Additive and multiplicative notions of leakage, and their capacities

Mário Alvim Kostas Chatzikokolakis
Annabelle McIver Carroll Morgan
Catuscia Palamidessi Geoffrey Smith

NSA Best Cybersecurity Paper Award Ceremony, 10 November 2015

Authors

Mário Alvim

Kostas Chatzikokolakis

Annabelle McIver

Carroll Morgan

Catuscia Palamidessi

Geoffrey Smith

Motivation

- Protecting confidential information from improper leakage is a fundamental issue in computer security.
- But perfection is often infeasible, because a system's observable output often depends on its secret input:
  - When a password checker rejects an incorrect guess, it reveals a value that the secret password is **not**.
  - The tally of votes in an election reveals information about the secret ballots cast.
  - The timing of cryptographic operations may reveal information about the secret key.
- To argue that certain leaks are "small", we want a theory of quantitative information flow.

Plan of the talk

- Motivation
- Technical background
  - Vulnerability, gain functions, channels, hyper-distributions, multiplicative and additive g-leakage
- This paper
  - Significance of additive leakage
  - Six channel capacity scenarios
  - "Dalenius" leakage

Secrets and their vulnerability

- A secret X is something about which the adversary knows only a probability distribution π.
  - π is clear if X is a **randomly-generated string of bits**.
  - If X is **my mother's maiden name**, then π must instead reflect the adversary's knowledge of the population I come from.
- We wish to quantify the vulnerability of a secret with distribution π.
- **Definition**: $V[\pi] = \max_x \pi_x$
  - This is an optimal adversary's probability of guessing X correctly in one try.
  - **Example**: If X is the sum of two fair dice, $V[\pi] = 1/6$.

Generalizing to g-vulnerability

- But maybe an adversary can benefit by guessing a value close to X or a property of X, or gets three tries, or is penalized for an incorrect guess.
- We can model each scenario with a gain function g.
  - Finite set $\mathcal{W}$ of "guesses" (or "actions") about X.
  - $g : \mathcal{W} \times \mathcal{X} \to [0, 1]$
  - g(w,x) gives the value of guess w if the secret is x.
  - Ordinary vulnerability implicitly uses $g_{id}(w,x) = \begin{cases} 1, \text{ if } w = x \\ 0, \text{ otherwise} \end{cases}$
- **Definition**: g-vulnerability $V_g[\pi] = \max_w \sum_x \pi_x g(w,x)$
  - the maximum expected gain over all possible guesses

Information-theoretic channels

- A (probabilistic) system taking secret input X to observable output Y is modeled with a channel matrix C of the conditional probabilities p(y|x):

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

- Both prior π and C are assumed known by the adversary.
- **Key insight**: The information-theoretic essence of C is a mapping from priors π to hyper-distributions $[\pi,C]$, which are distributions on posterior distributions.

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Scale rows with π.

Joint matrix

J	y_1	y_2	y_3	y_4
x_1	1/8	1/8	0	0
x_2	0	1/8	1/4	1/8
x_3	1/8	1/12	1/24	0

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Scale rows with π.

Joint matrix

J	y_1	y_2	y_3	y_4
x_1	1/8	1/8	0	0
x_2	0	1/8	1/4	1/8
x_3	1/8	1/12	1/24	0

Add up columns.

Distribution on Y

p_Y	1/4	1/3	7/24	1/8

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Scale rows with π.

Joint matrix

J	y_1	y_2	y_3	y_4
x_1	1/8	1/8	0	0
x_2	0	1/8	1/4	1/8
x_3	1/8	1/12	1/24	0

Add up columns.

Distribution on Y

p_Y	1/4	1/3	7/24	1/8

Normalize columns.

Posterior distributions

	$p_{X\|y1}$	$p_{X\|y2}$	$p_{X\|y3}$	$p_{X\|y4}$
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Scale rows with π.

Joint matrix

J	y_1	y_2	y_3	y_4
x_1	1/8	1/8	0	0
x_2	0	1/8	1/4	1/8
x_3	1/8	1/12	1/24	0

Add up columns.

Distribution on Y

p_Y	1/4	1/3	7/24	1/8

Normalize columns.

Posterior distributions

	$p_{X\|y1}$	$p_{X\|y2}$	$p_{X\|y3}$	$p_{X\|y4}$
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

Drop output labels.

Hyper-distribution on X

$[\pi, C]$	1/4	1/3	7/24	1/8
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

Example

Prior

π
1/4
1/2
1/4

Channel matrix

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

Scale rows with π.

Joint matrix

J	y_1	y_2	y_3	y_4
x_1	1/8	1/8	0	0
x_2	0	1/8	1/4	1/8
x_3	1/8	1/12	1/24	0

Abstractly, a channel is a mapping from priors to hyper-distributions [ICALP 10].

Add up columns.

Distribution on Y

p_Y	1/4	1/3	7/24	1/8

Normalize columns.

Hyper-distribution on X

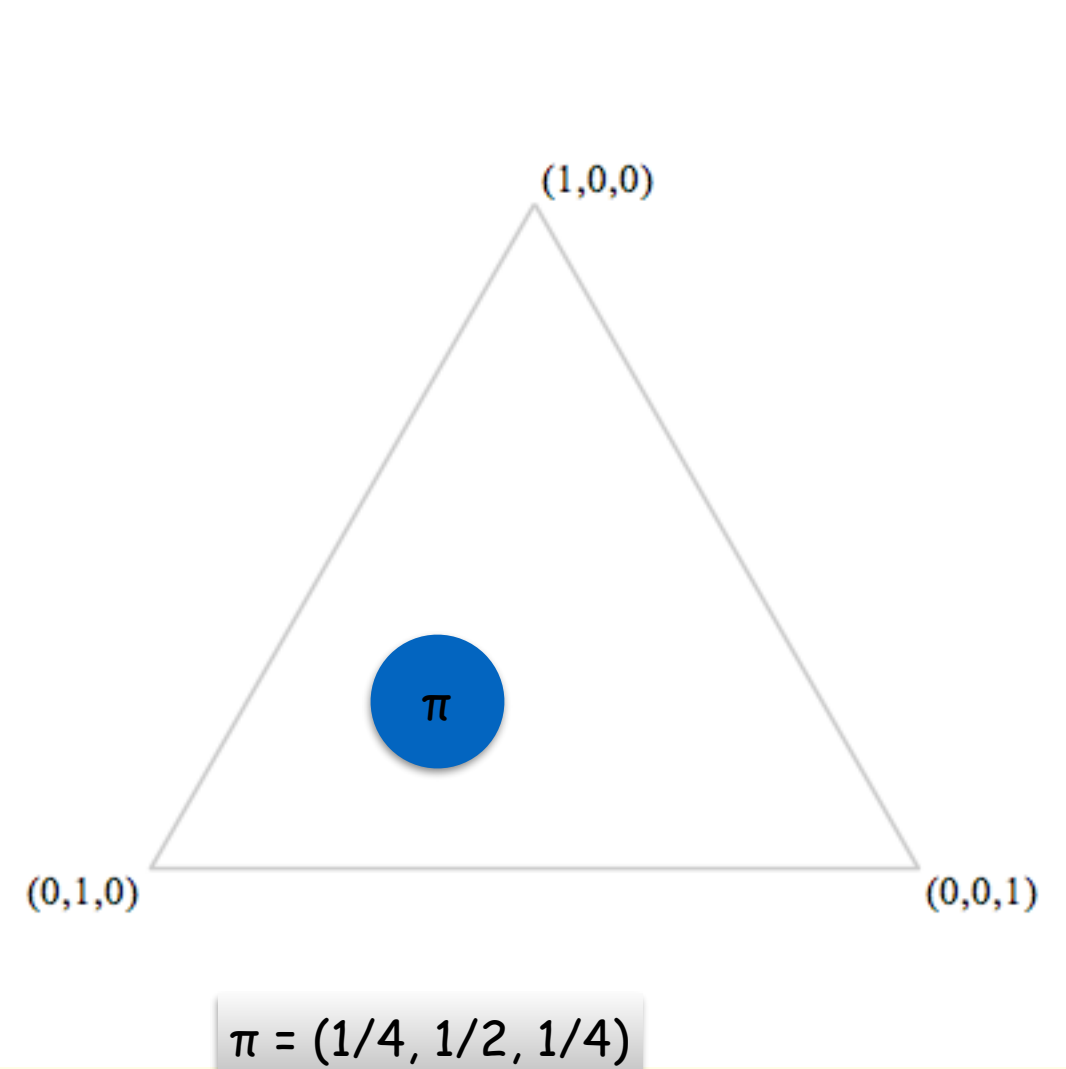
$[\pi,C]$	1/4	1/3	7/24	1/8
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

Drop output labels.

Posterior distributions

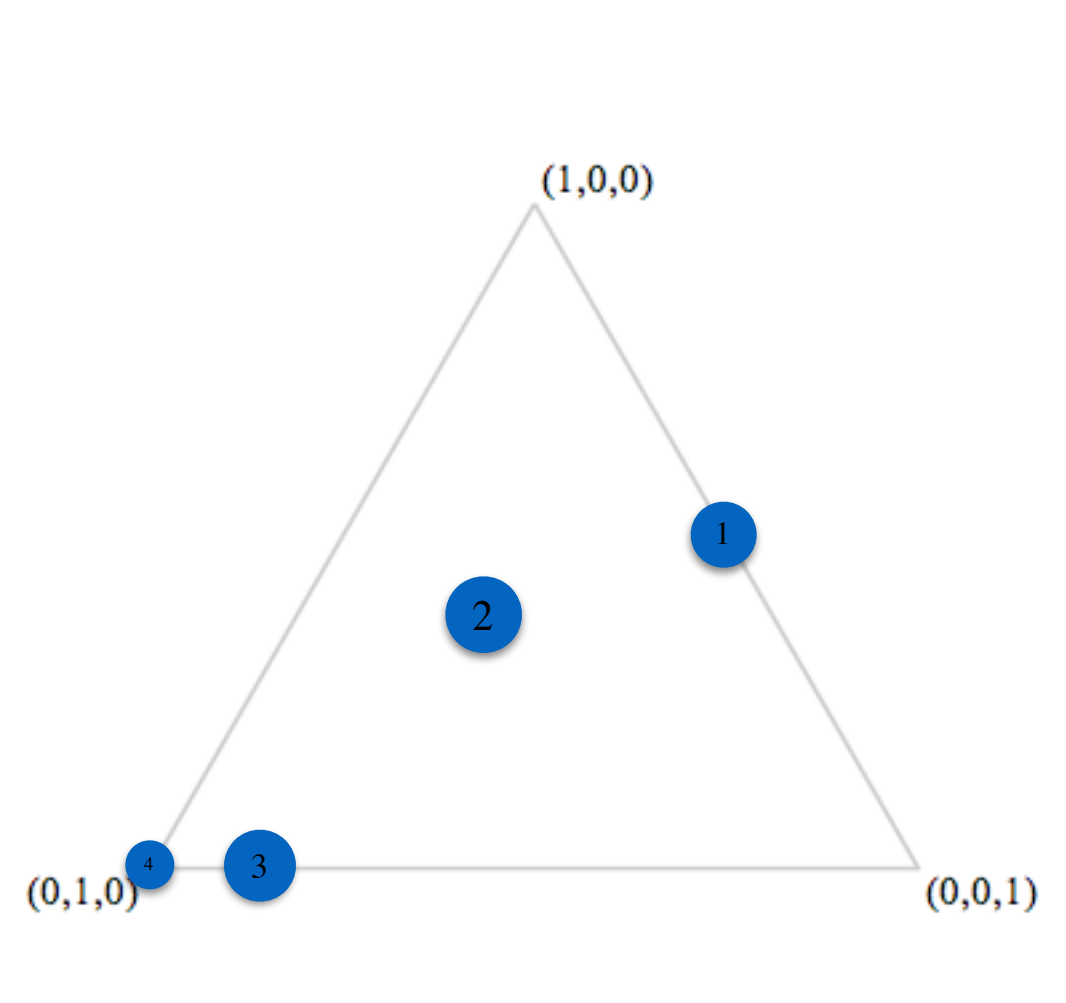
	$p_{X\|y1}$	$p_{X\|y2}$	$p_{X\|y3}$	$p_{X\|y4}$
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

Graphical representation of example



π = (1/4, 1/2, 1/4)

Graphical representation of example



[π,C] is a distribution on 4 posterior distributions

Posterior vulnerability and leakage

- **Definition**: Posterior g-vulnerability
  $V_g[\pi,C] = \sum_y p(y)\, V_g[p_{X|y}]$
  - the average g-vulnerability in the hyper-distribution

- Leakage is naturally defined by comparing the prior and posterior vulnerabilities.
- Multiplicative g-leakage: $\mathcal{L}_g(\pi,C) = \lg\,(V_g[\pi,C] \,/\, V_g[\pi])$
  - Multiplicative leakage using ordinary vulnerability V (equivalently, $V_{g_{id}}$) is called min-entropy leakage.
- Additive g-leakage: $\mathcal{L}^+_g(\pi,C) = V_g[\pi,C] - V_g[\pi]$

Plan of the talk

- Motivation
- Technical background
  - Vulnerability, gain functions, channels, hyper-distributions, multiplicative and additive g-leakage
- This paper
  - Significance of additive leakage
  - Six channel capacity scenarios
  - "Dalenius" leakage

Why additive leakage?

- Operational significance
  - A channel that increases g-vulnerability from 2^{-1000} to 2^{-700} has a huge multiplicative increase (2^{300}) but a negligible additive increase (less than 2^{-700}).
  - With g-vulnerability modeling the economic value to the adversary of the prior and posterior situations, additive leakage gives the expected monetary gain.
- Expressiveness of additive g-leakage
  - It can express guessing entropy leakage.
  - It can express Shannon leakage (mutual information) if we extend the allowable gain functions.

Robustness worries

- Using g-leakage, we can express precisely a rich variety of operational scenarios.
- But we could worry about the **robustness** of our conclusions about leakage.
- The g-leakage $\mathcal{L}_g(\pi,C)$ depends on both π and g.
  - π models adversary $\mathcal{A}$'s prior knowledge about X
  - g models (among other things) what is valuable to $\mathcal{A}$.
- How confident can we be about these?
- Can we minimize sensitivity to questionable assumptions about π and g?

Capacity

- Our paper's approach to robustness is to study capacity, the **maximum** leakage of a channel C over **all** priors π and/or gain functions g.
- This gives **worst-case** bounds on leakage.
- There are six capacity scenarios:
  - multiplicative or additive leakage
  - maximize over π, over g, or over **both** π and g.
- We also consider "Dalenius" scenarios.

Results on computing capacity of C

	$\forall\pi$, fixed g	fixed π, $\forall g$	$\forall\pi$, $\forall g$
Multiplicative capacity	?	**Linear time**	Linear time, by Miracle Theorem [CSF 12]
Additive capacity	**NP-complete**	**Linear time**	?

Additive capacity, $\forall\pi$, fixed g

- **Theorem**: With respect to g_{id}, the additive capacity of C over all priors π is NP-complete.
  - Notice that the input here is the **channel matrix** C, rather than a concise program.
  - Proof is by reduction from the Set Packing Problem ("Does a collection of sets contain k pairwise disjoint sets?").

- For general g, we can find a π that maximizes leakage by solving exponentially many linear-programming problems.

Additive capacity, fixed π, $\forall g$

- **Theorem**: If we fix π and universally quantify over g (ranging over "1-spanning" gain functions), then the additive capacity of C is the Kantorovich distance between the prior $[\pi]$ and hyper-distribution $[\pi,C]$.
- So, by the Kantorovich-Rubinstein Theorem, it is also the earth-moving distance between $[\pi]$ and $[\pi,C]$, which can be computed in time linear in $|C|$.

Example of earth-moving distance

π
1/4
1/2
1/4

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

$[\pi, C]$	1/4	1/3	7/24	1/8
x_1	1/2	3/8	0	0
x_2	0	3/8	6/7	1
x_3	1/2	1/4	1/7	0

- If we want the additive capacity over all g, we take the average earth-moving distance between π and each of the posterior distributions.
  - E.g. the distance between π and (1/2, 0, 1/2) is 1/2.
- Overall we get
  $1/4 \cdot 1/2 + 1/3 \cdot 1/8 + 7/24 \cdot 5/14 + 1/8 \cdot 1/2 = 1/3$

"Dalenius" leakage

- Consider a secret X with prior π, and a channel C from Y to Z, apparently having **nothing** to do with X.
- But if X and Y are correlated, having joint distribution Π, then C can cause leakage of X.
- This "Dalenius" channel can be described as a cascade of a channel B with C.
- The paper proves that the capacity of a cascade BC is upper bounded by the capacity of C.
- Hence the "Dalenius" leakage of X caused by C is upper bounded by the capacity of C, regardless of the correlations that may exist between X and Y.

Thanks!