Connections between g-leakage and the Dalenius desideratum

Geoffrey Smith
Florida International University

QIF Day (PRINCESS workshop), 16 December 2014

I. Concepts of Quantitative Information Flow (QIF)

- We wish to quantify the leakage of a secret input X to an observable output Y caused by a probabilistic channel C.
  - Example: Y = X & 0x1ff leaks 9 bits of X, intuitively.
- The possible values of X and Y are given by finite sets $\mathcal{X}$ and $\mathcal{Y}$.
- There is a prior distribution π on $\mathcal{X}$.
- Both π and C are assumed known by the adversary $\mathcal{A}$.
- Then the (information-theoretic) essence of C is a mapping from priors π to hyper-distributions [π,C].

Example

Prior

π
3/8
3/8
1/4

Channel matrix

C	y_1	y_2
x_1	2/3	1/3
x_2	2/3	1/3
x_3	1/4	3/4

Multiply each row by prior probability.

Joint matrix

J	y_1	y_2
x_1	1/4	1/8
x_2	1/4	1/8
x_3	1/16	3/16

Abstractly, channel C is a mapping from priors to hyper-distributions.

Add up each column.

Distribution on Y

p_Y	9/16	7/16

Normalize columns of joint matrix.

Posterior distributions

	$p_{X\|y_1}$	$p_{X\|y_2}$
x_1	4/9	2/7
x_2	4/9	2/7
x_3	1/9	3/7

Forget about column labels.

Hyper-distribution on X

	9/16	7/16
x_1	4/9	2/7
x_2	4/9	2/7
x_3	1/9	3/7

Vulnerability and min-entropy leakage

- [Smith09] proposed to measure leakage based on X's vulnerability to be guessed by $\mathcal{A}$ in one try.
- Prior vulnerability:
  $V[\pi] = \max_x \pi_x$
- Posterior vulnerability:
  $V[\pi,C] = \sum_y p(y)\, V[p_{X|y}]$
  - $V[\pi,C]$ is the average vulnerability in the hyper-distribution.
  - $V[\pi,C]$ is the complement of the Bayes risk.
- Min-entropy leakage:
  $\mathcal{L}(\pi,C) = \lg (V[\pi,C] / V[\pi])$

Operational significance of vulnerability

- $V[\pi]$ is an optimal adversary $\mathcal{A}$'s probability of winning the following game:

  $x \xleftarrow{\$} \pi$

  $w \xleftarrow{\$} \mathcal{A}(\pi)$

  if $w = x$ then **win** else **lose**

- $V[\pi, C]$ is an optimal adversary $\mathcal{A}$'s probability of winning the following game:

  $x \xleftarrow{\$} \pi$

  $y \xleftarrow{\$} C_{x,-}$

  $w \xleftarrow{\$} \mathcal{A}(\pi, C, y)$

  if $w = x$ then **win** else **lose**

Generalizing to g-vulnerability [ACPS12]

- Finite set $\mathcal{W}$ of guesses about X (or "actions").
- Gain (or "scoring") function $g : \mathcal{W} \times \mathcal{X} \rightarrow [0, 1]$
  - g(w,x) gives the value of w if the secret is x.
  - Can model scenarios where the adversary benefits by guessing X partially, approximately, in k tries, …
- Note: (Ordinary) vulnerability implicitly uses

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

- **Prior g-vulnerability**: $V_g[\pi] = \max_w \sum_x \pi_x \, g(w,x)$
- **Posterior g-vulnerability**: $V_g[\pi,C] = \sum_y p(y) \, V_g[p_{X|y}]$

g-leakage

- g-leakage is defined based on the prior and posterior g-vulnerability.
- But there are a number of plausible definitions:
  - "logged" multiplicative: $\lg (V_g[\pi,C] / V_g[\pi])$
  - additive: $V_g[\pi,C] - V_g[\pi]$
  - multiplicative: $V_g[\pi,C] / V_g[\pi]$
- Fortunately, if we just want to compare the leakage of two channels, these all give the same result!
- We **always** get

$$\mathcal{L}_g(\pi,A) \leq \mathcal{L}_g(\pi,B) \quad \text{iff} \quad V_g[\pi,A] \leq V_g[\pi,B] .$$

II. "Dalenius's Desideratum"

- [Dwork11]: "In 1977…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."
- "…The last hopes for Dalenius's goal evaporate in light of the following parable…"
- "Given the auxiliary information 'Turing is two inches taller than the average Lithuanian woman', access to the statistical database teaches Turing's height."
- (Actually, Dwork's account appears to be completely unfair to Dalenius…)

A "Dalenius" QIF scenario

- Imagine a secret X with prior π.
- Suppose adversary $\mathcal{A}$ is interested in learning X, measuring knowledge with a gain function g.
- Now imagine a channel C from Y to Z, apparently having **nothing** to do with X.
- But suppose there is an interesting joint matrix J on (X,Y), expressing a correlation between X and Y.
  - (J must give marginal distribution π to X.)
- Can we see C as leaking information about X?

The Dalenius scenario with g-leakage

- Given channel C from X to Y, we can construct C^* from (X,Y) to Z:
  - $C^*_{(x,y),z} = C_{y,z}$
  - C^* ignores X.
- Given gain function g from W to X, we can construct g^* from W to (X,Y):
  - $g^*(w,(x,y)) = g(w,x)$
  - g^* ignores Y.
- Hence $\mathcal{L}_{g^*}(J,C^*)$ can be seen as the leakage about X caused by C, given the correlations in J.

A neater formulation

- The joint matrix J can of course be converted into the prior π on X and a channel matrix B from X to Y.
- We can cascade B and C to get a channel BC from X to Z.
- And it turns out (a bit mysteriously, to me) that $\mathcal{L}_g(\pi, BC) = \mathcal{L}_{g^*}(J, C^*)$.
- One nice consequence (thanks to theorems about cascading) is that this "Dalenius" leakage of X cannot exceed the capacity of C, no matter what correlations J may ever be discovered to exist!

III. Another application of Dalenius scenarios

- Given channels A and B on input X, the question of which leaks more will ordinarily depend on π and g.
- Is there a **robust** ordering?
- Yes!
- **Coriaceous Theorem**:
  A never leaks more than B, regardless of π and g
  iff
  A can be factored into BR, for some channel R.
- Proved in [MMSEM14], but proved in the early 1950s by statistician David Blackwell.

Example

$A =$

	z_1	z_2
x_1	2/3	1/3
x_2	2/3	1/3
x_3	1/4	3/4

$B =$

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

- A cannot be factored into BR, for any R.
- Yet under ordinary vulnerability (min-entropy leakage), A never leaks more than B, regardless of π.
- But suppose that x_1 and x_2 are **male** and x_3 is **female**, and the adversary uses a gain function that cares only about the gender of the secret.
- In that case A leaks more than B.

A less convincing example

A =

	z_1	z_2	z_3
x_1	0.2	0.22	0.58
x_2	0.2	0.4	0.4
x_3	0.35	0.4	0.25

B =

	y_1	y_2	y_3	y_4
x_1	0.1	0.4	0.1	0.4
x_2	0.2	0.2	0.3	0.3
x_3	0.5	0.1	0.1	0.3

- Again, A cannot be factored into BR, for any R.
- Here's a gain function that makes A leak more than B:

g	x_1	x_2	x_3
w_1	153/296	0	1/2
w_2	0	289/296	63/296
w_3	21/148	1	0

- Why should we care about such weird gain functions?

The trace formulation of g-vulnerability

- Recall that we can express *g*-vulnerability as a **trace**.
- The trace of a square matrix is the sum of its diagonal entries.
- $V_g[\pi, C] = \max_S \mathrm{tr}(D_\pi C S G)$
  - D_π (indexed by X,X) is a diagonal matrix of the prior
  - C (indexed by X,Y) is the channel matrix
  - S (indexed by Y,W) is the strategy for choosing guess w from output y
  - G (indexed by W,X) is the gain function

Gain functions as Dalenius scenarios

- Amazingly, trace satisfies a cyclic property:
  $tr(ABC) = tr(BCA) = tr(CAB)$
- Hence we have
  $V_g[\pi,C] = \max_S tr(D_\pi CSG)$
  $= \max_S tr(GD_\pi CS)$
  $= \max_S tr((GD_\pi)CSI)$
- I (identity matrix) gives ordinary vulnerability.
- And note that GD_π can always be normalized to a **joint matrix** J between W and X!
- Hence we can see the g-leakage of X caused by C as the min-entropy leakage of W caused by C when W and X are correlated according to GD_π.

Example, revisited

A	z_1	z_2	z_3
x_1	0.2	0.22	0.58
x_2	0.2	0.4	0.4
x_3	0.35	0.4	0.25

B	y_1	y_2	y_3	y_4
x_1	0.1	0.4	0.1	0.4
x_2	0.2	0.2	0.3	0.3
x_3	0.5	0.1	0.1	0.3

g	x_1	x_2	x_3
w_1	153/296	0	1/2
w_2	0	289/296	63/296
w_3	21/148	1	0

- With a uniform prior, A's g-leakage of X exceeds B's.
- And if W is regarded as a secret, and it is correlated with X according to g, then A's min-entropy leakage of W exceeds B's.
- So **if** we care about min-entropy leakage under arbitrary correlations **then** we also need to care about g-leakage for all g, no matter how weird!

IV. [Dalenius77]

- The apparent source of Dwork's characterization of the "Dalenius Desideratum":
  "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."
- But Dalenius does not make this a desideratum!
- On the contrary:
  "A reasonable starting point is to discard the notion of **elimination** of disclosure."
  "It may be argued that elimination of disclosure is possible only by elimination of statistics."
  "[This] is the reason for our use of the term 'statistical disclosure **control**' rather than 'prevention' or 'avoidance'."
  "More specifically, we need two measures: M = the amount of disclosure associated with the release of some statistics; and B = the benefit associated with the statistics."

Questions?