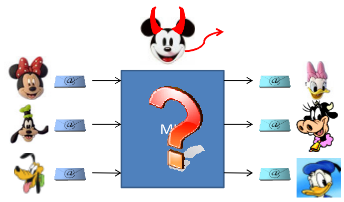
Revisiting a Combinatorial Approach Toward Measuring Anonymity

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

Anonymous 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

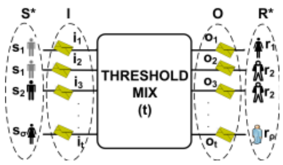


Metrics for anonymous communications

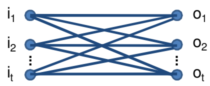
- Need for metrics to evaluate and compare different designs
- Numerous information-theoretic metrics:
  - Meausure the adversary's uncertainty about the sender/receiver of a single given message (entropy, rel. entropy, Rény entropy, etc.)
- A combinatorial approach [Edman et al.]
  - Don't analyze the 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

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)

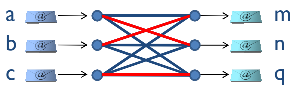


- Graph can be represented by its adjacency matrix, here $a_{ij} \in \{0,1\}$:

$$A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

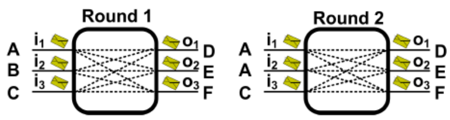
System's anonymity level [Edman et al.]

- Metric relies on the fact that there must be a one-to-one relation between inputs and outputs: a perfect matching on the graph G



  - 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 *per(A)* of the adjacency matrix A) and normalizes to [0;1]

Example, Limitations, Counterexample



- The graphs for both rounds allow 3! = 6 perfect matchings

$$\{i_1o_1, i_2o_2, i_3o_3\},\ \{i_1o_1, i_2o_3, i_3o_2\},\ \{i_1o_2, i_2o_3, i_3o_1\}$$
$$\{i_1o_2, i_2o_1, i_3o_3\},\ \{i_1o_3, i_2o_1, i_3o_2\},\ \{i_1o_3, i_2o_2, i_3o_1\}$$

- But: goal of adversary is to identify relationships between users

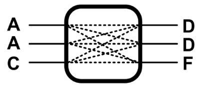
$$\{AD, BE, CF\},\ \{AD, BF, CE\},\ \{AE, BF, CD\}$$
$$\{AE, BD, CF\},\ \{AF, BD, CE\},\ \{AF, BE, CD\}$$
d(A) = 1

Equivalence classes $[M_p]$

$$\{AD, AE, CF\},\ \{AD, AF, CE\},\ \{AE, AF, CD\}$$
$$\{AE, AD, CF\},\ \{AF, AD, CE\},\ \{AF, AE, CD\}$$
$d(A) \overset{?}{=} 1$

Generalizing the system's anonymity level

- Senders and receivers form multisets



- Let θ 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 $Prob(M_c \in [M_1]) = 2/6$ and $Prob(M_c \in [M_2]) = 4/6$
- 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 $Pr(M_c \in [M_p])$

Computing the revised metric *d*(A)*

- Metric *d*(A)* computes this entropy and normalizes to [0;1]
- We need to obtain θ and C_p
- A naïve way is exhaustive search: generate all perfect matchings and classify them into equivalence classes
- This requires $\mathcal{O}(t!)$ 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