QASA 2012
(Pisa, Italy)

From Qualitative to Quantitative Information Erasure

Adedayo O. Adetoye Michael H. Goldsmith

Cyber Security Centre,
Department of Computer Science,
University of Oxford,
United Kingdom

The Need for Information Erasure in the Context of Information Release

We want to process sensitive information but not necessarily propagate (some parts of) the information.

- **Statistical Databases**
  - May release sufficient information to be useful for statistical purposes, but must erase sufficient information not to violate privacy
- **E-commerce**
  - There are regulations on what information can be displayed by a merchant on receipts and screens, and what must be masked (*erased*)
- **E-voting**
  - We want to release result of election but not individual votes
- ...

Outline

1. Intuition about Information Erasure Modelling
2. Quantifying Information Erasure
3. Weaknesses and Advantages of Quantification
4. What Qualitative model does not capture
5. Conclusion

E-commerce

- PCI stipulates that payment processing systems may display, on receipts and screens, at most first six and last four digits of CC *Primary Account Number* – other digits must be masked (*erased*)
- Desired erasure policy is $\mathit{all} \leftarrow R$, where $\forall c, c' \in CC.(c, c') \in R \iff c[1:6] = c'[1:6] \wedge c[13:16] = c'[13:16]$.

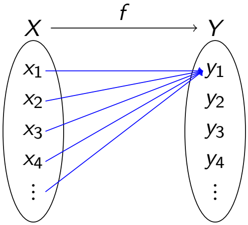
System and Attack Model

- We consider deterministic systems
  - Modelled as functions: $f : X \to Y$
  - System's input domain (contains secrets): X
  - System's output domain (the public observables): Y
- Attacker can observe Y but not necessarily X
- Attacker knows the system model f

- **How much of X is erased in Y?**

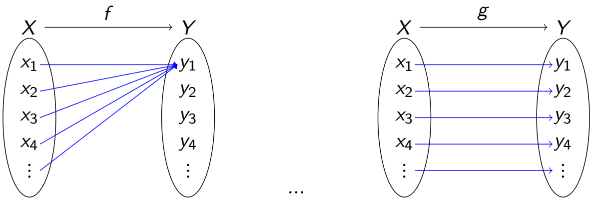
Information Erasure: An Extreme

System modelled by $f : X \to Y$ erases all information (is noninterferring) if $\forall x_1, x_2, f(x_1) = f(x_2)$



- The *observation* at Y is *independent* of the choice of input $x \in X$

Intuition about Information Erasure: The Extremes



(a) Total Information Erasure (b) No Erasure

Figure: *Extreme cases of Information Erasure*

Intuition about Information Erasure: The Extremes

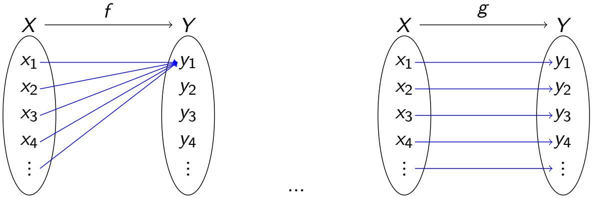


Figure: *Extreme cases of Information Erasure*

- The level of erasure can be characterised by kernels of $f, g, \cdots$ (or ERs)
  - (a) $\forall x, x' \in X, x\; all\; x'$ (total erasure of information in X)
  - (b) $\forall x, x' \in X, x\; id\; x' \iff x = x'$ (Input can be precisely determined from g and Y)
- Various other intermediate levels of information exist

Example: Knowledge of Colour

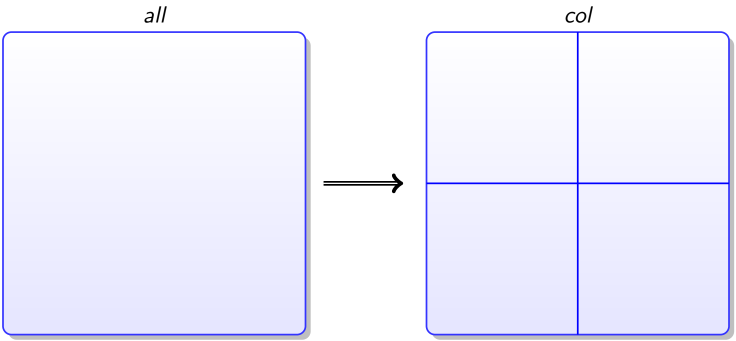


Figure: *Information about colour*

$$\forall b, b' \in Balls. \quad b\ col\ b' \iff b.colour = b'.colour$$
$$\forall b, b' \in Balls. \quad b\ all\ b'$$

The ER model of information

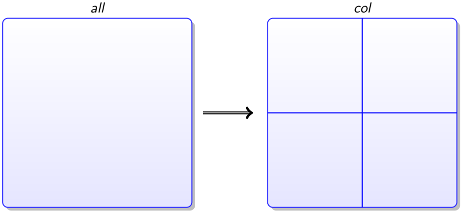


Figure: *Information about colour*

$all, col \in ER(Balls)$ are ERs over the set of Balls.

- A ER ($R \in ER(X)$) represents information by its ability to distinguish, or not, a pair of elements ($x, x' \in X$):
  - $(x, x') \in R$ means indistinguishability of pair: lack of knowledge
  - $(x, x') \notin R$ means distinguishability of pair: knowledge

Erasure Policies

Suppose $R, R' \in ER(X)$

1. **Release Policy**: $R \twoheadrightarrow R'$
   - Given initial knowledge R, agent may not learn more than $R \sqcup R'$
   - Release because $R \sqsubseteq R \sqcup R'$
2. **Erasure Policy**: $R \twoheadleftarrow R'$
   - Given some reference information R', then $R \sqcap R'$ is the maximum allowed to be propagated
   - Or, if a system conforms to $R \twoheadleftarrow R'$ then it ensures that no more than $R \sqcap R'$ may be learnt from its output.
   - Erasure because $R \sqcap R' \sqsubseteq R'$

Erasure Policy Satisfaction

- A system modelled by f satisfies the erasure policy $R \twoheadleftarrow R'$, written $f \vDash R \twoheadleftarrow R'$, if $\kappa_f \sqsubseteq R \sqcap R'$.
- Similarly, a system modelled by f satisfies the release policy $R \twoheadrightarrow R'$, written $f \vDash R \twoheadrightarrow R'$, if $\kappa_f \sqsubseteq R \sqcup R'$.

Quantifying Erasure

Suppose $R \in ER(X)$ and μ is a probability measure over X.

- The information content of X subject to its partitioning by R is $\mathcal{H}(\mu|R) \triangleq -\sum_{X' \in [X]_R} \mu(X') \log \mu(X')$
- A generalisation of the standard Shannon's entropy definition $\mathcal{H}(\mu) = \mathcal{H}(\mu|id_X) = -\sum_{x \in X} \mu(x) \log \mu(x)$

Erasure and Release Quantification

$$\begin{aligned} \mathcal{H}(\mu|R \leftarrow R') &\triangleq \mathcal{H}(\mu|R') - \mathcal{H}(\mu|R \sqcap R') \\ \mathcal{H}(\mu|R \rightarrow R') &\triangleq \mathcal{H}(\mu|R' \sqcup R) - \mathcal{H}(\mu|R) \end{aligned} \quad (1)$$

Illustration (Uniform μ)

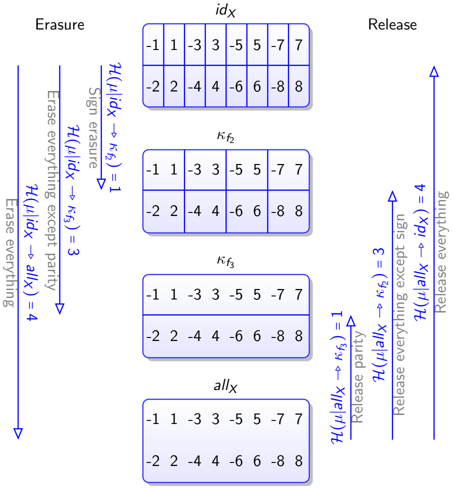


Figure: *Information Erasure and Release Policies*

Consider four systems modelled by the following functions:

1. $f_1(x) = x$
2. $f_2(x) = |x|$
3. $f_3(x) = x \mod 2$
4. $f_4(x) = 0$

Properties of Quantitative Erasure

Duality of Erasure and Release

Theorem

For any chain of equivalence relations $R_1, R_2, R_3 \in ER(X)$ *such that* $R_1 \sqsubseteq R_2 \sqsubseteq R_3$ *we have that*
$\mathcal{H}(\mu|R_1 \twoheadrightarrow R_2) + \mathcal{H}(\mu|R_2 \twoheadleftarrow R_3) = \mathcal{H}(\mu|R_1 \twoheadrightarrow R_3) = \mathcal{H}(\mu|R_1 \twoheadleftarrow R_3)$.

Corollary

For any set X*, equivalence relation* R *over* X*, and probability measure* μ *over* X*, we have that* $\mathcal{H}(\mu|all_X \twoheadrightarrow R) + \mathcal{H}(\mu|R \twoheadleftarrow id_X) = \mathcal{H}(\mu)$.

Properties of Quantitative Erasure (Contd.)

Agrees with existing definitions

Theorem ($\mathcal{H}(\mu|all_X \twoheadrightarrow \kappa_f)$ equals mutual information)

Let κ_f be the kernel of the function $f : X \to Y$, then
$\mathcal{H}(\mu|all_X \twoheadrightarrow \kappa_f) = \mathcal{I}(X; Y)$.
Furthermore, $\mathcal{H}(\mu|\kappa_f \twoheadleftarrow id_X) = \mathcal{I}(X) - \mathcal{I}(X; Y)$.

Lemma (Erasure and Release between two comparable levels are identical)

Let $R, R' \in ER(X)$ such that $R \sqsubseteq R'$, and let μ be a probability measure over X. Then, $\mathcal{H}(\mu|R \twoheadrightarrow R') = \mathcal{H}(\mu|R \twoheadleftarrow R')$.

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Caveat!

- Analysis requires μ, what of if we don't know it?
- Even with μ, what does the measure mean?

Release/Erasure Occlusion

Suppose $\mu(n) = \frac{1}{4}$ for all n.

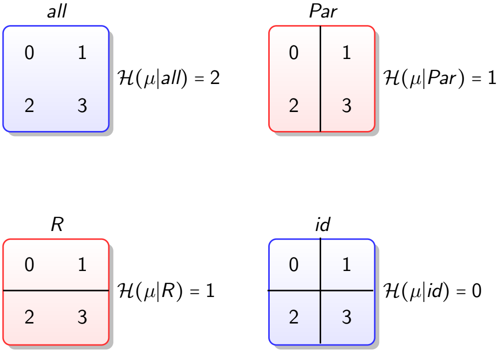


Figure: *Probability Permutation Problem of Quantitative Policies*

Occlusion due to μ effectively restricting the function domain

Suppose $\mu(-2) = \mu(2) = \mu(-1) = \mu(1) = \frac{1}{4}$ and $\mu(-4) = \mu(4) = \mu(-3) = \mu(3) = 0$

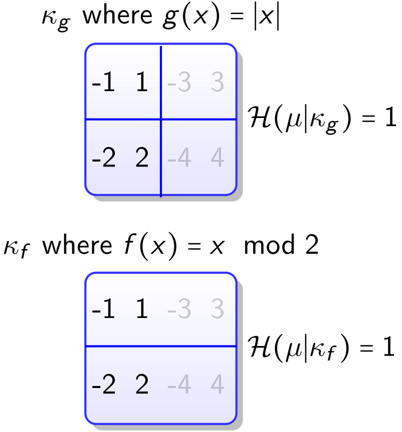


Figure: *Information Erasure and Release Policies*

Conclusion

- Information erasure is important in practice
- We can model what information is erased in systems
- Care should be taken with the interpretation of quantitative measures: what impact does *prob* (or our assumption about it) have on risk to information vis-a-vis the quantitative measure?
- We may be able to constrain, via policies on μ, the probabilistic behaviour of systems and their environments as a statement of required system security to guarantee desired assurance
- Many more interesting open issues: **Hybrid Qualitative + Quantitative Policies**, Reasoning about erasure of components of structured inputs, **nondeterminism**, system composition and structuring ...

Last Slide!

Thank You!

Questions?