Comparing systems: max-case refinement orders and application to differential privacy

Kostas Chatzikokolakis, Natasha Fernandes and Catuscia Palamidessi

University of Athens

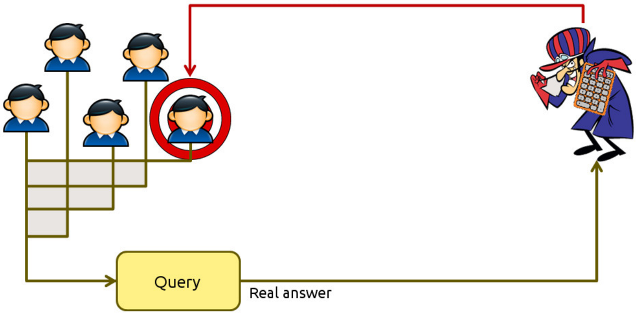
Computer Security Foundations Symposium (CSF)
June 28th, 2019

Topic of the talk

How can we compare systems that unavoidably leak some information?

I. Leakage that happens intentionally

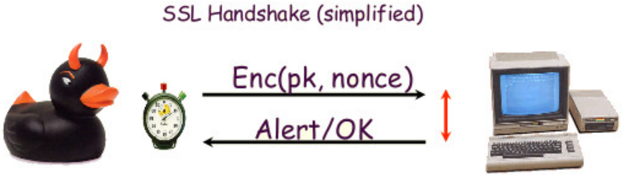
- eg: extract statistics from a dataset



- Problem: inference of personal information
- eg: “what is the median age of cancer patients”

II. Leakage due to side channels

- ge: OpenSSL timing attack [BonehBrumley03]



- Also: cache misses, power, radiation, faults, …
- Completely preventing such channels is costly/impossible

III. Leakage in exchange to a service

- eg: Location Based Services
  - Retrieval of Points Of Interest (POI)
  - Dating
  - Finding friends / social networks
  - …

Channels

Simple probabilistic model of the behavior of a system

- Input : secret event
- Output : observable event

Channels

Simple probabilistic model of the behavior of a system

- Input : secret event
- Output : observable event
- Channel matrix: C_{xy} is the probability that x produces y

$$
\begin{array}{cc}
 & \begin{array}{ccc} y_1 & \cdots & y_n \end{array} \\
\begin{array}{c} x_1 \\ \vdots \\ x_m \end{array} & \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ & \ddots & \\ C_{m1} & \cdots & C_{mn} \end{bmatrix}
\end{array}
$$

Fundamental question

How can we quantify information leakage in such systems?

Quantitative Information Flow (QIF)

Study of different leakage measures, quantifying the adversary's success in achieving some goal.

 $(A) \;=\;$ probability to fully guess the secret $\;=\; 0.2$

 $(A) \;=\;$ exp. error of optimal location infer. $\;=\; 400\text{m}$

Another fundamental question

When can we say that a system B is safer than A? $(A \sqsubseteq B)$

- Can we safely replace A by B?
- Needs to be robust wrt different adversaries!

 $(\; A \;) \quad \geq$  $(\; B \;)$

 $(\; A \;) \quad \leq$  $(\; B \;)$

Another fundamental question

When can we say that a system B is safer than A? $(A \sqsubseteq B)$

- Can we safely replace A by B?
- Needs to be robust wrt different adversaries!
- Needs to be robust wrt different contexts!

 $(\ A\ )\ \geq$  $(\ B\ )$

 $(R[A])\ \leq$  $(R[B])$

Example : Differential Privacy

- $\varepsilon \cdot \mathbf{d}(x, x')$: now much do we want to distinguish x and x'?
  - **d** : "kind" of privacy, ε : "amount" of privacy
- **d**-privacy

$$C \text{ satisfies } \varepsilon\cdot\mathbf{d}\text{-privacy} \quad \text{iff} \quad \frac{C_{x\,y}}{C_{x'\,y}} \le e^{\varepsilon\cdot\mathbf{d}(x,x')} \quad \forall x, x', y$$

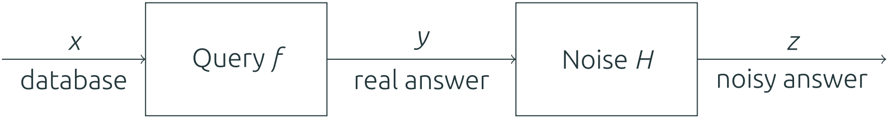
- Differential privacy
  - Hamming $\mathbf{d}_{\mathrm{H}}(x, x')$: # of users with different value in dbs x, x'

Example : Differential Privacy

- Oblivious mechanism $H \circ f$
  - Compute f then apply noise mechanism H to the real answer
  - $\varepsilon \cdot \mathbf{d}_E$-privacy can be proven for H alone

- A variety of noise mechanisms, eg
  - RR^{ε} : randomized response
  - TG^{ε} : geometric (truncated)

Both satisfy $\varepsilon \cdot \mathbf{d}_E$-privacy (same ε)
Are they equivalent ?



Example : Differential Privacy

Are TG^ε and RR^ε equivalent ?

- f : minimum age of people in the database
  - $RR^\varepsilon \circ f$ is ε-diff. private
  - $TG^\varepsilon \circ f$ is not ε-diff. private
  - We cannot replace RR^ε by TG^ε in this context!
- In the other direction
  - We can prove that $TG^\varepsilon \sqsubseteq RR^\varepsilon$ (for a suitable $\sqsubseteq$)
  - For any query f :
    - if $TG^\varepsilon \circ f$ is ε-diff. private
    - then $RR^\varepsilon \circ f$ is also ε-diff. private
  - RR^ε is safer than TG^ε

Example : Differential Privacy

In the context of local differential privacy

- Noise applied to the data
- We can construct mechanisms A and B such that
  - A is $\log 3$-LDP
  - B is $\log 2$-LDP so B looks safer
- But B is not safer for all adversaries
  -  fully guess the secret x  $(A) \geq$  (B)
  -  guess whether $x = x_0$ or not  $(A) \leq$  (B)

How we can apply QIF to this problem?

QIF : Vulnerability

- Prior π on the secrets
  - probabilistic knowledge of the adversary
- Vulnerability $V(\pi)$
  - how happy the adversary is to have π
  - eg. Bayes vulnerability : prob. of correctly guessing the secret
- Axiomatic view
  - V can be any continuous convex function
  - All of them expressible in the g-leakage framework

QIF : Posterior vulnerability

With probability $p(y)$ the vulnerability of the system becomes $V(\delta^y)$

$$\xrightarrow[\text{prior}]{\pi} \boxed{\text{Channel}} \xrightarrow[\text{posterior}]{\delta^y,\ \text{with pb } p(y)}$$

Average-case

 $(C) \quad = \quad V[\pi, C] \quad = \quad \sum_y p(y)\, V(\delta^y)$

Max-case

 $(C) \quad = \quad V^{\max}[\pi, C] \quad = \quad \max_{p(y)>0} V(\delta^y)$

QIF : Comparing channels

- Leakage order

$$A \sqsubseteq_{\mathbb{G}}^{\mathsf{avg}} B \qquad \text{iff} \qquad V[\pi, A] \geq V[\pi, B] \quad \forall \pi, V$$

Intuitive but hard to verify

- Refinement order

$$A \sqsubseteq^{\mathsf{avg}} B \qquad \text{iff} \qquad AR = B \quad \text{for some } R$$

Structural property of the channels

Theorem [CSF'12, POST'14]

$$\sqsubseteq^{\mathsf{avg}} \quad \Leftrightarrow \quad \sqsubseteq_{\mathbb{G}}^{\mathsf{avg}}$$

Is refinement enough?

- Refinement is robust
  - $A \sqsubseteq^{\text{avg}} B \quad \Rightarrow \quad$ no adversary prefers B
  - $A \not\sqsubseteq^{\text{avg}} B \quad \Rightarrow \quad$ at least one adversary V prefers B
    - And we can compute V!
- But what if we care about the max-case V^{max}?
  - $A \sqsubseteq^{\text{avg}} B \quad \Rightarrow \quad ?$
  - $A \not\sqsubseteq^{\text{avg}} B \quad \Rightarrow \quad ?$
- What if we care only about differential privacy
  - A max-case measure!

This work answers these questions (and some more)

Max-case refinement

- We can easily define a max-case leakage order

$$A \sqsubseteq^{\max}_{\mathbb{Q}} B \quad \text{iff} \quad V^{\max}[\pi, A] \geq V^{\max}[\pi, B] \quad \forall \pi, V$$

Again, intuitive but hard to verify

- Max-case refinement order

$$A \sqsubseteq^{\max} B \quad \text{iff} \quad R\tilde{A} = \tilde{B} \quad \text{for some } R$$

Again, structural property of the channels

Theorem

$$\sqsubseteq^{\max} \quad \Leftrightarrow \quad \sqsubseteq^{\max}_{\mathbb{Q}}$$

Max-case refinement

- Max-case refinement is robust
  - $A \sqsubseteq^{\max} B \quad \Rightarrow \quad$ no max-case adversary prefers B
  - $A \not\sqsubseteq^{\max} B \quad \Rightarrow \quad$ at least one max-case adversary V prefers B
    - And we know such a V
- We can also show: $\sqsubseteq^{\text{avg}} \quad \Rightarrow \quad \sqsubseteq^{\max}$ (strictly)
  - So $\sqsubseteq^{\text{avg}}$ also provides max-case guarantees!
  - But it might be too strong
- What about differential privacy?
  - $A \sqsubseteq^{\max} B \quad \Rightarrow \quad ?$
  - $A \not\sqsubseteq^{\max} B \quad \Rightarrow \quad ?$

Differential privacy vs QIF

- DP is a max-case notion
  - Treats every y equally, independently from its probability
  - Can we express it as a QIF measure?

Theorem

C satisfies $\varepsilon{\cdot}\mathbf{d}$-privacy iff $V_{\mathbf{d}}^{\max}[\pi^u, C] \le \varepsilon$
for a suitably constructed $V_{\mathbf{d}}$.

- So $\sqsubseteq^{\max}$ imposes a DP order
  - But is it too strong?

Privacy-based refinement

- We can also easily define a privacy-based order

$$A \sqsubseteq^{\text{prv}}_{\mathbb{M}} B \quad \text{iff} \quad A \text{ sat. } \mathbf{d}\text{-privacy} \Rightarrow B \text{ sat. } \mathbf{d}\text{-privacy} \quad \forall \mathbf{d}$$

Again, intuitive but hard to verify

- Privacy-case refinement order

$$A \sqsubseteq^{\text{prv}} B \quad \text{iff} \quad \mathbf{d}_A \geq \mathbf{d}_B$$

Again, structural property of the channels

Theorem

$$\sqsubseteq^{\text{prv}} \quad \Leftrightarrow \quad \sqsubseteq^{\text{prv}}_{\mathbb{M}}$$

Privacy-based refinement

- Privacy-case refinement is robust
  - $A \sqsubseteq^{\text{prv}} B \quad \Rightarrow \quad$ no DP adversary prefers B
  - $A \not\sqsubseteq^{\text{prv}} B \quad \Rightarrow \quad$ at least one DP adversary **d** prefers B
    - And we know such a **d**
- We can also show: $\sqsubseteq^{\text{max}} \quad \Rightarrow \quad \sqsubseteq^{\text{prv}}$ (strictly)
  - So $\sqsubseteq^{\text{avg}}, \sqsubseteq^{\text{max}}$ also provide privacy guarantees!
  - But they might be too strong

Privacy-based refinement

What about query composition?

Theorem

$$A \quad \sqsubseteq^{\mathrm{prv}} \quad B \qquad \Leftrightarrow \qquad A \circ f \quad \sqsubseteq^{\mathrm{prv}} \quad B \circ f \quad \text{for all queries } f$$

Not true if we compare A, B on a single $\mathbf{d}$

Comparison of leakage/refinement orders

Leakage orders		Refinement orders
$\sqsubseteq_{\mathbb{G}}^{\mathsf{avg}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathsf{avg}}$
$\Downarrow$		$\Downarrow$
$\sqsubseteq_{\mathbb{Q}}^{\mathsf{max}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathsf{max}}$
$\Downarrow$		$\Downarrow$
$\sqsubseteq_{\mathbb{M}}^{\mathsf{prv}}$	$\Leftrightarrow$	$\sqsubseteq^{\mathsf{prv}}$
$\searrow$		$\swarrow$
	$\sqsubseteq_{\mathbf{d}}^{\mathsf{prv}}$	

All implications are strict

Application: comparing DP mechanisms

Same family, different ε

$$C^{\varepsilon} \quad \sqsubseteq^{\text{avg}} \quad C^{\varepsilon'} \qquad \text{iff} \qquad \varepsilon \geq \varepsilon' \qquad \text{for } C \in \{G, TG, RR, E\}$$

- Decreasing ε is safe in a very strong sense
- But surprisingly, for the "overly truncated" geometric:
  - $OTG^{\varepsilon} \quad \not\sqsubseteq^{\text{avg}} \quad OTG^{\varepsilon'}$
  - $OTG^{\varepsilon} \quad \not\sqsubseteq^{\text{max}} \quad OTG^{\varepsilon'}$
  - $OTG^{\varepsilon} \quad \sqsubseteq^{\text{prv}} \quad OTG^{\varepsilon'}$ still holds!

Application: comparing DP mechanisms

Different families, same ε

$\mathrm{TG} \not\sqsubseteq^{\mathrm{avg}} \mathrm{RR}$ $\mathrm{RR} \not\sqsubseteq^{\mathrm{avg}} \mathrm{TG}$	$\mathrm{TG} \not\sqsubseteq^{\max} \mathrm{RR}$ $\mathrm{RR} \not\sqsubseteq^{\max} \mathrm{TG}$	$\mathrm{TG} \sqsubseteq^{\mathrm{prv}} \mathrm{RR}$ $\mathrm{RR} \not\sqsubseteq^{\mathrm{prv}} \mathrm{TG}$
$\mathrm{TG} \not\sqsubseteq^{\mathrm{avg}} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\mathrm{avg}} \mathrm{TG}$	$\mathrm{TG} \not\sqsubseteq^{\max} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\max} \mathrm{TG}$	$\mathrm{TG} \sqsubseteq^{\mathrm{prv}} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\mathrm{prv}} \mathrm{TG}$
$\mathrm{RR} \not\sqsubseteq^{\mathrm{avg}} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\mathrm{avg}} \mathrm{RR}$	$\mathrm{RR} \not\sqsubseteq^{\max} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\max} \mathrm{RR}$	$\mathrm{RR} \not\sqsubseteq^{\mathrm{prv}} \mathrm{E}$ $\mathrm{E} \not\sqsubseteq^{\mathrm{prv}} \mathrm{RR}$

Other results

Verification

- $\sqsubseteq^{\text{avg}}, \sqsubseteq^{\text{max}}, \sqsubseteq^{\text{prv}}$ can be verified in time polynomial in the size of C
- We obtain counterexamples when they fail

Lattice properties

- It is known that $\sqsubseteq^{\text{avg}}$ is not a lattice
- But $\sqsubseteq^{\text{max}}$ is !
  - $A \vee^{\text{max}} B$: intersection of the convex-hull of posteriors
- So is $\sqsubseteq^{\text{prv}}$
  - $A \vee^{\text{prv}} B$: sup in the lattice of metrics

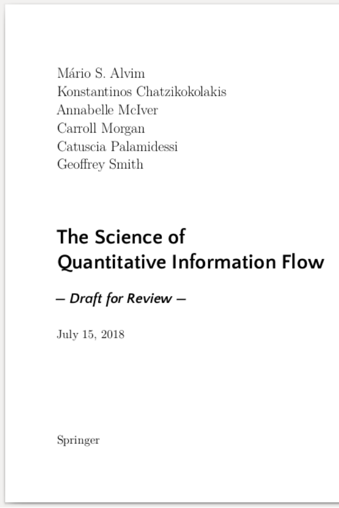
Shameful advertisement

We have a QIF book!

- Ask me for a draft

Postdoc / Research Assistant positions

- HYPATIA
  - Statistical utility from noisy data
  - Optimal privacy-utility trade-off
  - Generation of optimal mechanism via ML
- DATAiA
  - Analysis of privacy threats in ML



Mário S. Alvim
Konstantinos Chatzikokolakis
Annabelle McIver
Carroll Morgan
Catuscia Palamidessi
Geoffrey Smith

The Science of Quantitative Information Flow

– Draft for Review –

July 15, 2018

Springer

Conclusion

- QIF provides rich, robust tools for comparing leaky systems
- Leakage-based (intruitive) and structural (verifiable) characterizations
- DP: (mostly) safe to decrease ε within a family, but not to change family

Future directions

- Comparison with other channel orders
- Study the behavior under different contexts
- Conditions for refinement in different models
- Use refinement to verify complex programs

Questions?