|)$ occurrences of absolute value seem potentially to make the calculation more complicated in general than e.g. a quadratic optimisation.

VIII. Example calculations

Here are some examples of the techniques of Section VII.

A. Additive capacity, fixed g — $\mathcal{L}^+_g(\forall, C)$

Let channel C and gain function g be

C	y_1	y_2	y_3
x_1	0.9	0.1	0
x_2	0.1	0.7	0.2
x_3	0.1	0.5	0.4

g	x_1	x_2	x_3
w_1	0.7	0.1	0.4
w_2	0.1	0.4	0.9

On this example, the algorithm described in §VII-A needs to solve $|\mathcal{W}|^{|\mathcal{Y}|+1} = 2^4 = 16$ linear programming problems. It finds that $\mathcal{L}^+_g(\forall, C) = {}^{12}/_{55}$, realized on prior $\pi = ({}^5/_{11}, 0, {}^6/_{11})$.

Using g_{id}, in contrast, the algorithm needs to solve $3^4 = 81$ linear programming problems, and it finds that $\mathcal{L}^+_{g_{id}}(\forall, C) = {}^2/_5$, realized on prior $\pi = ({}^1/_2, {}^1/_2, 0)$.

B. Additive capacity, fixed π — $\mathcal{L}^+_\forall(\pi, C)$

Cor. 18 and Thm. 17 provide two equivalent ways to calculate $\mathcal{L}^+_\forall(\pi, C)$, which we illustrate using the cancer channel.

Given general prior π such that $\pi_{x_1} = a$ and $\pi_{x_2} = 1{-}a$, we use Cor. 18 to calculate $\mathcal{L}^+_\forall(\pi, C)$ using the expected earth moving metric as follows. We calculate first the $\mathcal{Y}$ marginal probabilities: the chance of observing a positive result is $a{\times}0.9 + (1{-}a) \times 0.07 = 0.07{+}0.83a$, and for negative result it is $0.93{-}0.83a$. The corresponding posterior for observing a positive result, which we denote ν^p, is given by $\nu^p_{x_1} = 0.9a/(0.07{+}0.83a)$ and $\nu^p_{x_2} = 0.07(1{-}a)/(0.07{+}0.83a)$. There is a similar calculation for the negative posterior ν^n. Next Cor. 18 directs us to calculate the Manhattan distance between π and ν^p and π and ν^n, and then to take the expected value over the above marginal. The distance between π and ν^p is $(|a{-}\nu^p_{x_1}| + |(1{-}a){-}\nu^p_{x_2}|)/2$, which simplifies to $0.83a(1{-}a)/(0.07{+}0.83a)$. Similarly the distance between π and ν^n is $0.83a(1{-}a)/(0.93{-}0.83a)$. Finally taking the expected move we obtain $2 \times 0.83a(1{-}a) = 1.66a(1{-}a)$. For the prior with $a = \frac{1}{125}$ this is approximately 0.0132.

An alternative but equivalent way to calculate $\mathcal{L}^+_\forall(\pi, C)$ is given in the proof of Thm. 17, where we construct an optimising gain function that achieves additive capacity. Working through the definitions for this example we see that $\mathcal{W}$ is given by the four functions $\mathcal{X}{\to}\{-1, 1\}$. For each such function we construct $g(w, x) = [w(x) - \sum_{x'} \pi_{x'} w(x')]/2{+}1$. Two of the four possibilities result in constants for $g(w, x)$ (e.g. assigning 1 everywhere as x varies), and this corresponds to an overall additive leakage of 0. Hence only two functions from $\mathcal{W}$ correspond to options that participate in maximising over all leakage calculations. One maps x_1 to 1 and x_2 to -1, and the other is its "reflection" (multiply by -1). For the first, we see that $g(w, x_1) = 2{-}a$, and $g(w, x_2) = 1{-}a$. We summarise the situation by the following matrix:

g	*cancer*	*no cancer*
don't treat	a	$a{+}1$
treat	$2{-}a$	$1{-}a$

One possible operational interpretation for this gain function is as follows. The result of the cancer test provides information for the health practitioner and the patient in order to make an informed decision about the best treatment plan. In this case the plan would normally consist of whether to recommend some form of direct medical intervention. The two gain strategies can therefore be thought of as putting a "cost" on either treating, or not treating the patient; that cost has either a high- or low gain for the patient depending on whether cancer is actually present. In the scenario where the patient is given direct medical intervention (row labelled "treat" in the matrix) there is therefore a strong positive gain for the patient who does have cancer, but relatively a low gain for the patient who does not have cancer but is given treatment anyway. Similarly in the scenario corresponding to deciding not to intervene (row labelled "don't treat"), there is a relatively low gain for the patient who does have cancer, but a positive gain for the patient who does not. With this interpretation, the additive capacity of $1.66a(1{-}a)$ gives the expected overall gain for the patient, and corresponds to the strategy of always offering treatment whenever the test is positive, but never when the test is negative.

We note that this gain function is 1-spanning and non-negative; in a more realistic costing the difference between treating and not-treating would be obtained by shifting and scaling (see §IV), with the former having the option to assign negative gains in the case of no treatment being given when in fact the cancer is present.

C. General additive capacity — $\mathcal{L}^+_\forall(\forall, C)$

Computing the general additive capacity $\mathcal{L}^+_\forall(\forall, C)$ involves solving the optimisation problem set out at §VII-C. For $|\mathcal{X}|{=}2$, this problem can be solved exactly and indeed the optimal prior is always uniform. This can be seen directly in the cancer channel above, where the capacity is the maximum value of $1.66a(1{-}a)$, where $0{\le}a{\le}1$, giving a value of $1.66/4 = 0.415$.

When $|\mathcal{X}|{>}2$, however, the optimal prior does not seem to be uniform in all cases: a graph plotted for the channel

C	y_1	y_2	y_3
x_1	0.4	0.5	0.1
x_2	0.6	0.1	0.3
x_3	0.2	0.2	0.6

shows that the optimal additive capacity occurs at approximately $(0.44, 0.09, 0.47)$.

Finally, let C be the *identity matrix* on $\mathcal{X}$ (i.e. a perfect channel). Given prior $\pi{:}\,\mathbb{D}\mathcal{X}$, the hyper $[\pi, C]$ produced by C consists of point distributions $[x]$, each one with probability π_x. Now for any $x{:}\,\mathcal{X}$ the earth-moving distance between π and $[x]$ is $1{-}\pi_x$, since all of the mass of π needs to be moved onto x, except for what is already there. Hence the overall earth-moving distance, obtained by considering all x's in $\mathcal{X}$, is

$$\begin{array}{rcl} & \sum_x (\textit{amount moved to } [x])(\textit{distance moved to } [x]) & \\ = & \sum_x \pi_x(1{-}\pi_x) \quad = \quad 1 - \sum_x (\pi_x)^2 \ , & \end{array}$$

and that is maximized by a uniform π, giving just $1{-}{}^1/_{|\mathcal{X}|}$. (Interestingly, this is also the maximum additive leakage with respect to g_{id}.)

IX. Bounding "Dalenius" Leakage

Here we consider the "Dalenius" scenario described in §V-C, establishing bounds on the leakage of a secret X that may be caused by a channel C from Y to Z, no matter what correlations may exist between X and Y. Our results, given in

Cor. 22 and Cor. 23 below, follow from general results that we now develop for capacities of *cascades* of channels.

In the following lemmas etc. we assume that channel $R{:}\mathcal{X}{\to}\mathcal{Z}$ is a cascade of channels $S{:}\mathcal{X}{\to}\mathcal{Y}$ and $T{:}\mathcal{Y}{\to}\mathcal{Z}$. The proofs will be presented for additive capacity; but, once established additively, the results hold easily for multiplicative capacity as well (Cor. 23).

Lemma 19: Left-capacity bound for cascade For any prior $\rho{:}\mathbb{D}\mathcal{X}$ and gain function g on $\mathcal{X}$ we have the inequality $\mathcal{L}^+_g(\rho,R) \leq \mathcal{L}^+_g(\rho,S)$.

Proof: This is a direct consequence of properties of a relation comparing information-flow between channels, described in recent and related work [7], [11], [23], [37]; it is a partial order which here we will call "secures."[14] The order has the property that, for any two channels C, C', whenever C' secures C, we have $\mathcal{L}^+_g(\pi,C'){\leq}\mathcal{L}^+_g(\pi,C)$ for all π, g (a corollary of [11, Thm. 6.2 generalised] and of [37, Thm. 8]). Furthermore, and conveniently for our purposes here, one of several (equivalent) presentations of that order [11] is as follows:

> C' secures C just when for some conformal channel M we have $C'{=}CM$ (with the juxtaposition representing matrix multiplication, i.e. channel cascade).

Thus we can calculate

$$\begin{array}{ll} & R = ST \\ \Rightarrow & R \text{ secures } S \qquad \text{"definition of information-flow order [11]"} \\ \Rightarrow & V_g[\rho,R] \leq V_g[\rho,S] \qquad \text{" [11, Thm. 6.2], above"} \\ \Rightarrow & V_g[\rho,R]{-}V_g[\rho] \leq V_g[\rho,S]{-}V_g[\rho] \\ \Rightarrow & \mathcal{L}^+_g(\rho,R) \leq \mathcal{L}^+_g(\rho,S)\ . \qquad \text{"definition } \mathcal{L}^+_g(\forall,)\text{"} \end{array}$$

■

Lemma 20: Right-capacity bound for cascade For any prior $\rho{:}\mathbb{D}\mathcal{X}$ we have $\mathcal{L}^+_\forall(\rho,R) \leq \mathcal{L}^+_\forall(\tau,T)$, where distribution $\tau{:}\mathbb{D}\mathcal{Y}$ is the distribution of outputs in $\mathcal{Y}$ produced by channel S with prior ρ.

Proof: Let gain function g on $\mathcal{X}$ be arbitrary. We calculate

$$\begin{array}{ll} & V_g[\rho,R]\ . \\ = & \sum_z \max_w \sum_x \rho_x R_{x,z} g(w,x) \qquad \text{"definition } V\text{"} \\ = & \sum_z \max_w \sum_x \rho_x (\sum_y S_{x,y} T_{y,z}) g(w,x) \qquad \text{"}R = ST\text{"} \\ = & \sum_z \max_w \sum_y T_{y,z} \sum_x \rho_x S_{x,y} g(w,x) \qquad \text{"}\textstyle\sum_x \sum_y = \sum_{x,y}\text{"} \\ = & \qquad \text{"define } \tau,h \text{ so that } \tau_y h(w,y) = \textstyle\sum_x \rho_x S_{x,y} g(w,x)\text{"} \\ & \sum_z \max_w \sum_y T_{y,z} \tau_y h(w,y) \\ = & V_h[\tau,T]\ . \qquad \text{"definition } V\text{"} \end{array}$$

The key move obviously is the definition of τ, h in terms of ρ, g. We use a "temporary" $\Pi{:}\mathbb{D}(\mathcal{X}{\times}\mathcal{Y})$ like this:

$$\begin{array}{ll} & \Pi_{x,y} = \rho_x S_{x,y} \\ \text{then} & \tau_y = \textstyle\sum_x \Pi_{x,y} \\ \text{and} & h(w,y) = \textstyle\sum_x \Pi_{x,y} g(w,x)/\tau_y\ . \end{array}$$

Note that we have that h is 1-spanning, because g is. For each w there are L_w, H_w with $H_w{-}L_w{\leq}1$ and $L_w{\leq}g(w,x){\leq}H_w$ for all x, so that $L_w = \sum_x \Pi_{x,y} L_w/\tau_y \leq h(w,y) \leq \sum_x \Pi_{x,y} H_w/\tau_y = H_w$, from the definition of τ_y.

Now if you set $T{=}\mathbb{O}$ in the above, [15] then $R{=}S\,\mathbb{O}{=}\mathbb{O}$ also, and you get

$$V_g[\rho] \quad = \quad V_g[\rho,\mathbb{O}] \quad = \quad V_h[\tau,\mathbb{O}] \quad = \quad V_h[\tau]\ ,$$

so that for any ρ, g there are τ, h with $\mathcal{L}^+_g(\rho,R) = \mathcal{L}^+_h(\tau,T)$. And so $\mathcal{L}^+_\forall(\rho,R) \leq \mathcal{L}^+_\forall(\tau,T)$ as claimed. ■

Those two lemmas give us this theorem about cascades:

Theorem 21: Capacity of cascades If $R{:}\mathcal{X}{\to}\mathcal{Z}$ is a cascade of channels $S{:}\mathcal{X}{\to}\mathcal{Y}$ and $T{:}\mathcal{Y}{\to}\mathcal{Z}$ (as above) then

$$\mathcal{L}^+_\forall(\forall,R) \quad \leq \quad \mathcal{L}^+_\forall(\forall,S) \ \min\ \mathcal{L}^+_\forall(\forall,T)\ ,$$

where we recall that $\mathcal{L}^+_\forall(\forall,\cdot)$ is a supremum over all priors and all (1-spanning) gain-functions.

Proof: The result follows directly from Lemmas 19,20: simply take suprema on both sides to give $\mathcal{L}^+_\forall(\forall,R) \leq \mathcal{L}^+_\forall(\forall,S)$ and $\mathcal{L}^+_\forall(\forall,R) \leq \mathcal{L}^+_\forall(\forall,T)$ respectively. ■

Now we can use Lem. 20 to give a bound on any "collateral" Dalenius leakage about some other $\mathcal{X}$ even when there is no direct link between C and $\mathcal{X}$, i.e. if we are given only that there is some interesting joint distribution on $\mathcal{X}$ and $\mathcal{Y}$ that is somehow known to the observer.

Corollary 22: Dalenius bound Consider as above a channel $C{:}\mathcal{Y}{\to}\mathcal{Z}$ and let $\pi{:}\mathbb{D}\mathcal{Y}$ a prior for C. Assume further that there is some other space $\mathcal{X}$ and a known joint distribution $\Pi{:}\mathbb{D}(\mathcal{X}{\times}\mathcal{Y})$ describing a correlation between $\mathcal{X}$ and $\mathcal{Y}$. Note that π must in that case be the $\mathcal{Y}$-marginal of Π.

Imagine an extended channel C^* from $\mathcal{X}{\times}\mathcal{Y}$ to $\mathcal{Z}$ defined by $C^*_{(x,y),z} = C_{y,z}$. Then (addressing the concern raised in §V-C) the "collateral leakage" that the imaginary C^* causes wrt. the arbitrary $\mathcal{X}$ can be bounded independendently of $\mathcal{X}$ and the joint distribution Π it shares with $\mathcal{Y}$. The bound depends only on C and π: we have $\mathcal{L}^+_\forall(\Pi,C^*) \leq \mathcal{L}^+_\forall(\pi,C)$, where the right-hand side –the upper bound– does not depend on $\mathcal{X}$ and its relationship with $\mathcal{Y}$ beyond the (necessary) fact that π is the $\mathcal{Y}$-marginal of Π.

Proof: Construct a deterministic "projection channel" $P{:}\mathcal{X}{\times}\mathcal{Y}{\to}\mathcal{Y}$ so that $P_{(x,y),y'} = (1 \text{ if } y{=}y' \text{ else } 0)$. Observe then that $C^*{=}PC$, that π is the $\mathcal{Y}$-output of C^* wrt. prior Π, and apply Lem. 20. ■

Note from our conclusion $\mathcal{L}^+_\forall(\Pi,C^*) \leq \mathcal{L}^+_\forall(\pi,C)$ we have in particular, for any "$\mathcal{Y}$-ignoring" $g^*(w,(x,y)) = g(w,x)$ encoding an attacker's interest in $\mathcal{X}$ alone (i.e. not in $\mathcal{Y}$ at all), that the bound $\mathcal{L}^+_{g^*}(\Pi,C^*) \leq \mathcal{L}^+_\forall(\pi,C)$ applies.

Finally, we note that the same results hold for multiplicative capacities:

Corollary 23: For channels $R{:}\mathcal{X}{\to}\mathcal{Z}$ and $S{:}\mathcal{X}{\to}\mathcal{Y}$ and $T{:}\mathcal{Y}{\to}\mathcal{Z}$, with $R{=}ST$, we have

$$\begin{array}{ll} \textit{Left-capacity bound} & \mathcal{L}_g(\rho,R) \leq \mathcal{L}_g(\rho,S) \\ \textit{Right-capacity bound} & \mathcal{L}_\forall(\rho,R) \leq \mathcal{L}_\forall(\tau,S) \\ \textit{Capacity of cascades} & \mathcal{L}_\forall(\forall,R) \leq \mathcal{L}_\forall(\forall,S) \ \min\ \mathcal{L}_\forall(\forall,T) \\ \textit{Dalenius bound} & \mathcal{L}_\forall(\Pi,C^*) \leq \mathcal{L}_\forall(\pi,C)\ , \end{array}$$

[14] In [11], [23] we wrote *more-secure* $\sqsubseteq_\circ$ *less-secure*, by analogy with refinement of partitions making them *less* secure. In [7], [37] we wrote *less-secure* $\sqsubseteq$ *more-secure*, since refinement of programs makes them *more* secure. This unlucky opposition is why we just call it "secures" here.

[15] Recall that that $\mathbb{O}$ is a single-column channel, a right-zero of cascading.

where $g, \rho, \tau, C^*, \Pi, \pi$ are as defined above.

Proof: The proofs of Lems. 19,20, of Thm. 21 and of Cor. 22 hold *mutatis mutandis* for multiplicative capacities.[16] ■

X. Related work

As early as the 1950's, Shannon himself observed [39] that "[i]t is hardly expected that a single concept of information would satisfactorily account for the numerous possible applications of [information theory]." In the 90's, Cachin advocated [40] that information metrics for security, or entropies, not only include a way to calculate some numeric value but also offer an *operational interpretation*, which describes what aspect of interest is being quantified. In recent years various metrics have been proposed for capturing the leakage of information in different operational scenarios. Work has been done on Shannon entropy [1], [4], [41]–[45] guessing entropy [28], [46] and vulnerability or Bayes risk [6], [7], [29].

This plurality of metrics motivated researchers to look for some essential notion of information that would permeate all the various notions of entropy. Originally proposed by Landauer and Redmond [47], the Lattice of Information has been studied as an underlying algebraic structure for deterministic channels. And indeed, for deterministic channels, Yasuoka and Terauchi [18] showed that the orderings on induced by Shannon entropy, guessing entropy, and vulnerability are all equivalent, and Malacaria [46] showed that they coincide with the partition-refinement order in the lattice of partitions induced by such channels.

However meaningful, these traditional entropies fail to capture a range of other relevant scenarios of interest, and the g-leakage framework proposed by Alvim et al. [11], and McIver el al. [7], models the operational scenario in which the channel executes by means of gain-functions. Given the wide range of possible gain functions and contexts, a robust comparison between channels is made via a strong g-leakage pre-order, which requires a channel never leak more than another, for *any* prior and gain function. It was shown in [7], [11] that composition refinement is a sufficient condition for the strong g-leakage pre-order in probabilistic channels. The converse was proved by McIver et al. [7], [23], so establishing the complete coincidence between composition refinement and the strong g-leakage pre-order. Hence that ordering is a compelling generalization to probabilistic channels of the partition refinement on deterministic channels. Furthermore, [23] introduces abstract channels, which quotient over the abstract redundancies of the traditional channel representations. The authors show that composition refinement is a partial order on abstract channels, but not on channel matrices, and hence that the abstraction provides a more adequate representation for probabilistic systems in general.

Finally, Braun, Chatzikokolakis, and Palamidessi [29] show that multiplicative and additive min-entropy leakage orderings are equivalent when comparing two channels on a given prior. However, when channels are compared with respect to their capacity (over all priors), multiplicative and additive leakage can produce inconsistent results.

[16] In fact the capacity of cascades in the multiplicative case also follows directly from Theorem 6 of [38] and the Miracle Theorem.

XI. Conclusions and future work

We have studied robustness for leakage measurements by defining additive- and multiplicative capacities; and we have argued that both play an important role in assessing the severity of information flows. Our definitions are based on the relatively new approach of *gain functions*, which have been shown to capture a surprisingly wide range of attacker scenarios; and the corresponding capacity measurements thus apply automatically to all of those. Further, we have argued that robustness of leakage measurements is especially useful when our knowledge is limited, either about the prior or about the cost/benefit scenarios of an attacker, or both.

Surprisingly, even in the case of unknown and essentially unpredictable correlations with apparently *unrelated* data –the "Dalenius" scenario– we have been able to contribute novel limits on the information flow that can result.

Our current definitions of additive capacity use 1-spanning gain functions; this is primarily due to a 1-Lipschitz condition associated with the Kantorovich metric [35], [37] and is what enables our connection of the latter with the additive g-capacity. In future work we plan to investigate other metrics which would yield similarly efficient algorithms for computing capacity, but which could express more conveniently other notions of leakage (such as that based on Shannon). Finally we plan to investigate approximation algorithms for $\mathcal{L}^+_\forall(\forall, C)$.

Acknowledgments

Geoffrey Smith was partially supported by the National Science Foundation under grant CNS-1116318. McIver and Morgan were supported by the Australian Research Council under grant DP120101413. NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program. Also, the authors are grateful for support from Digiteo and the INRIA équipe associée Princess.

References

[1] D. Clark, S. Hunt, and P. Malacaria, "Quantitative information flow, relations and polymorphic types," *Journal of Logic and Computation*, vol. 18, no. 2, pp. 181–199, 2005.

[2] M. Clarkson, A. Myers, and F. Schneider, "Belief in information flow," in *Proc. 18th IEEE Computer Security Foundations Workshop (CSFW '05)*, 2005, pp. 31–45.

[3] B. Köpf and D. Basin, "An information-theoretic model for adaptive side-channel attacks," in *Proc. 14th ACM Conference on Computer and Communications Security (CCS '07)*, 2007, pp. 286–296.

[4] P. Malacaria, "Assessing security threats of looping constructs," in *Proc. 34th Symposium on Principles of Programming Languages (POPL '07)*, 2007, pp. 225–235.

[5] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, "On the Bayes risk in information-hiding protocols," *Journal of Computer Security*, vol. 16, no. 5, pp. 531–571, 2008.

[6] G. Smith, "On the foundations of quantitative information flow," in *Proc. 12th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS '09)*, ser. Lecture Notes in Computer Science, L. de Alfaro, Ed., vol. 5504, 2009, pp. 288–302.

[7] A. McIver, L. Meinicke, and C. Morgan, "Compositional closure for Bayes risk in probabilistic noninterference," in *Proc. ICALP'10*, 2010, pp. 223–235.

[8] M. R. Clarkson and F. B. Schneider, "Quantification of integrity," in *Proc. 23nd IEEE Computer Security Foundations Symposium (CSF '10)*, 2010, pp. 28–43.

[9] M. S. Alvim, M. Andrés, and C. Palamidessi, "Probabilistic information flow," in *Proc. 25th IEEE Symposium on Logic in Computer Science (LICS 2010)*, 2010, pp. 314–321.

[10] G. Barthe and B. Köpf, "Information-theoretic bounds for differentially private mechanisms," in *Proc. 24th IEEE Computer Security Foundations Symposium (CSF 2011)*, 2011, pp. 191–204.

[11] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *Proc. 25th IEEE Computer Security Foundations Symposium (CSF 2012)*, Jun. 2012, pp. 265–279.

[12] D. Clark, S. Hunt, and P. Malacaria, "A static analysis for quantifying information flow in a simple imperative language," *Journal of Computer Security*, vol. 15, pp. 321–371, 2007.

[13] M. Backes, B. Köpf, and A. Rybalchenko, "Automatic discovery and quantification of information leaks," in *Proc. 30th IEEE Symposium on Security and Privacy*, 2009, pp. 141–153.

[14] J. Newsome, S. McCamant, and D. Song, "Measuring channel capacity to distinguish undue influence," in *Proc. Fourth Workshop on Programming Languages and Analysis for Security (PLAS '09)*, 2009, pp. 73–85.

[15] M. Andrés, C. Palamidessi, P. van Rossum, and G. Smith, "Computing the leakage of information-hiding systems," in *Tools and Algorithms for the Construction and Analysis of Systems (TACAS '10)*, ser. Lecture Notes in Computer Science, J. Esparza and R. Majumdar, Eds., vol. 6015, 2010, pp. 373–389.

[16] K. Chatzikokolakis, T. Chothia, and A. Guha, "Statistical measurement of information leakage," in *Tools and Algorithms for the Construction and Analysis of Systems (TACAS '10)*, 2010, pp. 390–404.

[17] B. Köpf and A. Rybalchenko, "Approximation and randomization for quantitative information-flow analysis," in *Proc. 23nd IEEE Computer Security Foundations Symposium (CSF '10)*, 2010, pp. 3–14.

[18] H. Yasuoka and T. Terauchi, "Quantitative information flow — verification hardness and possibilities," in *Proc. 23nd IEEE Computer Security Foundations Symposium (CSF '10)*, 2010, pp. 15–27.

[19] Z. Meng and G. Smith, "Calculating bounds on information leakage using two-bit patterns," in *Proc. Sixth Workshop on Programming Languages and Analysis for Security (PLAS '11)*, 2011, pp. 1:1–1:12.

[20] B. Köpf and G. Smith, "Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks," in *Proc. 23nd IEEE Computer Security Foundations Symposium (CSF '10)*, 2010, pp. 44–56.

[21] J. Heusser and P. Malacaria, "Quantifying information leaks in software," in *Proc. ACSAC '10*, 2010, pp. 261–269.

[22] B. Köpf, L. Mauborgne, and M. Ochoa, "Automatic quantification of cache side-channels," in *Proc. 24th International Conference on Computer-Aided Verification (CAV '12)*, 2012, pp. 564–580.

[23] A. McIver, C. Morgan, G. Smith, B. Espinoza, and L. Meinicke, "Abstract channels and their robust information-leakage ordering," in *Proc. 3rd Conference on Principles of Security and Trust (POST 2014)*, 2014, pp. 83–102.

[24] T. Dalenius, "Towards a methodology for statistical disclosure control," *Statistik Tidskrift*, vol. 15, pp. 429–444, 1977.

[25] C. Dwork, "Differential privacy," in *Proc. 33rd International Colloquium on Automata, Languages, and Programming (ICALP 2006)*, 2006, pp. 1–12.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.

[27] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[28] J. L. Massey, "Guessing and entropy," in *Proc. 1994 IEEE International Symposium on Information Theory*, 1994, p. 204.

[29] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," in *Proc. 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009)*, ser. ENTCS, vol. 249, 2009, pp. 75–91.

[30] G. Smith, "Quantifying information flow using min-entropy," in *Proc. QEST 2011: 8th International Conference on Quantitative Evaluation of SysTems*, 2011, pp. 159–167.

[31] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[32] J. Bonneau, "The science of guessing: Analyzing an anonymized corpus of 70 million passwords," in *Security and Privacy (SP), 2012 IEEE Symposium on*, 2012, pp. 538–552.

[33] C. Dwork, "A firm foundation for private data analysis," *Communications of the ACM*, vol. 54, no. 1, 2011.

[34] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[35] Y. Deng and W. Du, "The Kantorovich Metric in computer science: A brief survey," *Electron. Notes Theor. Comput. Sci.*, vol. 253, no. 3, pp. 73–82, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.entcs.2009.10.006

[36] F. van Breugel, "The metric monad for probabilistic nondeterminism," 2005, draft available at *http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf*.

[37] A. McIver, L. Meinicke, and C. Morgan, "A Kantorovich-monadic powerdomain for information hiding, with probability and nondeterminism," in *Proc. 27th IEEE Symposium on Logic in Computer Science (LICS 2012)*, 2012, pp. 461–470.

[38] B. Espinoza and G. Smith, "Min-entropy as a resource," *Information and Computation (Special Issue on Information Security as a Resource)*, vol. 226, pp. 57–75, Apr. 2013.

[39] C. Shannon, "The lattice theory of information," *Information Theory, Transactions of the IRE Professional Group on*, vol. 1, no. 1, pp. 105 –107, February 1953.

[40] C. Cachin, "Entropy measures and unconditional security in cryptography," Ph.D. dissertation, Swiss Federal Institute of Technology, 1997.

[41] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden, "Anonymity protocols as noisy channels," *Information and Computation*, vol. 206, pp. 378–401, 2008.

[42] P. Malacaria and H. Chen, "Lagrange multipliers and maximum information leakage in different observational models," in *Proc. of the 2008 Workshop on Programming Languages and Analysis for Security (PLAS 2008)*. ACM, June 2008, pp. 135–146.

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

[44] I. S. Moskowitz, R. E. Newman, D. P. Crepeau, and A. R. Miller, "Covert channels and anonymizing networks," in *Proceedings of the 2003 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '03. New York, NY, USA: ACM, 2003, pp. 79–88. [Online]. Available: http://doi.acm.org/10.1145/1005140.1005153

[45] M. S. Alvim, M. E. Andrés, and C. Palamidessi, "Quantitative information flow in interactive systems," *Journal of Computer Security*, vol. 20, no. 1, pp. 3–50, 2012.

[46] P. Malacaria, "Algebraic foundations for information theoretical, probabilistic and guessability measures of information flow," *CoRR*, vol. abs/1101.3453, 2011.

[47] J. Landauer and T. Redmond, "A lattice of information," in *Proc. 6th IEEE Computer Security Foundations Workshop (CSFW'93)*, Jun. 1993, pp. 65–70.

Appendix

A. Technical lemmas on gain functions (from §IV and §VII)

Lem. 24 shows how to express the vulnerability as a standard expected value of a function over a hyper, and is a general result we use repeatedly.

Lemma 24: Let $C: \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ be a channel, and g a gain function and π a prior. Then we have $V_g[\pi, C] = \mathcal{E}_{[\pi,C]}V_g$.

Proof: The left-hand side just above is interpreted at (4) as an expected value on the distribution p_Y of a random variable $F: \mathcal{Y}\to\mathbb{R}$ defined $F(y) = V_g(p_{X|y})$, where the joint distribution Π wrt. which p_Y, $p_{X|y}$ are defined is the usual one induced by channel C's action on the prior π.

We want to establish that the construction can equally well be interpreted as acting on a hyper directly: the $[\pi, C]$ on the

right-hand side just above is the hyper defined by Π, and one way of formulating that hyper is as the *push forward* of a function $G(y) = p_{X|y}$ acting on p_Y, typically written $G_*(p_Y)$. And then the entire right-hand side is the expected value $\mathcal{E}_{G_*(p_Y)}V_g$.

Once we observe that F is the functional composition $V_g \circ G$, the result follows from a general equality relating expected value and push-forward, that $\mathcal{E}_\delta(A \circ B) = \mathcal{E}_{B_*(\delta)}A$.

Further details about hypers and their relation to joint distributions can be found at [7], [23], [37]. ■

Now we can use standard properties of expected values to verify our gain function algebra.

Lemma 25: Gain function algebra (§IV) Recall the definitions from (5) and (6) in §IV:

$$\begin{array}{lrcl} & g_{\times k}(w,x) & = & g(w,x) \times k \\ & g_{+k}(w,x) & = & g(w,x) + k \\ & g_{+k@x'}(w,x) & = & g(w,x) + (k \text{ if } x{=}x' \text{ else } 0) \\ x\text{-Shift} & V_{g_{+k@x'}}[\pi,C] & = & V_g[\pi,C] + k\pi_{x'} \\ \text{Global Shift} & V_{g_{+k}}[\pi,C] & = & V_g[\pi,C] + k \\ \text{Scale} & V_{g_{\times k}}[\pi,C] & = & k \times V_g[\pi,C] \text{ , for } k{\geq}0 \end{array}$$

The general proof strategy for these properties is to use distribution of addition and multiplication through the expectation operator. Note however that this expectation operator acts at the level of hypers and therefore most of the proofs are concerned with translating $V_g[\pi,C]$ into that form.

We use $\underline{0}$ in gain function expressions to represent the gain function that returns the value 0 for all values of w, x. This allows us to write $V_{g_{+k@x'}}$ as $V_g + V_{\underline{0}_{+k@x'}}$, because the modification of g to $g_{+k@x'}$ does not affect the (maximising) choice of w. Our three proofs are now as follows:

x-Shift:

$$\begin{array}{lll} & V_{g_{+k@x'}}[\pi,C] & \\ = & \mathcal{E}_{[\pi,C]}V_{g_{+k@x'}} & \text{“Lem. 24”} \\ = & \mathcal{E}_{[\pi,C]}(V_g + V_{\underline{0}_{+k@x'}}) & \text{“remark above”} \\ = & \mathcal{E}_{[\pi,C]}V_g + \mathcal{E}_{[\pi,C]}V_{\underline{0}_{+k@x'}} & \text{“averaging distributes addition”} \\ = & \mathcal{E}_{[\pi,C]}V_g + k\pi_{x'} & \text{“}\mathcal{E}_{[\pi,C]}V_{0_{+k@x'}} = \textstyle\sum_y k\pi_{x'}C_{x',y} = k\pi_{x'}\sum_y C_{x',y} = k\pi_{x'}\text{”} \\ = & V_g[\pi,C] + k\pi_{x'} \text{ .} & \text{“Lem. 24”} \end{array}$$

Global Shift: This follows from the proof of x-Shift since a global shift is simply an x-shift for each value of x in $\mathcal{X}$.

Scaling: This follows from observing that $V_{g_{\times k}} = kV_g$ when $k{>}0$, and then using the fact that multiplication distributes through the expectation operator.

The next result Lem. 26 is used for Lem. 14 from §VII.

Lemma 26: 1-spanning implies 1-Lipschitz
If g is 1-spanning then V_g is 1-Lipschitz (wrt $\mathbb{K}_{\mathbb{D}\mathcal{X}}$), considered as a function $\mathbb{D}\mathcal{X}{\to}\mathbb{R}$.

Proof: (Note that the proof does not assume a finite $\mathcal{W}$.)

Fix $\pi, \pi' : \mathbb{D}\mathcal{X}$; define $\mathcal{X}^+ = \{x : \mathcal{X} \mid \pi_x \geq \pi'_x\}$ and $\mathcal{X}^- = \mathcal{X} \setminus \mathcal{X}^+$. From $\sum\pi = \sum\pi' = 1$ we have

$$\textstyle\sum_{x:\mathcal{X}^+}(\pi_x - \pi'_x) = \sum_{x:\mathcal{X}^-}(\pi'_x - \pi_x) = {}^1\!/_2 \sum_x |\pi_x - \pi'_x| = \mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi,\pi'). \quad (10)$$

Given a guess $w : \mathcal{W}$, we define

$$g(w,\top) \;=\; \max_x g(w,x) \quad \text{and} \quad g(w,\bot) \;=\; \min_x g(w,x) \text{ ,}$$

and then define the function $f_w : \mathbb{D}\mathcal{X}{\to}\mathbb{R}$ as

$$f_w(\pi) \quad = \quad \mathcal{E}_\pi g(w,\cdot) \quad = \quad \textstyle\sum_x \pi_x g(w,x) \text{ .}$$

We first show that f_w is 1-Lipschitz, as follows:

$$\begin{array}{lll} & |f_w(\pi) - f_w(\pi')| & \\ = & |\sum_x(\pi_x - \pi'_x)g(w,x)| & \text{“definition of } f_w\text{”} \\ = & & \text{“rewrite sum”} \\ & |\sum_{x:\mathcal{X}^+}(\pi_x - \pi'_x)g(w,x) \;-\; \sum_{x:\mathcal{X}^-}(\pi'_x - \pi_x)g(w,x)| & \\ \leq & & \text{“definition } g(w,\top), g(w,\bot) \text{ above”} \\ & |\sum_{x:\mathcal{X}^+}(\pi_x - \pi'_x)g(w,\top) \;-\; \sum_{x:\mathcal{X}^-}(\pi'_x - \pi_x)g(w,\bot)| & \\ = & \mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi,\pi')\,|g(w,\top) \;-\; g(w,\bot)| & \text{“ (10) ”} \\ \leq & \mathbb{K}_{\mathbb{D}\mathcal{X}}(\pi,\pi') \text{ .} & \text{“}g \text{ is 1-spanning Def. 3”} \end{array}$$

Now we recall that $V_g[\pi] = \sup_w f_w(\pi)$, and so it remains only to show that the supremum preserves the 1-Lipschitz property.

So let F be a set of functions $\mathcal{A}{\to}\mathbb{R}$ that are 1-Lipschitz wrt some metric d on $\mathcal{A}$. We show that $F(a) = \sup_{f:F} f(a)$ is also 1-Lipschitz. Fixing $a, a' : \mathcal{A}$, assume wlog that $F(a){\geq}F(a')$; we have

$$\begin{array}{lll} & |F(a) - F(a')| & \\ = & \sup_f f(a) - \sup_f f(a') & \text{“definition } F\text{; } F(a) \geq F(a')\text{”} \\ \leq & \sup_f (f(a) - f(a')) & \text{“}-\sup_{f'} f'(a') \leq -f(a')\text{”} \\ \leq & \sup_f |f(a) - f(a')| & \\ \leq & \sup_f d(a,a') & \text{“}f \text{ is 1-Lipschitz wrt } d\text{”} \\ = & d(a,a') \text{ .} & \end{array}$$

■

B. Calculation of Kantorovich distance from $[\pi]$ to $[\pi, C]$ (from §VII-C)

The *Manhattan distance* between two vectors of equal dimension (e.g. distributions in $\mathbb{D}\mathcal{X}$) is the sum of the non-negative differences between them in each dimension, by analogy with the distance needed to travel between two intersections in a city of regular blocks (like Manhattan). Thus we have

$$\begin{array}{ll} & \mathbb{K}_{\mathbb{D}^2\mathcal{X}}([\pi],[\pi,C]) \\ = & 1/2 \sum_y p(y)(\text{Manhattan distance from } p_X \text{ to } p_{X|y}) \\ = & 1/2 \sum_y p(y) \sum_x |p_X(x) - p_{X|y}(x)| \\ = & 1/2 \sum_y p(y) \sum_x |\pi_x - p(x,y)/p(y)| \\ = & 1/2 \sum_{x,y} |\pi_x p(y) - p(x,y)| \\ = & 1/2 \sum_{x,y} \pi_x |p(y) - C_{x,y}| \\ = & 1/2 \sum_{x,y} \pi_x |C_{x,y} - \sum_{x'} \pi_{x'} C_{x',y}| \text{ .} \end{array}$$

C. Existence of g-function for Thm. 17 (from §VII-B)

Here we show that the g-function quoted in the proof of Thm. 17 has the property claimed:

$$\begin{array}{lll} & \sum_x \tau_x g(w,x) & \\ = & \sum_x \tau_x (1 + (w(x) - \sum_{x'} \pi_{x'} w(x'))/2) & \text{“definition } g\text{”} \\ = & (\sum_x \tau_x w(x) - \sum_x \tau_x \sum_{x'} \pi_{x'} w(x'))/2 + 1 & \text{“}\textstyle\sum_x \tau_x = 1\text{”} \\ = & (\sum_x \tau_x w(x) - \sum_{x'} \pi_{x'} w(x'))/2 + 1 & \text{“}\textstyle\sum_x \tau_x = 1\text{”} \\ = & \sum_x (\tau_x - \pi_x) w(x)/2 + 1 \text{ .} & \text{“rename } x' \text{ to } x\text{”} \end{array}$$