Revisiting a Combinatorial Approach Toward Measuring Anonymity

B. Gierlichs, C. Troncoso, C. Diaz, B. Preneel, I. Verbauwhede
K.U. Leuven, Esat - COSIC, Belgium

Anonymity communication systems

- Anonymous communication systems aim at hiding relations between communication partners
- Many designs, typically built with mixes or onion routers
- Adversary’s goal is to discover relations between users

Example, Limitations, Counterexample

- More possible perfect matchings
- If only one perfect matching is possible, zero anonymity
- More possible perfect matchings, more anonymity
- Metric \( d(A) \) counts the number of perfect matchings on \( G \) (equivalent to the permanent \( \text{perm}(A) \) of the adjacency matrix \( A \)) and normalizes to [0,1]

System’s anonymity level [Edman et al.]

- Measures the amount of information required to reveal the full set of relations between the inputs and outputs of a mix
- Can be modeled as a bipartite graph \( G = ([I,O], E) \)
- Graph can be represented by its adjacency matrix, here \( a_{ij} \in \{0,1\} \):

\[
A = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
1 & 1 & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & 1
\end{bmatrix}
\]

Metrics for anonymous communications

- Need for metrics to evaluate and compare different designs
- Numerous information-theoretic metrics:
  - Measure the adversary’s uncertainty about the sender/receiver of a single given message (entropy, rel. entropy, Rény entropy, etc.)
- A combinatorial approach [Edman et al.]
  - Don’t analyze the anonymity of a single given message but consider all inputs and outputs simultaneously
  - Metric gives a good picture of the anonymity provided by the system as a whole
  - But it is not able to express the anonymity of a single given message
  - Conclusion: use both
Generalizing the system's anonymity level

- Senders and receivers form multisets

- Let $\theta$ denote the number of equivalence classes and let $C_p$ denote the number of perfect matchings in class $[M_p]$

- $3! = 6$ perfect matchings, but only 2 classes: $[M_1] = \{AD,AD,CF\}$ with $C_1 = 2$ and $[M_2] = \{AD,AF,CD\}$ with $C_2 = 4$

- Let $M_c$ be the correct perfect matching; we have
  \[ \text{Prob}(M_c \in [M_1]) = \frac{2}{6} \text{ and } \text{Prob}(M_c \in [M_2]) = \frac{4}{6} \]

- The amount of additional information required to identify the equivalence class that contains $M_c$ is given by the Shannon entropy of the RV with probability distribution $\text{Pr}(M_c \in [M_p])$

Computing the revised metric $d^*(A)$

- Metric $d^*(A)$ computes this entropy and normalizes to $[0;1]$

- We need to obtain $\theta$ and $C_p$

- A naive way is exhaustive search: generate all perfect matchings and classify them into equivalence classes

- This requires $\theta(t!)$ operations and quickly becomes infeasible

- In the paper we present 2 alternatives
  - A divide-and-conquer algorithm to compute the exact metric
  - An easy way to compute upper and lower bounds if the graph associated to the system is complete, i.e. the system is a threshold mix

Conclusions

- We revisited Edman et al.'s combinatorial approach towards measuring anonymity

- We argue that a metric should focus on the relationships between users rather than inputs and outputs

- We show how the System's anonymity level as defined by Edman et al. focuses on inputs and outputs and thus cannot reflect the reduction of anonymity due to multiplicities

- We generalize the metric in scenarios where user relations can be modeled by yes/no

- We propose an algorithm to compute the metric and show how to easily obtain bounds if the system is a threshold mix

Thanks for your attention!

benedikt.gierlichs@esat.kuleuven.be