2016 IEEE 29th Computer Security Foundations Symposium

# Correlated Secrets in Quantitative Information Flow

Nicolás E. Bordenabe
Department of Computing
Macquarie University, Sydney
Email: nicolas.bordenabe@mq.edu.au

Geoffrey Smith
School of Computing and Information Sciences
Florida International University, Miami
Email: smithg@cis.fiu.edu

***Abstract*—A fundamental challenge in controlling the leakage of sensitive information by computer systems is the possibility of correlations between different secrets, with the result that leaking information about one secret may also leak information about a different secret. We explore such leakage, here called Dalenius leakage, within the context of the $g$-leakage family of leakage measures. We prove a fundamental equivalence between Dalenius min-entropy leakage under arbitrary correlations and $g$-leakage under arbitrary gain functions, and show how this equivalence increases the significance of the composition refinement relation. We also consider Dalenius leakage in the case when the marginal distributions induced by the correlation are known, giving techniques to compute stronger upper bounds in this case.**

***Index Terms*—Information leakage, Dalenius's Desideratum, channel capacity.**

## I. Introduction

In the context of *differential privacy* [1], [2], an important motivation is the possibility that the adversary knows interesting *correlations* among secrets. Recall Dwork's discussion of what she calls *Dalenius's Desideratum*:

> In 1977 the statistician Tore Dalenius articulated an "*ad omnia*" (as opposed to *ad hoc*) privacy goal for statistical databases: Anything that can be learned about a respondent from the statistical database should be learnable without access to the database. [2, p. 90]

Dwork argues for the impossibility of this goal by imagining the auxiliary information "Turing is two inches taller than the average Lithuanian woman." With such information, revealing statistical information about Lithuanian women reveals information about Alan Turing.[1]

Similar concerns arise in the context of *quantitative information flow* [4], [5], [6], [7], [8], [9]. Here we are concerned with a secret $Y$, about which an adversary knows only a *prior probability distribution* $\psi$. A system $C$ (modeled as an information-theoretic *channel*) takes $Y$ as input and produces observable output $Z$, and we wish to quantify how much information about $Y$ is leaked by $C$ through $Z$. But, as in the scenario imagined by Dwork, there could be a *different* secret $X$, apparently having nothing to do with channel $C$, which is *correlated* with $Y$; in this case, we can see channel $C$ as leaking information about $X$.

[1]We remark here that the name "Dalenius's Desideratum" actually appears to us to be a misnomer that is quite unfair to Dalenius [3]; we elaborate on this point in Section VII.

To illustrate, let us consider an example. Suppose that a number of *beneficiaries* (of some sort) reside in a region comprising three counties (A, B, and C), and have a variety of ages. If a beneficiary is selected at random, then we might regard the beneficiary's *county* and *age* as secrets $X$ and $Y$, respectively. Now suppose that a census has been taken and the following table of "macrostatistics" (from Dalenius [3]) has been published:

*Number of beneficiaries by county and age*

| County | Age class | | | |
|---|---|---|---|---|
| | Under 65 | 65–69 | 70–74 | 75 & over |
| A | 3 | 15 | 11 | 8 |
| B | 7 | 60 | 34 | 20 |
| C | 0 | 4 | 0 | 0 |

We can normalize this table by dividing each entry by 162 (since there are a total of 162 beneficiaries), giving a *joint distribution* $J$ between $X$ and $Y$:

| $J$ | Under 65 | 65–69 | 70–74 | 75 & over |
|---|---|---|---|---|
| A | $3/162$ | $15/162$ | $11/162$ | $8/162$ |
| B | $7/162$ | $60/162$ | $34/162$ | $20/162$ |
| C | 0 | $4/162$ | 0 | 0 |

Moreover by summing the rows and columns of $J$ we can now compute marginal distributions $\pi$ and $\psi$ for $X$ and $Y$, respectively:

$$\pi = (37/162, 121/162, 4/162)$$

and

$$\psi = (10/162, 79/162, 45/162, 28/162).$$

These marginals are of course useful to an adversary interested in discovering $X$ or $Y$. More interestingly, the correlation $J$ implies that a channel $C$ that leaks information about $Y$ will also leak information about $X$. For instance, if an output of $C$ reveals that $Y$ is not in the range 65–69, then the adversary can deduce that $X$ is not C. We call such leakage the *Dalenius leakage* of $X$ caused by channel $C$ under correlation $J$.

The main goal of this paper is to develop theory allowing us to *quantify* Dalenius leakage within the *g-leakage* family of leakage measures [10], [11], [12].[2] While some study of

[2]We review $g$-leakage, and its use of *gain functions* $g$ to model the operational scenario, in Section II below.


DOI 10.1109/CSF.2016.14

Dalenius leakage was done already in [12], this paper makes several significant and novel contributions:

- Given a correlation $J$ between $X$ and $Y$ giving marginal distribution $\pi$ on $X$ and a channel $C$ from $Y$ to $Z$, we show that the Dalenius $g$-leakage of $X$ caused by $C$ under $J$ is equal to the ordinary $g$-leakage of $X$ under $\pi$ and the cascade $BC$, where $B$ is a channel matrix formed by normalizing the rows of $J$. (Section III).
- We establish a fundamental equivalence between *Dalenius min-entropy leakage under arbitrary correlations* and *g-leakage under arbitrary gain functions*, and we show how this equivalence offers additional justification for the significance of the *composition refinement* relation studied in [10], [11] (Section IV).
- We extend the *capacity* bounds on Dalenius leakage shown in [12] by considering situations where something is known about the correlation $J$, giving for instance techniques to compute stronger capacity bounds in the case when the marginal distributions induced by $J$ are known (Section V).

The rest of the paper is structured as follows. Section II gives a brief summary of important concepts and results in the theory of quantitative information flow and $g$-leakage. Section III formally introduces Dalenius leakage, and shows how it can be expressed in terms of $g$-leakage. Section IV presents a striking equivalence between Dalenius min-entropy leakage and $g$-leakage, and uses it to further justify the composition refinement relation. Section V presents a series of results and techniques to obtain improved bounds on Dalenius leakage when something is known about the correlation between secrets. Section VI briefly explores the applicability of our results to additive leakage. Finally, Section VII discusses related work and Section VIII concludes.

## II. Technical Background

In this section, we briefly review the key concepts of quantitative information flow and $g$-leakage, as developed in [10], [11], [12].

### A. Secrets and their vulnerability

A *secret* $Y$ is something about which an adversary knows only a probability distribution $\psi \in \mathbb{D}\mathcal{Y}$, where $\mathcal{Y}$ is the set (assumed to be finite) of possible values of $Y$. For example, if $Y$ is a *randomly-generated string of bits*, then $\psi$ is determined by the random process used to generate $Y$. In contrast, if $Y$ is a secret like *the beneficiary's age*, then $\psi$ is determined by the adversary's knowledge of the population that the beneficiary comes from.

To quantify the threat to a secret with distribution $\psi$, it is natural to consider the *Bayes vulnerability*[3] of $\psi$, denoted $V[\psi]$, and defined as the maximum probability within $\psi$:

$$V[\psi] = \max_{y \in \mathcal{Y}} \psi_y.$$

[3]In [8], it is called simply "vulnerability".

Note that $V[\psi]$ is an optimal adversary's probability of guessing $Y$ correctly in one try.

While Bayes vulnerability is clearly a basic security concern, we can easily imagine operational scenarios where it is not appropriate. For instance, an adversary might benefit by guessing a value *close* to $Y$ or a *property* of $Y$, or guessing $Y$ within *three tries*; or the adversary might be *penalized* for making an incorrect guess. For this reason, [10] generalizes Bayes vulnerability to *g-vulnerability* $V_g$, parameterizing $V$ with a *gain function* $g$ that models the operational scenario.

For each scenario, there is a finite set $\mathcal{W}$ of "guesses" (or "actions") that the adversary could make about the secret, and for any guess $w$ and secret value $y$, there will be some *gain* $g(w, y)$ that the adversary gets by having chosen $w$ when the secret's actual value was $y$. Formally, $g : \mathcal{W} \times \mathcal{Y} \to [0, 1]$, where $\mathcal{W}$ is a finite, non-empty set.[4] All the scenarios listed above, and many others, can be realized using gain functions.

Given a gain function $g$, the prior *g-vulnerability* is defined as the maximum expected gain over all possible guesses:

$$V_g[\psi] = \max_w \sum_y \psi_y g(w, y)\,.$$

Note that $g$-vulnerability generalizes Bayes vulnerability since Bayes vulnerability is the same as $g_{id}$-vulnerability, where $g_{id}$ is the *identity gain function*, defined by

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

Note that $g_{id}$ has $\mathcal{W} = \mathcal{X}$ and, when written as a matrix, it is the *identity matrix* $I$.

### B. Information-theoretic channels

A channel $C$ that (perhaps probabilistically) maps secret input $Y$ to some observable output $Z$ can be modeled as an information-theoretic *channel matrix*, whose rows show the distribution of outputs corresponding to each possible input. For example,

| $C$ | $z_1$ | $z_2$ | $z_3$ | $z_4$ |
|---|---|---|---|---|
| $y_1$ | 1 | 0 | 0 | 0 |
| $y_2$ | 0 | $1/4$ | $1/2$ | $1/4$ |
| $y_3$ | $1/2$ | $1/3$ | $1/6$ | 0 |

$C$ is *deterministic* if each row contains exactly one 1, with all other entries 0.

Now we consider the question of what the output of $C$ reveals about $Y$ to an adversary who knows channel $C$ and prior $\psi = (1/4, 1/2, 1/4)$. First we form the *joint matrix* by multiplying each row of $C$ by the prior probability. Alternatively, if we let $\Psi$ (indexed by $\mathcal{Y} \times \mathcal{Y}$) be a diagonal matrix with the prior $\psi$ on its diagonal, then the joint matrix is the matrix product $\Psi C$:

| $\Psi C$ | $z_1$ | $z_2$ | $z_3$ | $z_4$ |
|---|---|---|---|---|
| $y_1$ | $1/4$ | 0 | 0 | 0 |
| $y_2$ | 0 | $1/8$ | $1/4$ | $1/8$ |
| $y_3$ | $1/8$ | $1/12$ | $1/24$ | 0 |

[4]Sometimes gain functions are generalized by allowing them to return arbitrary real numbers and allowing $\mathcal{W}$ to be infinite.

Summing the columns of $\Psi C$ gives the marginal distribution $p_Z = (3/8, 5/24, 7/24, 1/8)$. Each possible value of $Z$ gives rise to a *posterior distribution* on $Y$, by Bayesian updating; these can be calculated by normalizing the columns of $\Psi C$:

| | $p_{Y\|z_1}$ | $p_{Y\|z_2}$ | $p_{Y\|z_3}$ | $p_{Y\|z_4}$ |
|---|---|---|---|---|
| $y_1$ | $2/3$ | $0$ | $0$ | $0$ |
| $y_2$ | $0$ | $3/5$ | $6/7$ | $1$ |
| $y_3$ | $1/3$ | $2/5$ | $1/7$ | $0$ |

Notice that the particular output *labels* ($z_1$, $z_2$, ...) do not matter to the adversary. All that matters are the *posterior distributions*, which can be seen as the possible "worlds" that the adversary will be in, along with their *probabilities*, given by the marginal distribution $p_Z$. The adversary thus obtains a *distribution on posterior distributions*, called a *hyper-distribution* [9] and denoted by $[\psi, C]$:

| $[\psi, C]$ | $3/8$ | $5/24$ | $7/24$ | $1/8$ |
|---|---|---|---|---|
| $y_1$ | $2/3$ | $0$ | $0$ | $0$ |
| $y_2$ | $0$ | $3/5$ | $6/7$ | $1$ |
| $y_3$ | $1/3$ | $2/5$ | $1/7$ | $0$ |

In conclusion, the information-theoretic essence of $C$ is a mapping from priors $\psi$ to hyper-distributions $[\psi, C]$.[5]

### C. Posterior vulnerability and leakage

Now we consider the vulnerability of $Y$ after the adversary sees the output of $C$. The *posterior vulnerability* is naturally defined as the weighted average of the vulnerabilities of each of the posterior distributions:

$$V_g[\psi, C] = \sum_z p_Z(z) V_g[p_{Y|z}] \,.$$

(Notice that this is simply the *expectation* of $V_g$ with respect to the hyper-distribution $[\psi, C]$.) An equivalent formulation, which we will use subsequently, is

$$V_g[\psi, C] = \sum_z \max_w \sum_y \psi_y C_{yz} g(w, y) \,.$$

Now *g-leakage* is naturally defined in terms of the *ratio* between the posterior- and prior vulnerabilities:

$$\mathcal{L}_g(\psi, C) = \log \frac{V_g[\psi, C]}{V_g[\psi]} \,.$$

In the special case when the gain function is $g_{id}$ (giving Bayes vulnerability $V$), we get what is called *min-entropy leakage*.

The leakages that we have just defined are *multiplicative*, and they are our main focus in this paper. However, in Section VI we briefly consider *additive g-leakage*, defined by

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

To achieve more *robust* leakage analysis of a channel $C$, we can abstract away from a particular prior $\psi$ and consider instead the maximum leakage over all possible priors; this is called the *capacity*. Particularly useful is *min-capacity* $\mathcal{ML}(C)$, the maximum min-entropy leakage over all priors:

$$\mathcal{ML}(C) = \sup_\psi \log \frac{V[\psi, C]}{V[\psi]} \,.$$

The following theorems (from [13], [10], [14]) give three important properties of min-capacity.

First, min-capacity is easy to compute:

*Theorem 2.1:* $\mathcal{ML}(C)$ is the logarithm of the sum of the column maximums of $C$, and is always realized on a uniform prior.

Second, min-capacity is an upper bound on $g$-leakage, for any prior and any gain function:

*Theorem 2.2 (Miracle):* For all priors $\psi$ and gain functions $g$, $\mathcal{ML}(C) \geq \mathcal{L}_g(\psi, C)$.

Third, the min-capacity of a *cascade* $CD$ (formed by multiplying channel matrices $C$ and $D$) is upper bounded by the min-capacities of $C$ and $D$:

*Theorem 2.3:* $\mathcal{ML}(CD) \leq \min\{\mathcal{ML}(C), \mathcal{ML}(D)\}$.

### D. Composition Refinement

Given channels $C$ and $D$, both taking input $Y$, the question of *which leaks more* will ordinarily depend on the prior and gain function used. However, a robust channel ordering is given by the *composition refinement* relation.

*Definition 2.1:* $C$ is *composition refined* by $D$, denoted $C \sqsubseteq_\circ D$, if there exists a channel $R$ such that $D = CR$.

This definition requires that $D$ can be expressed as the cascade of $C$ and $R$ for some channel $R$, which intuitively means that $D$ can be understood as $C$ followed by some "post-processing" $R$.[6]

While composition refinement is only a pre-order on channel matrices, in [11] it is shown that on *abstract channels*, composition refinement ($\sqsubseteq_\circ$) is a *partial order*.

But the main interest in composition refinement is its relation to $g$-leakage. First, composition refinement implies a strong $g$-leakage ordering; this can be seen as an analogue of the classic *data-processing inequality*.

*Theorem 2.4 (Data-processing inequality):* If $C \sqsubseteq_\circ D$ then the $g$-leakage of $D$ never exceeds that of $C$, for any prior $\psi$ and any gain function $g$. (We denote this by $C \geq_\mathcal{G} D$.)

More interestingly, the converse implication holds as well.

*Theorem 2.5 (Coriaceous):* If $C \geq_\mathcal{G} D$ then $C \sqsubseteq_\circ D$.

The Coriaceous Theorem was proved in [11], but it was subsequently learned that this theorem was in fact discovered already in the 1950s by statistician David Blackwell [15], [16].

Hence composition refinement ($\sqsubseteq_\circ$) is a partial order on abstract channels with both structural and leakage-testing significance.

[5] Note that this implies that classic channel matrices have structural redundancies: *labels of outputs*, *columns that are multiples of one another*, and *zero columns*, which are irrelevant with respect to (information-theoretic) leakage. Eliminating these redundancies gives *abstract channels*, as studied in [11].

[6] Note that in [10] the order of the channels was *reversed*.

## III. Dalenius Scenarios

In this section, we show how $g$-leakage can quantify the "surprising" information leakage that can result from correlations between secrets; we call this *Dalenius leakage*.

Consider an adversary interested in learning a secret $X$, assumed to have prior distribution $\pi \in \mathbb{D}\mathcal{X}$. Assume that the adversary's gain function is $g : \mathcal{W} \times \mathcal{X} \to [0,1]$. Suppose further that there is a channel $C$ from $\mathcal{Y}$ to $\mathcal{Z}$, apparently having nothing to do with $X$.

But suppose that there is a joint distribution $J \in \mathbb{D}(\mathcal{X} \times \mathcal{Y})$ expressing a *correlation* between $X$ and $Y$. Note that $J$ must give marginal distribution $\pi$ to $X$: for every $x$, $\sum_y J_{xy} = \pi_x$.

Then we can see $C$ as leaking information about $X$. Following [12], we can extend $C$ to a channel $C^*$ from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{Z}$ by defining

$$C^*_{(x,y),z} = C_{y,z}$$

(this means that $C^*$ ignores $X$). And we can extend $g$ to a gain function $g^* : \mathcal{W} \times (\mathcal{X} \times \mathcal{Y}) \to [0,1]$ by defining

$$g^*(w,(x,y)) = g(w,x)$$

(this means that $g^*$ ignores $Y$). Now we can define Dalenius vulnerability and leakage.

*Definition 3.1:*

- $DV_g[J]$, the *prior Dalenius $g$-vulnerability* of $X$ under correlation $J$, is $V_{g^*}[J]$.
- $DV_g[J,C]$, the *posterior Dalenius $g$-vulnerability* of $X$ under correlation $J$ and channel $C$, is $V_{g^*}[J,C^*]$.
- $\mathcal{DL}_g(J,C)$, the *Dalenius $g$-leakage* of $X$ under correlation $J$ and channel $C$, is $\log(DV_g[J,C]/DV_g[J])$.

When referring to *Dalenius Bayes vulnerability* or *Dalenius min-entropy leakage* (i.e. with respect to $g_{id}$), we will omit the $g$, so that $DV[J] = DV_{g_{id}}[J]$, $DV[J,C] = DV_{g_{id}}[J,C]$, and $\mathcal{DL}(J,C) = \mathcal{DL}_{g_{id}}(J,C)$.

Also, in our studies of capacity in Section V we will write $\mathcal{DL}_\forall(J,C)$ to denote quantification over *all* gain functions:

$$\mathcal{DL}_\forall(J,C) = \sup_g \mathcal{DL}_g(J,C).$$

We now develop some important properties of Dalenius vulnerability and leakage. First, we show that the correlation $J$ does not, by itself, have any effect on the prior vulnerability of $X$.

*Theorem 3.1:* $DV_g[J] = V_g[\pi]$.

*Proof:* We have

$$\begin{array}{lll} & DV_g[J] & \\ = & V_{g^*}[J] & \text{(definition of } DV_g\text{)} \\ = & \max_w \sum_{x,y} J_{xy} g^*(w,(x,y)) & \text{(definition of } V_{g^*}\text{)} \\ = & \max_w \sum_{x,y} J_{xy} g(w,x) & \text{(definition of } g^*\text{)} \\ = & \max_w \sum_x (\sum_y J_{xy}) g(w,x) & \text{(reorganizing sum)} \\ = & \max_w \sum_x \pi_x g(w,x) & (\pi \text{ is marginal of } J) \\ = & V_g[\pi] & \text{(definition of } V_g\text{)} \end{array}$$

■

More interestingly, we can achieve a neater formulation for the posterior Dalenius $g$-vulnerability. First, observe that we can view $J$ as the joint matrix resulting from prior $\pi$ and a channel matrix $B$ from $\mathcal{X}$ to $\mathcal{Y}$; that is, we can define a channel matrix $B$ such that for all $x$ and $y$, $J_{xy} = \pi_x B_{xy}$. When $\pi$ is full support, channel $B$ is uniquely obtained by normalizing the rows of $J$, giving $B_{xy} = J_{xy}/\pi_x$. If however we have $\pi_x = 0$, then row $x$ of $J$ is all zero and therefore cannot be normalized; but in this case row $x$ of $B$ can be chosen arbitrarily.

Now we show that the posterior Dalenius $g$-vulnerability of $X$ under $J$ and $C$ is simply the posterior $g$-vulnerability of $X$ under $\pi$ and the *cascade* $BC$.

*Theorem 3.2:* If $B$ is any channel from $\mathcal{X}$ to $\mathcal{Y}$ satisfying $J_{xy} = \pi_x B_{xy}$ for all $x$ and $y$, then $DV_g[J,C] = V_g[\pi, BC]$.

*Proof:* We have

$$\begin{array}{lll} & DV_g[J,C] & \\ = & V_{g^*}[J,C^*] & \text{(def. } DV_g\text{)} \\ = & \sum_z \max_w \sum_{x,y} J_{xy} C^*_{(x,y)z} g^*(w,(x,y)) & \text{(def. } V_{g^*}\text{)} \\ = & \sum_z \max_w \sum_{x,y} \pi_x B_{xy} C_{yz} g(w,x) & \text{(def. } B, C^*, g^*\text{)} \\ = & \sum_z \max_w \sum_x \pi_x (\sum_y B_{xy} C_{yz}) g(w,x) & \text{(reorganize sum)} \\ = & \sum_z \max_w \sum_x \pi_x (BC)_{xz} g(w,x) & \text{(def. cascade)} \\ = & V_g[\pi, BC] & \text{(def. } V_g\text{)} \end{array}$$

■

Finally, we get a neater formulation of Dalenius leakage as an immediate corollary.

*Corollary 3.3:* If $B$ is any channel from $\mathcal{X}$ to $\mathcal{Y}$ satisfying $J_{xy} = \pi_x B_{xy}$ for all $x$ and $y$, then $\mathcal{DL}_g(J,C) = \mathcal{L}_g(\pi, BC)$.

*Proof:* We have

$$\begin{array}{ll} & \mathcal{DL}_g(J,C) \\ = & \log(DV_g[J,C]/DV_g[J]) \\ = & \log(V_g[\pi,BC]/V_g[\pi]) \\ = & \mathcal{L}_g(\pi, BC) \end{array}$$

■

The fact that Dalenius leakage under $C$ is equal to $g$-leakage under a cascade $BC$ is useful for obtaining bounds on Dalenius leakage, because it enables us to use existing bounds on the leakage of cascades, such as Theorem 2.3. We will explore such bounds in Section V.

Corollary 3.3 also directly implies the following important result, which shows that composition refinement ($\sqsubseteq_\circ$) also implies a *strong Dalenius leakage ordering*:

*Corollary 3.4:* If $C \sqsubseteq_\circ D$, then the Dalenius $g$-leakage of $D$ never exceeds that of $C$, for any correlation $J$ and any gain function $g$.

*Proof:* If $C \sqsubseteq_\circ D$, then $D = CR$, for some $R$. Hence, for any $B$ (with appropriate dimensions) we have $BD = BCR$, whence $BC \sqsubseteq_\circ BD$. Hence, by Theorem 2.4, the $g$-leakage of $BD$ never exceeds that of $BC$. And so, by Corollary 3.3, the Dalenius $g$-leakage of $D$ never exceeds than of $C$. ■

We conclude this section with an example calculation of Dalenius leakage.

*Example 3.1:* Suppose that we have a 2-bit secret $Y$ and a channel $C$ that leaks the least significant bit of $Y$:

| $C$ | 0 | 1 |
|---|---|---|
| 00 | 1 | 0 |
| 01 | 0 | 1 |
| 10 | 1 | 0 |
| 11 | 0 | 1 |

Now suppose that there is a 1-bit secret $X$ that is correlated to $Y$ according to the following joint distribution $J$:

| $J$ | 00 | 01 | 10 | 11 |
|---|---|---|---|---|
| 0 | $1/8$ | $1/16$ | $1/4$ | $1/16$ |
| 1 | $1/16$ | $1/4$ | $1/16$ | $1/8$ |

According to $J$, the least significant bit of $Y$ is uniformly distributed, that is, the probability of $Y$ being either 00 or 10 is the same as the probability of its being either 01 or 11. Furthermore, $X$ is equal to the least significant bit of $Y$ with probability $3/4$.

We obtain the marginal distributions $\pi$ and $\psi$ on $\mathcal{X}$ and $\mathcal{Y}$ respectively:

| | 0 | 1 |
|---|---|---|
| $\pi$ | $1/2$ | $1/2$ |

| | 00 | 01 | 10 | 11 |
|---|---|---|---|---|
| $\psi$ | $3/16$ | $5/16$ | $5/16$ | $3/16$ |

Now, to calculate the min-entropy leakage of $C$ (about the value of $Y$), we first calculate the hyper-distribution $[\psi, C]$:

| $[\psi, C]$ | $1/2$ | $1/2$ |
|---|---|---|
| 00 | $3/8$ | 0 |
| 01 | 0 | $5/8$ |
| 10 | $5/8$ | 0 |
| 11 | 0 | $3/8$ |

Hence we have $V[\psi] = 5/16$ and

$$V[\psi, C] = 1/2 \cdot 5/8 + 1/2 \cdot 5/8 = 5/8,$$

which implies that

$$\mathcal{L}(\psi, C) = \log\left(\frac{5/8}{5/16}\right) = \log 2 = 1.$$

A min-entropy leakage of 1 bit is intuitively correct here, since $C$ reveals the last bit of $Y$, which was uniformly distributed under $\psi$.

Turning our attention now to leakage about $X$, we now calculate the Dalenius min-entropy leakage of $C$ under $J$. For that, we first get $B$ by normalizing the rows of $J$, and then we calculate the cascade $BC$:

| $B$ | 00 | 01 | 10 | 11 |
|---|---|---|---|---|
| 0 | $1/4$ | $1/8$ | $1/2$ | $1/8$ |
| 1 | $1/8$ | $1/2$ | $1/8$ | $1/4$ |

$\cdot$

| $C$ | 0 | 1 |
|---|---|---|
| 00 | 1 | 0 |
| 01 | 0 | 1 |
| 10 | 1 | 0 |
| 11 | 0 | 1 |

$=$

| $BC$ | 0 | 1 |
|---|---|---|
| 0 | $3/4$ | $1/4$ |
| 1 | $1/4$ | $3/4$ |

The Dalenius min-entropy leakage of $C$ under $J$ can now be calculated as follows:

$$\mathcal{DL}(J, C) = \mathcal{L}(\pi, BC) = \log\left(\frac{3/4}{1/2}\right) = \log(3/2) \approx 0.585$$

So in the end we find that $C$ leaks 1 bit of $Y$ and 0.585 bits of $X$. While this might make it appear that $C$ is worse for $Y$ than for $X$, it is worth noting that the posterior Bayes vulnerability of $Y$ (i.e. $5/8$) is less than that of $X$ (i.e. $3/4$). This situation is possible because the prior Bayes vulnerability of $Y$ (i.e. $5/16$) is less than that of $X$ (i.e. $1/2$), and (multiplicative) leakage tells only the *factor* by which vulnerability is increased.

## IV. A Fundamental Equivalence

In this section, we establish a fundamental equivalence between Dalenius min-entropy leakage under arbitrary correlations and $g$-leakage under arbitrary gain functions.

To begin with, we recall that [10] shows that posterior $g$-vulnerability can be formulated as a matrix *trace*. (Recall that the *trace* of a square matrix is the sum of its diagonal entries.) The formulation is

$$V_g[\pi, C] = \max_S tr(\Pi C S G),$$

where

- $\Pi$ (indexed by $\mathcal{X} \times \mathcal{X}$) is a diagonal matrix with the prior $\pi$ on its diagonal,
- $C$ (indexed by $\mathcal{X} \times \mathcal{Y}$) is the channel matrix,
- $S$ (indexed by $\mathcal{Y} \times \mathcal{W}$) is a channel matrix giving the *strategy* for choosing guess $w$ from output $y$, and
- $G$ (indexed by $\mathcal{W} \times \mathcal{X}$) is the matrix representation of gain function $g$.

For we have

$$\begin{array}{ll} & V_g[\pi, C] \\ = & \sum_y \max_w \sum_x \pi_x C_{xy} g(w, x) \\ = & \max_S \sum_y \sum_w S_{yw} \sum_x \pi_x C_{xy} g(w, x) \\ = & \max_S \sum_x \sum_y \sum_w \pi_x C_{xy} S_{yw} G_{wx} \\ = & \max_S \sum_x \pi_x \sum_y C_{xy} \sum_w S_{yw} G_{wx} \\ = & \max_S \sum_x \pi_x \sum_y C_{xy} (SG)_{yx} \\ = & \max_S \sum_x \pi_x (CSG)_{xx} \\ = & \max_S \sum_x \sum_{x'} \Pi_{xx'} (CSG)_{x'x} \\ = & \max_S \sum_x (\Pi C S G)_{xx} \\ = & \max_S tr(\Pi C S G) \end{array}$$

(Note that the "max" in the third step is realized on any $S$ such that $S_{yw} > 0$ only if $w$ is a best guess given $y$.)

Next, we recall that trace satisfies a remarkable *cyclic property*: if a matrix product $AB$ is square, then

$$\begin{array}{ll} & tr(AB) \\ = & \sum_x (AB)_{xx} \\ = & \sum_x \sum_y A_{xy} B_{yx} \\ = & \sum_y \sum_x B_{yx} A_{xy} \\ = & \sum_y (BA)_{yy} \\ = & tr(BA) \end{array}$$

And, by the associativity of matrix multiplication, the cyclic property generalizes to a product of any number of matrices:

$$tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC).$$

Now we are ready to present our fundamental equivalence between *g-leakage* and *Dalenius min-entropy leakage*. More

precisely, we show that $g$-leakage with respect to an arbitrary gain function $g$ can always be converted into Dalenius min-entropy leakage under some correlation $J$, and (conversely) Dalenius min-entropy leakage under an arbitrary correlation $J$ can always be converted into $g$-leakage with respect to some gain function $g$.

*Theorem 4.1:* Let $C$ be a channel from $\mathcal{Y}$ to $\mathcal{Z}$.

- For any $\psi \in \mathbb{D}\mathcal{Y}$ and gain function $g : \mathcal{W} \times \mathcal{Y} \to [0, 1]$, there exists a correlation $J \in \mathbb{D}(\mathcal{W} \times \mathcal{Y})$ such that the $g$-leakage of $Y$ is equal to the Dalenius min-entropy leakage of $W$: $\mathcal{L}_g(\psi, C) = \mathcal{DL}(J, C)$.
- For any $J \in \mathbb{D}(\mathcal{X} \times \mathcal{Y})$, giving marginal distribution $\psi \in \mathbb{D}\mathcal{Y}$, there exists a gain function $g : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ such that the Dalenius min-entropy leakage of $X$ is equal to the $g$-leakage of $Y$: $\mathcal{DL}(J, C) = \mathcal{L}_g(\psi, C)$.

*Proof:* For the first implication, recall first that

$$\mathcal{L}_g(\psi, C) = \log \frac{V_g[\psi, C]}{V_g[\psi]}.$$

Next, by the trace formulation of posterior $g$-vulnerability and the cyclic property of trace, we have

$$V_g[\psi, C] = \max_S tr(\Psi CSG) = \max_S tr(G\Psi CS).$$

Now consider the matrix $G\Psi$, which is indexed by $\mathcal{W} \times \mathcal{Y}$. The key idea is that we can normalize $G\Psi$ to get a *joint distribution* $J$ on $\mathcal{W} \times \mathcal{Y}$, allowing us to view $W$ as a secret random variable that is correlated with $Y$.[7] Thus we get a scalar $\lambda$ and $J \in \mathbb{D}(\mathcal{W} \times \mathcal{Y})$ such that $G\Psi = \lambda J$. Moreover we can let $\pi \in \mathbb{D}\mathcal{W}$ be the marginal distribution on $W$ and choose channel $B$ from $\mathcal{W}$ to $\mathcal{Y}$ such that $J_{wy} = \pi_w B_{wy}$ for all $w$ and $y$, which is equivalent to $J = \Pi B$. Hence, continuing the reasoning above, we have

$$\begin{array}{lll} & V_g[\psi, C] & \\ = & \max_S tr(G\Psi CS) & \\ = & \max_S tr(\lambda JCS) & \text{(normalizing } G\Psi) \\ = & \lambda \max_S tr(\Pi BCSI) & \text{(factoring } J\text{, rearranging)} \\ = & \lambda V[\pi, BC] & \text{(trace formulation of Bayes vulnerability)} \\ = & \lambda DV[J, C] & \text{(Theorem 3.2)} \end{array}$$

Turning next to $V_g[\psi]$, we note that $V_g[\psi] = V_g[\psi, \mathbf{0}]$, where $\mathbf{0}$ is any channel with only 1 column, which leaks nothing. Then we can repeat the above reasoning to get $V_g[\psi, \mathbf{0}] = \lambda V[\pi, B\mathbf{0}]$. Since $B\mathbf{0} = \mathbf{0}$, we conclude that

$$V_g[\psi] = \lambda V[\pi] = \lambda DV[J].$$

Putting these two results together, we get

$$\mathcal{L}_g(\psi, C) = \log \frac{V_g[\psi, C]}{V_g[\psi]} = \log \frac{\lambda DV[J, C]}{\lambda DV[J]} = \mathcal{DL}(J, C),$$

completing the proof of the first implication.

The proof of the second implication is quite similar. Here we use the cyclic property of trace to move the correlation $J$ to the end, and viewing it as the gain function $g$. ■

[7] We are excluding the "pathological" case where $G\Psi$ is an all-zero matrix.

### A. Justifying composition refinement

Theorem 4.1 tells us that *g-leakage under arbitrary gain functions* is equivalent to *Dalenius min-entropy leakage under arbitrary correlations*. As we discuss here, this equivalence gives additional significance to the composition refinement relation (reviewed in Section II-D).

Given two channels $C$ and $D$, both taking input $Y$, recall that $C$ is composition refined by $D$, written $C \sqsubseteq_\circ D$, if $D$ can be factored into $CR$ for some $R$. If we think in terms of *program refinement* and imagine $C$ as a component of a system, then we would like $C \sqsubseteq_\circ D$ to mean that it is *safe* to replace $C$ with $D$, in the sense that the security of the system will not thereby be decreased. From this perspective, note that Theorem 2.4 expresses the *soundness* of composition refinement: if $C \sqsubseteq_\circ D$, then $D$ never leaks more than $C$, no matter the prior or gain function. And Theorem 2.5 expresses the *completeness* of composition refinement: if $C \not\sqsubseteq_\circ D$, then there exists a prior $\psi$ and gain function $g$ that make $D$ leak more than $C$, which means that it would *not* be safe to replace $C$ with $D$.

As an example, consider the following channels (from [11]):

| $C$ | $z_1$ | $z_2$ | $z_3$ |
|---|---|---|---|
| $y_1$ | $1/2$ | $1/2$ | $0$ |
| $y_2$ | $1/2$ | $0$ | $1/2$ |
| $y_3$ | $0$ | $1/2$ | $1/2$ |

| $D$ | $v_1$ | $v_2$ |
|---|---|---|
| $y_1$ | $2/3$ | $1/3$ |
| $y_2$ | $2/3$ | $1/3$ |
| $y_3$ | $1/4$ | $3/4$ |

It turns out that $C \not\sqsubseteq_\circ D$, since $D$ cannot be factored into $CR$, for any $R$. Yet under Bayes vulnerability (min-entropy leakage), $D$ never leaks more than $C$, regardless of $\psi$. But suppose that $y_1$ and $y_2$ are *male* and $y_3$ is *female*, and the adversary cares only about the *sex* of the secret, as specified by the following gain function $g_{sex}$:

| $g_{sex}$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| *male* | 1 | 1 | 0 |
| *female* | 0 | 0 | 1 |

Under a uniform prior $\psi$ and gain function $g_{sex}$, it turns out that $D$ leaks more than $C$. Gain function $g_{sex}$ seems quite reasonable, so we would surely accept that $D$ should not be considered to be as secure as $C$.

But here is an example (from [10]) where the completeness of composition refinement seems less convincing:

| $C$ | $z_1$ | $z_2$ | $z_3$ | $z_4$ |
|---|---|---|---|---|
| $y_1$ | 0.1 | 0.4 | 0.1 | 0.4 |
| $y_2$ | 0.2 | 0.2 | 0.3 | 0.3 |
| $y_3$ | 0.5 | 0.1 | 0.1 | 0.3 |

| $D$ | $v_1$ | $v_2$ | $v_3$ |
|---|---|---|---|
| $y_1$ | 0.2 | 0.22 | 0.58 |
| $y_2$ | 0.2 | 0.4 | 0.4 |
| $y_3$ | 0.35 | 0.4 | 0.25 |

Again, $C \not\sqsubseteq_\circ D$, since $D$ cannot be factored into $CR$, for any $R$. But here it is not so easy to find a gain function that make $D$ leak more that $C$! In [10], a linear-programming technique is used to find the following gain function, which (under a uniform prior $\psi$) makes $D$ leak more than $C$:

| $g_{weird}$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $w_1$ | $153/296$ | $0$ | $1/2$ |
| $w_2$ | $0$ | $289/296$ | $63/296$ |
| $w_3$ | $21/148$ | $1$ | $0$ |

But should we care about such a weird gain function? Or should $g_{weird}$ just be seen as a "monster" that no adversary would ever use?

In fact Theorem 4.1 provides an important justification for the significance of composition refinement here. For it tells us that the *$g_{weird}$-leakage of $Y$* is exactly equivalent to the *Dalenius min-entropy leakage of $W$* when $W$ and $Y$ are correlated according to $G_{weird}\Psi$. And, while we might conceivably be willing to assume that no adversary would ever use the gain function $g_{weird}$, it would seem foolish to assume that there could *never* exist a secret $W$ correlated with $Y$ according to $G_{weird}\Psi$. Given that, we should not consider $D$ to be as secure as $C$.

The conclusion is very striking: if we agree that Bayes vulnerability and min-entropy leakage are important, and we agree that Dalenius leakage due to arbitrary correlations among secrets is also important, then we *must* care about $g$-leakage under arbitrary gain functions, no matter how weird. And this means that composition refinement $\sqsubseteq_\circ$ is of fundamental importance.

## V. Bounding Dalenius Leakage

As we have seen, a channel $C$ taking a secret input $Y$ to an observable output $Z$ may cause leakage of any apparently unrelated secret $X$ that happens to have an interesting correlation $J$ with $Y$. Such Dalenius leakage is worrisome, because it seems so difficult to foresee the correlations that might be discovered to hold between secrets. For this reason, upper bounds on Dalenius leakage would be very desirable.

One very general bound can be proved from Corollary 3.3, which shows that the Dalenius leakage of $C$ can be formulated as the $g$-leakage of a *cascade* $BC$. (A very similar result was shown in [12, Corollary 22].)

*Theorem 5.1:* For any channel $C$, gain function $g$, and correlation $J$, we have $\mathcal{DL}_g(J,C) \leq \mathcal{ML}(C)$.

*Proof:* We have

$$\begin{array}{lll} & \mathcal{DL}_g(J,C) & \\ = & \mathcal{L}_g(\pi, BC) & \text{(Corollary 3.3)} \\ \leq & \mathcal{ML}(BC) & \text{(Miracle Theorem)} \\ \leq & \mathcal{ML}(C). & \text{(Theorem 2.3)} \end{array}$$

■

This is an important and robust result, since it allows us to make *no* assumptions about the correlation or the preferences of the adversary (given by the gain function). However, there are many situations in which this result might not provide a very useful bound. Consider, for instance, the case where channel $C$ leaks a large amount of information, which might be allowed if the secret $Y$ is not considered to be very sensitive. If we consider $X$ to be sensitive, then the previous bound would discourage us from using $C$, due to possible correlations between $X$ and $Y$. However, if we knew *something* about this correlation, we might reach a different conclusion. In particular, if we have reasons to belive that there is little correlation between $X$ and $Y$, then using $C$ should not harm the secrecy of $X$ too much.

As a running example, consider the following channel:

| $C$ | $z_1$ | $z_2$ | $z_3$ |
|---|---|---|---|
| $y_1$ | 0 | 0 | 1 |
| $y_2$ | $1/4$ | $3/4$ | 0 |
| $y_3$ | $3/4$ | $1/8$ | $1/8$ |

and assume there is a secret $X$ from the set $\mathcal{X} = \{x_1, x_2, x_3\}$ that is correlated to $Y$, as given by the joint probability distribution $J$. We can use Theorem 5.1 above to get a bound on the Dalenius leakage of $C$ about $X$. Since we are not going to make assumptions about the gain function $g$, we limit our analysis to bounding $\mathcal{DL}_\forall(J,C)$:

$$\mathcal{DL}_\forall(J,C) \leq \mathcal{ML}(C) = \log 2.5\,.$$

But this number is not much lower than the trivial bound (following from the fact that $|\mathcal{X}| = 3$) of $\log 3$. We could, however, aim at obtaining a better bound by making assumptions about some characteristics of the correlation $J$.

In particular, we assume that we know the *marginal probabilities* of $J$. In the case of our running example, let the marginal distributions $\pi$ and $\psi$ respectively on $\mathcal{X}$ and $\mathcal{Y}$ be as follows:

| | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $\pi$ | $1/3$ | $1/3$ | $1/3$ |

| | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $\psi$ | $1/4$ | $1/8$ | $5/8$ |

By using this information we can derive a new bound on the Dalenius leakage of $C$.

*Lemma 5.2 (Bound with marginals):* Let $\pi$ and $\psi$ be the marginal distributions on $\mathcal{X}$ and $\mathcal{Y}$ respectively, derived from the correlation $J$, and assume that $\pi$ is full support. Then

$$\mathcal{DL}_\forall(J,C) \leq \log \sum_{y\in\mathcal{Y}} \min\left(\frac{\psi_y}{\pi_{\min}}, 1\right)$$

where $\pi_{\min} = \min_{x\in\mathcal{X}} \pi_x$.

*Proof:* First, we have

$$\begin{array}{lll} & \mathcal{DL}_\forall(J,C) & \\ = & \mathcal{L}_\forall(\pi, BC) & \text{(Corollary 3.3)} \\ \leq & \mathcal{ML}(BC) & \text{(Miracle Theorem)} \\ \leq & \mathcal{ML}(B) & \text{(Theorem 2.3)} \\ = & \log \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} B_{xy} & \text{(Theorem 2.1)} \end{array}$$

Now, we note that $J_{xy} = \pi_x B_{xy}$, and therefore

$$\psi_y = \sum_{x\in\mathcal{X}} J_{xy} = \sum_{x\in\mathcal{X}} \pi_x B_{xy} \tag{1}$$

Finally, if we define $\pi_{\min} = \min_{x\in\mathcal{X}} \pi_x$, and recall that $\pi_{\min} > 0$ (since $\pi$ is full support), then:

$$\begin{array}{lll} & \max_{x\in\mathcal{X}} B_{xy} & \\ \leq & \sum_{x\in\mathcal{X}} B_{xy} & \text{(algebra)} \\ = & (1/\pi_{\min}) \sum_{x\in\mathcal{X}} \pi_{\min} B_{xy} & \text{(Mult. by 1)} \\ \leq & (1/\pi_{\min}) \sum_{x\in\mathcal{X}} \pi_x B_{xy} & (\pi_{\min} \leq \pi_x, \quad \forall x) \\ = & \psi_y/\pi_{\min} & \text{(eq. 1)} \end{array}$$

We must note however that even though $\psi_y/\pi_{\min}$ can be greater than 1, we know $\max_{x\in\mathcal{X}} B_{xy}$ is not. Therefore, we can conclude that

$$\max_{x\in\mathcal{X}} B_{xy} \leq \min\left(\frac{\psi_y}{\pi_{\min}}, 1\right)$$

and hence

$$\mathcal{DL}_\forall(J,C) \leq \log \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} B_{xy} \leq \log \sum_{y\in\mathcal{Y}} \min\left(\frac{\psi_y}{\pi_{\min}}, 1\right).$$

■

In the case of our running example, we can use this new result to obtain a better bound:

$$\mathcal{DL}_\forall(J,C) \leq \log\left(\frac{1/4}{1/3} + \frac{1/8}{1/3} + 1\right) = \log 2.125\,.$$

It is worth noting that although in this particular case the new bound is lower than $\mathcal{ML}(C)$, there might be cases where it is actually higher. Of course, since both are upper bounds on the Dalenius leakage, we can take the minimum of the two.

The previous result arises as an attempt to find an upper bound on the min-capacity of channel $B$. However, this is a pessimistic bound, in the sense that there are cases where there is no valid $B$ with such min-capacity. Therefore, in order to obtain an optimal bound we would need to calculate the actual channel $B_{\max}$ with the maximum min-capacity that agrees with marginal distributions $\pi$ and $\psi$.

It is indeed possible to obtain such a channel by using a method that solves an exponential number of linear optimization problems. The method relies on fixing the location of the maximum of each column of channel $B$ (recall from Theorem 2.1 that the min-capacity is equal to the logarithm of the sum of the column maximums) and then calculates, using linear optimization, the entries of $B$ such that

- $B$ agrees with the given marginal distributions
- $B$ has the column maximums in the locations specified
- the sum of the column maximums of $B$ is maximized.[8]

If we now iterate over all the possible placements of the column maximums on $B$ (we have $|\mathcal{X}|^{|\mathcal{Y}|}$ possibilities), and solve the corresponding linear program for each one, we can retrieve the channel $B$ with maximum min-capacity.

More concretely, given a mapping $f : \mathcal{Y} \to \mathcal{X}$, where $f(y)$ represents the value of $x$ that holds the maximum of column $y$, the corresponding linear program is as follows:

| | | |
|---|---|---|
| **Given:** | $\pi_x$ | $\forall x$ |
| | $\psi_y$ | $\forall y$ |
| | $f$ | |
| **Calculate:** | $B_{xy}$ | $\forall x, y$ |
| **That maximize:** | $\sum_y B_{f(y)y}$ | |
| **Subject to:** | $B_{xy} \leq B_{f(y)y}$ | $\forall x, y$ |
| | $\sum_x \pi_x B_{xy} = \psi_y$ | $\forall y$ |
| | $\sum_y B_{xy} = 1$ | $\forall x$ |
| | $B_{xy} \geq 0$ | $\forall x, y$ |

Solving this linear program will give us a $B$ of maximum min-capacity that is consistent with the mapping $f$. Since there are $|\mathcal{X}|^{|\mathcal{Y}|}$ possible mappings, we need to solve the linear program for each of them in order to get a $B$ of maximum min-capacity overall. In the case of our running example, we get the following result, which we call $B_{\max}$:

| $B_{\max}$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $x_1$ | $3/4$ | $0$ | $1/4$ |
| $x_2$ | $0$ | $3/8$ | $5/8$ |
| $x_3$ | $0$ | $0$ | $1$ |

Now we can use the min-capacity of $B_{\max}$ as a bound. Surprisingly, here we get the same result as with the previous bound:

$$\mathcal{DL}_\forall(J,C) \leq \mathcal{ML}(B_{\max}) = \log(3/4 + 3/8 + 1) = \log 2.125\,.$$

This means that, in this particular example, the bound shown in Lemma 5.2 is as good as it can possibly be, since there actually is a correlation with marginals $\pi$ and $\psi$ such that matrix $B$ derived from it has a min-capacity of $\log 2.125$.

Now, since we have found a channel $B_{\max}$ with the maximum min-capacity, we might be tempted to go one step further and calculate the min-capacity of the cascade $B_{\max}C$ in order to obtain a better upper bound. However, this would be *incorrect*: the fact that $B_{\max}$ has the greatest min-capacity among all the possible "valid" channels $B$ does not mean that it is the channel that maximizes the min-capacity of the *cascade* with $C$.

In fact we can calculate a channel $B$ that yields the maximum min-capacity of the cascade $BC$ by again solving an exponential number of linear programming problems, here slightly different than the ones presented before. In particular, in each linear program we need to fix a mapping $h : \mathcal{Z} \to \mathcal{X}$, where $h(z)$ denotes the value of $x$ that holds the maximum value of the $z$ column of the cascade $BC$. Each of the linear

[8] Note that maximizing this quantity will maximize the min-capacity, since logarithm is a monotonic function.

programs then is as follows:

| | | |
|---|---|---|
| **Given:** | $\pi_x$ | $\forall x$ |
| | $\psi_y$ | $\forall y$ |
| | $C_{yz}$ | $\forall y, z$ |
| | $h$ | |
| **Calculate:** | $B_{xy}$ | $\forall x, y$ |
| **That maximize:** | $\sum_{y,z} B_{f(z)y} C_{yz}$ | |
| **Subject to:** | $\sum_y B_{xy} C_{yz} \leq \sum_y B_{f(z)y} C_{yz}$ | $\forall x, z$ |
| | $\sum_x \pi_x B_{xy} = \psi_y$ | $\forall y$ |
| | $\sum_y B_{xy} = 1$ | $\forall x$ |
| | $B_{xy} \geq 0$ | $\forall x, y$ |

(It should be noted however that these new linear programs might be much more expensive than the original ones, particularly in the case where $C$ has far more columns than rows.) Again, by solving each linear program we get a $B$ that makes the min-capacity of $BC$ as large as possible when $BC$ is consistent with $h$. There are $|\mathcal{X}|^{|\mathcal{Z}|}$ possible such mappings, and we have to solve the linear programs for all of them in order to get the value of $B$ that maximizes the min-capacity of $BC$ in general. In our running example, we calculate this value and call it $B^\star$:

| $B^\star$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $x_1$ | 0 | 0 | 1 |
| $x_2$ | 0 | $3/8$ | $5/8$ |
| $x_3$ | $3/4$ | 0 | $1/4$ |

The corresponding cascade and its min-capacity are as follows:

| $B^\star C$ | $z_1$ | $z_2$ | $z_3$ |
|---|---|---|---|
| $x_1$ | $3/4$ | $1/8$ | $1/8$ |
| $x_2$ | $9/16$ | $23/64$ | $5/64$ |
| $x_3$ | $3/16$ | $1/32$ | $25/32$ |

and

$$\begin{aligned} \mathcal{DL}_\forall(J, C) &\leq \mathcal{ML}(B^\star C) \\ &= \log(3/4 + 23/64 + 25/32) \approx \log 1.891\,. \end{aligned}$$

Note that this is the *exact* Dalenius capacity of $C$ over all correlations $J$ that are consistent with the marginals $\pi$ and $\psi$.

As a final possibility, we can assume we *know* the correlation $J$. In this case, it is possible to calculate $\mathcal{DL}_\forall(J, C)$ exactly. In our running example, let the correlation $J$ be as follows:

| $J$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $x_1$ | $1/6$ | 0 | $1/6$ |
| $x_2$ | 0 | $1/24$ | $7/24$ |
| $x_3$ | $1/12$ | $1/12$ | $1/6$ |

(Note that the marginal distributions of $J$ match the distributions $\pi$ and $\psi$ presented before.) The corresponding channel $B$ for this correlation is

| $B$ | $y_1$ | $y_2$ | $y_3$ |
|---|---|---|---|
| $x_1$ | $1/2$ | 0 | $1/2$ |
| $x_2$ | 0 | $1/8$ | $7/8$ |
| $x_3$ | $1/4$ | $1/4$ | $1/2$ |

With this, we can now calculate the cascade $BC$:

| $BC$ | $z_1$ | $z_2$ | $z_3$ |
|---|---|---|---|
| $x_1$ | $3/8$ | $1/6$ | $9/16$ |
| $x_2$ | $11/16$ | $13/64$ | $7/64$ |
| $x_3$ | $7/16$ | $1/4$ | $5/16$ |

Finally, we use it to calculate the exact value of the Dalenius leakage:

$$\mathcal{DL}_\forall(J, C) = \mathcal{ML}(BC) = \log(11/16 + 1/4 + 9/16) = \log 1.5\,.$$

We conclude by comparing the different bounds achieved throughout this running example:

| | |
|---|---|
| Bound using $\mathcal{ML}(C)$ | $\log 2.5$ |
| Bound on $\mathcal{ML}(B)$ with marginals | $\log 2.125$ |
| Bound on $\mathcal{ML}(B)$ from linear programming | $\log 2.125$ |
| Bound on $\mathcal{ML}(BC)$ from linear programming | $\log 1.891$ |
| Actual value of $\mathcal{DL}_\forall(J, C)$ | $\log 1.5$ |

## VI. Additive Dalenius Leakage

As mentioned in Section II, leakage can also be defined *additively*, as the difference between the posterior and prior vulnerabilities:

*Definition 6.1 (Additive Leakage):*

- $\mathcal{L}_g^+(\pi, C) = V_g[\pi, C] - V_g[\pi]$ is the additive leakage of $C$ with respect to prior $\pi$ and gain function $g$.
- $\mathcal{DL}_g^+(J, C) = DV_g[\pi, C] - DV_g[J]$ is the additive Dalenius leakage with respect to correlation $J$ and gain function $g$.

In this section, we briefly consider the extent to which our results about multiplicative Dalenius leakage can be carried over to the additive case.

First, it is easy to see that Corollary 3.3 carries over directly:

*Corollary 6.1:* If $B$ is any channel from $\mathcal{X}$ to $\mathcal{Y}$ satisfying $J_{xy} = \pi_x B_{xy}$ for all $x$ and $y$, then $\mathcal{DL}_g^+(J, C) = \mathcal{L}_g^+(\pi, BC)$.

Unfortunately, the fundamental equivalence between standard $g$-leakage and Dalenius min-entropy leakage (Theorem 4.1) is not true for additive leakage. The trouble is that the normalization factor $\lambda$, which cancels out in the multiplicative case, *scales* the leakage in the additive case. Since this normalization factor is somewhat arbitrary, it is unclear whether anything useful can be achieved in this case.

Bounds on additive Dalenius $g$-leakage are given in [12], which includes a bound that holds under any correlation $J$ and gain function $g$:

$$\mathcal{DL}_\forall^+(J, C) \leq \mathcal{L}_\forall^+(\forall, C)\,.$$

However, the authors note that the most efficient way to calculate this value seems to be by solving a quadratic optimization problem. A more promising scenario is introduced by assuming that the marginal distribution on $\mathcal{Y}$, denoted by $\psi$, is known. Here, the authors discover a bound that can be calculated efficiently, using the "earth-moving" distance metric for hyper-distributions.

$$\mathcal{DL}_\forall^+(J, C) \leq \mathcal{L}_\forall^+(\psi, C)\,.$$

Here we go one step further and present a new bound, this time for the case in which both marginal distributions $\pi$ and $\psi$ are known.

We start by proving a useful property of the Manhattan distance between probability distributions:

*Lemma 6.2:* Let $\pi$ be a probability distribution on set $\mathcal{X}$, and choose $\hat{x} \in \mathcal{X}$ such that $\pi_{\hat{x}} = \min_{x \in \mathcal{X}} \pi_x$. Then, the probability distribution that maximizes the Manhattan distance with $\pi$ is the dirac distribution $\pi^*$, where $\pi^*_{\hat{x}} = 1$ and $\pi^*_x = 0, \forall x \neq \hat{x}$. The Manhattan distance between $\pi$ and $\pi^*$ is then $2(1 - \pi_{\hat{x}})$.

*Proof:* We will first see that $\pi^*$ is a dirac distribution. Suppose it is not, that is, there are $x, x' \in \mathcal{X}$ such that $\pi^*_x > 0$ and $\pi^*_{x'} > 0$. Then, it can be seen that distribution $\psi$ such that

$$\begin{aligned} &\psi_x = \pi^*_x + \pi^*_{x'} \\ &\psi_{x'} = 0 \\ &\psi_{x''} = \pi^*_{x''} \qquad\qquad \forall x'' \neq x, x' \end{aligned}$$

has at least the same Manhattan distance to $\pi$ as $\pi^*$. Therefore, we can consider $\pi^*$ to be $\psi$ instead, and that way it would be "closer" to be a dirac distribution (since the probability of $x'$ has been changed to 0). Since we can repeat this reasoning until $\pi^*$ has only one non-zero element, it is safe to assume that $\pi^*$ is a dirac distribution.

Finally, the dirac distribution that has the greatest Manhattan distance with $\pi$ is such that the probability is concentrated in the element with least mass in $\pi$, which we know is $\hat{x}$. ■

With the previous lemma, we can derive the following bound for the case of additive Dalenius leakage:

*Lemma 6.3:* Let $\pi$ and $\psi$ be the marginal distributions corresponding to correlation matrix $J$, and assume they are full-support. Let $B$ be a channel such that $J_{xy} = \pi_x B_{xy}$, and let $\psi_{\min} = \min_{y \in \mathcal{Y}} \psi_y$. Then,

$$\mathcal{DL}^+_\forall(J, C) \leq 1 - \max(\pi_{\min}, \psi_{\min})$$

*Proof:* By using the fact that the additive leakage is equal to the Kantorovich distance between the prior and the posterior hypers, we can get the following formula:

$$\mathcal{L}^+_\forall(\pi, B) = \frac{1}{2} \sum_{y \in \mathcal{Y}} \psi_y \sum_{x \in \mathcal{X}} \left| \pi_x - \frac{\pi_x B_{xy}}{\psi_y} \right| \tag{2}$$

We can simplify the previous expression and derive the following:

$$\mathcal{L}^+_\forall(\pi, B) = \frac{1}{2} \sum_{x \in \mathcal{X}} \pi_x \sum_{y \in \mathcal{Y}} |\psi_y - B_{xy}| \tag{3}$$

Now, if we consider Equation 2, we can see that the expression

$$\sum_{x \in \mathcal{X}} \left| \pi_x - \frac{\pi_x B_{xy}}{\psi_y} \right|$$

represents the Manhattan distance between $\pi$ and the posterior distribution corresponding to $y$. By using Lemma 6.2, we know that this must be at most $2(1 - \pi_{\min})$. By replacing this in equation 2, we get the following:

$$\begin{aligned} \mathcal{L}^+_\forall(\pi, B) &\leq \frac{1}{2} \sum_{y \in \mathcal{Y}} 2\psi_y (1 - \pi_{\min}) \\ &= (1 - \pi_{\min}) \sum_{y \in \mathcal{Y}} \psi_y \\ &= 1 - \pi_{\min} \end{aligned}$$

By using similar reasoning for equation 3 (where now the distributions are $\psi$ and $B_{x-}$) we can derive that

$$\mathcal{L}^+_\forall(\pi, B) \leq 1 - \psi_{\min}$$

Finally, we can now conclude that

$$\mathcal{L}^+_\forall(\pi, B) \leq 1 - \max(\pi_{\min}, \psi_{\min})$$

■

Similar to the case with multiplicative leakage, this bound might be overly pessimistic, in the sense that there might be no valid correlation that achieves the leakage derived in the bound. However, as in the multiplicative case, an "optimal" bound could be obtained by calculating the channel matrix $B$ that yields the maximum additive leakage with respect to the given marginals. This could be modeled as an optimization problem with objective function $\frac{1}{2} \sum_{x,y} \pi_x |\psi_y - B_{xy}|$. Unfortunately, as noted in [12], the presence of $O(|\mathcal{X}||\mathcal{Y}|)$ absolute value terms in the objective function seems to potentially make this problem harder to solve in general than a quadratic optimization problem. In fact, solving an optimization problem that requires maximization of an objective function with absolute values is in general NP-hard, since one particular instance of such a problem (namely, finding the $\|\cdot\|_{\infty,1}$ norm of a matrix) has been proven to be [17].

## VII. Related Work

We have adopted the name "Dalenius leakage" in honor of Tore Dalenius, whose 1977 paper [3] anticipated many of the concerns of modern quantitative information flow. Working in the context of census data, Dalenius considered the sorts of information leakage (which he called "disclosure") that could result from releasing census statistics, and gave the following definition:

> If the release of statistics $S$ makes it possible to determine the value $D_K$ more accurately than is possible without access to $S$, a disclosure has taken place. [3, p. 433]

This is the apparent source of Dwork's characterization of the "Dalenius Desideratum", but in fact Dalenius does *not* make elimination of disclosure a *desideratum*. Indeed, contrary to the impression given by Dwork's account,[9] Dalenius explicitly recognizes the need to accept some disclosure. He writes,

[9]Consider, for example, Dwork's sentence "The last hopes for Dalenius's goal evaporate in light of the following parable, which again involves auxiliary information." [2, p. 90]

> A reasonable starting point is to discard the notion of *elimination* of disclosure. Two arguments for doing so are:
>
> i. it would be unrealistic to aim at elimination: such a goal is not operationally feasible;
> ii. it would place unreasonable restrictions on the kind of statistics that can be released; it may be argued that elimination of disclosure is possible only by elimination of statistics.
>
> What has just been said is in fact the reason for our use of the term "statistical disclosure *control*" rather than "prevention" or "avoidance". [3, pp. 439–440]

Moreover, Dalenius was well aware of the possibility of information leakage due to auxiliary information (as in Dwork's parable about Lithuanian woman and Alan Turing). Denoting by $S$ the statistics released from the survey and by $E$ the "extra-objective data", he writes,

> $S \times E$-based disclosure may easily prove to be a much more serious problem that $S$-based disclosure: the statistician may not know about $E$, or—if he does—he may not have the authority to control it. [3, pp. 441–442]

Finally, Dalenius even suggests the need for two measures: $M$, "giving the amount of disclosure associated with the release of some statistics and the extra-objective data", and $B$, giving "the benefit associated with the statistics", and suggests using a criterion for statistical disclosure control that maximizes $B$ for some accepted level $M_0$ of disclosure. [3, p. 440]

As already mentioned, *differential privacy* [1] was largely motivated by concerns about correlations among secrets, and is aimed at ensuring that an individual does not suffer significant harm by agreeing to participate in a statistical database. An important cautionary note about the privacy protections ensured by differential privacy is given by Kifer and Machanavajjhala [18]. This paper argues that, for datasets where there is correlation between data, differential privacy guarantees are not sufficient in the case where an adversary is trying to determine whether or not an individual has participated in the database. This correlation may come in different forms. Examples studied in the paper include data generation procedures (e.g. in a social network, the presence of an individual may cause edges to form between this person's friends), and deterministic query answers released in the past (e.g. counts of the number of records having values within a given set). In both cases, the correlation allows as adversary to guess accurately whether an individual participated in the dataset or not, even when the queries are answered with a mechanism satisfying differential privacy.

More recently, Liu, Chakraborty, and Mittal [19] argue that standard differential privacy is insufficient when correlations are present in the database. They propose an alternative definition, *Dependent Differential Privacy*, which takes into account how many entries of a database are affected when a change is made to any particular record. The authors also present an obfuscation mechanism that satisfies this definition, based on a "dependence coefficient" between databases entries. However, the construction of the mechanism requires the dependence coefficient to be known, or at least accurately estimated.

An important inspiration for our Section IV-A was given by McIver, Meinicke, and Morgan [9]. That work identifies the crucial *composition refinement* relation $\sqsubseteq_\circ$ (defined there on sequential imperative programs, rather than on channels) and justifies its strength by arguing that if program $P$ is not composition refined by program $Q$, $P \not\sqsubseteq_\circ Q$, then there is a context $\mathcal{C}$ (involving Markov updates to secret variables) such that the posterior Bayes vulnerability of $\mathcal{C}(Q)$ is greater than that of $\mathcal{C}(P)$. Analogously, our Theorem 4.1 allows us to conclude that if $C \not\sqsubseteq_\circ D$, then there is a "Dalenius context" in which the posterior Bayes vulnerability (and min-entropy leakage) of $D$ is greater than that of $C$. In both cases, we find that if we accept the importance of Bayes vulnerability over a sufficiently rich set of contexts, then we are compelled also to accept the importance of composition refinement.

The first study of Dalenius leakage is given by Alvim et al. [12]. It shows that Dalenius leakage can be formulated as the $g$-leakage of a cascade, allowing bounds on the capacity of cascades to be applied to Dalenius leakage, both multiplicative and additive. Our paper goes beyond that work by establishing several new fundamental properties of Dalenius leakage and its relation to composition refinement, such as Corollary 3.3 and Theorem 4.1. Also new is our exploration of the case when the marginals induced by the correlation $J$ are known, where we show that stronger bounds can be achieved.

## VIII. Conclusion

This paper has explored information leakage resulting from correlations between secrets, which we call Dalenius leakage. The equivalence that we have established between *Dalenius leakage under arbitrary correlations* and *$g$-leakage under arbitrary gain functions* lets us understand Dalenius leakage within the framework of $g$-leakage. It enables us to establish general bounds on Dalenius capacity or even to compute precise bounds in certain situations (though only on very small channels) if we assume that some properties of the correlation are known. In future work, it will be important to apply these results to real-world scenarios in order to assess their practical usefulness.


## Acknowledgments

The authors are grateful to the anonymous referees for their careful reading of our paper and their insightful comments. We also wish to thank Mário Alvim, Kostas Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Marco Stronati for fruitful discussions of this work; we especially thank Marco for pointing out the cyclic property of trace to us and discussing its significance for posterior $g$-vulnerability. Geoffrey Smith and Nicolás Bordenabe were partially supported by the National Science Foundation under grant CNS-1116318. Nicolás Bordenabe was also supported by the Australian government under ARC grant DP140101119.

## REFERENCES


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

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

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *Proc. 27th IEEE Computer Security Foundations Symposium (CSF 2014)*, 2014, pp. 308–322.

[13] 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.

[14] 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.

[15] D. Blackwell, "Comparison of experiments," in *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 93–102.

[16] J. Crémer, "A simple proof of Blackwell's 'Comparison of experiments' theorem," *Journal of Economic Theory*, vol. 27, pp. 439–443, 1982.

[17] J. Rohn, "Computing the norm $\|A\|_{\infty,1}$ is NP-hard," *Linear and Multilinear Algebra*, vol. 47, no. 3, pp. 195–204, 2000.

[18] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proc. 2011 ACM SIGMOD International Conference on Management of Data*, 2011, pp. 193–204.

[19] C. Liu, S. Chakraborty, and P. Mittal, "Dependence makes you vulnerable: Differential privacy under dependent tuples," in *Proc. Network and Distributed System Security Symposium (NDSS '16)*, Feb. 2016.