Revisiting a Combinatorial Approach Toward Measuring Anonymity

Benedikt Gierlichs Carmela Troncoso Claudia Diaz Bart Preneel
Ingrid Verbauwhede
Katholieke Universiteit Leuven, ESAT/SCD-COSIC and IBBT
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
firstname.lastname@esat.kuleuven.be

ABSTRACT

Recently, Edman et al. proposed the *system's anonymity level* [10], a combinatorial approach to measure the amount of additional information needed to reveal the communication pattern in a mix-based anonymous communication system as a whole. The metric is based on the number of possible bijective mappings between the inputs and the outputs of the mix. In this work we show that Edman et al.'s approach fails to capture the anonymity loss caused by subjects sending or receiving more than one message. We generalize the *system's anonymity level* in scenarios where user relations can be modeled as yes/no relations to cases where subjects send and receive an arbitrary number of messages. Further, we describe an algorithm to compute the redefined metric.

Categories and Subject Descriptors: D.2.8 [Software Engineering]: Metrics: complexity measures, performance measures

General Terms: Algorithms, Measurement, Theory

Keywords: Anonymity metric, Combinatorics, Graph theory, Privacy

1. INTRODUCTION

The goal of anonymous communication systems is to hide the correspondence between communication partners, such that an adversary cannot determine who is sending messages to whom. Anonymous communication systems are usually built with mixes [2, 4, 19] or onion routers [9, 12, 14], black boxes whose objective is to hide the correspondence between input and output messages or streams.

The emergence of anonymous communication systems led to the need for anonymity metrics to evaluate and compare different designs. Based on the definition of anonymity proposed by Pfitzmann and Hansen [13] "Anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*" information-theoretic metrics were independently proposed by Diaz et al. and Serjantov and Danezis in [7, 16]. These metrics are based on Shannon entropy [17], and express the uncertainty of an adversary with respect to the sender or recipient of a given message. Several variations of information-theoretic metrics for anonymity have followed: Tóth et al. [18] propose using min-entropy and max-entropy for measuring local anonymity; Clauß and Schiffner [3] propose to use Rényi entropy [15] as a generalization of Shannon, min- and max-entropy; Deng et al. [5] suggest using relative entropy; and Zhu and Bettati [20] propose an anonymity metric based on mutual information.

A different approach was followed by Edman et al. [10]. Instead of computing the size of the (sender or recipient) anonymity set for a given message, they consider simultaneously all incoming and outgoing messages in an anonymous communication system. Combinatorial approaches have also been used to model unlinkability [11] and in the context of disclosure attacks [1].

We revisit Edman et al.'s *system's anonymity level* and show that it does not capture the anonymity loss caused by subjects sending or receiving multiple messages. We propose a generalization of the metric in scenarios where user relations can be modeled as yes/no relations taking multiple messages per subject into account. We provide a divide and conquer algorithm to compute the redefined *system's anonymity level*. The key difference between our approach and Edman et al.'s is that we consider relationships between senders and recipients, rather than between individual input and output messages.

While our observation also applies to anonymity metrics that measure the size of sender or recipient anonymity sets, we note that it is trivial to adapt these metrics to account for multiple messages per subject. This was not explicitly addressed in some of the first works [6, 7, 16], but later papers [8] do consider multiple messages per subject.

The next section introduces the *system's anonymity level* as defined in [10]. We show in Sect. 3 that this metric does not reflect the reduction of anonymity due to multiple messages per sender and/or receiver. Sections 4 and 5 present the generalization of the metric and an algorithm to compute it. Finally, we offer our conclusions in Sect. 6.

WPES'08, October 27, 2008, Alexandria, Virginia, USA.

2. A COMBINATORIAL APPROACH TO MEASURING ANONYMITY

In [10] Edman et al. present an anonymity metric that measures the amount of information needed to reveal the full set of relationships between the inputs and the outputs of a mix.

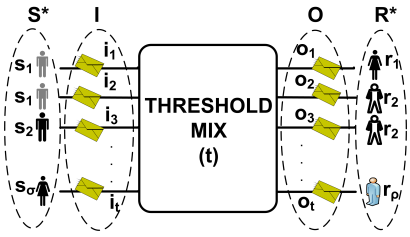

Figure 1: System modeled as a threshold mix

2.1 Notation

Edman et al. consider messages (or streams) as inputs and outputs of an anonymous communication system and model this system as a bipartite graph $G = (I, O, E)$, where $I = \{i_j\}$ is the set of t inputs, $O = \{o_l\}$ is the set of t outputs, and E is the set of edges $\{e_{i_j,o_l}\}$ between inputs and outputs. The graph G can be represented by its adjacency matrix A, where the elements $a_{j,l}$ of the matrix are 1 if the edge linking i_j and o_l exists in G, and 0 if it does not exist.

2.2 The metric

In this setting, Edman et al. exploit the fact that in an anonymity system there exists a one-to-one relation between inputs and outputs (i.e., a perfect matching on the associated bipartite graph G) to evaluate the amount of information a global adversary can infer about the relations between senders and receivers. If only one perfect matching is possible in G, the adversary can uniquely identify the relations between inputs and outputs, thus the anonymity provided by the system is zero. When the number of possible perfect matchings grows, so does the uncertainty of the attacker.

In order to measure the anonymity provided by the system, Edman et al. propose to count the number of possible perfect matchings in G, which is equivalent to computing the permanent $per(A)$ of the adjacency matrix A. They define the *system's anonymity level* as:

$$d(A) = \begin{cases} 0 & \text{if } t = 1 \\ \frac{log(per(A))}{log(t!)} & \text{if } t > 1\,. \end{cases}$$

The permanent of a matrix A whose entries are all 1 (i.e., a fully connected graph) is $per(A) = t!$. They note that in general computing the permanent of a matrix is NP-hard and provide upper and lower bounds.

3. LIMITATIONS

In this section we illustrate with an example how the *system's anonymity level* correctly measures the amount of information required to reveal the whole communication pattern if we consider messages or sets of subjects (i.e., all subjects send or receive exactly one message). We also show, however, that this metric over-estimates the anonymity if we consider multisets of senders and/or receivers (i.e., subjects send and/or receive multiple messages).

In the example we abstract the anonymity system as a round of a threshold mix with threshold t as shown in Fig. 1. We note that the same model can be applied to a router that mixes t streams.

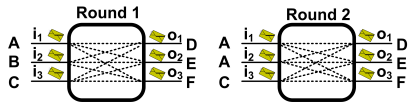

Figure 2: Two rounds of the mix

3.1 Notation

We define $S^* = (S, N)$ as the multiset of the senders of all messages, where the underlying set $S = \{s_1, s_2, ..., s_\sigma\}$ contains the identities of the σ distinct senders, and the vector $N = \langle n_1, n_2, ..., n_\sigma \rangle$ contains in its i^{th} position the multiplicity (number of occurrences) of sender s_i in the multiset S^*. Similarly, we define on the receiver side the multiset $R^* = (R, K)$, the set R of ρ unique receivers and the vector $K = \langle k_1, k_2, ..., k_\rho \rangle$ containing their multiplicities. We represent the fact that an input i_j corresponds to an output o_l by $i_j o_l$. A sender s_j communicating with a receiver r_l is denoted as $s_j r_l$.

In Fig. 2, Round 1, $S = \{A, B, C\}$, $N = \langle 1, 1, 1 \rangle$, $R = \{D, E, F\}$ and $K = \langle 1, 1, 1 \rangle$; while in Round 2, $S = \{A, C\}$, $N = \langle 2, 1 \rangle$, $R = \{D, E, F\}$ and $K = \langle 1, 1, 1 \rangle$.

3.2 Example and counterexample

In Fig. 2 we can see two different communication rounds of a threshold mix with threshold $t = 3$. Given the properties of a threshold mix, any of the inputs to the mix is equally likely to correspond to any of the outputs. If no further restrictions exist, the underlying bipartite graph $G = (I, O, E)$ is complete and all elements of the adjacency matrix A are equal to 1 (i.e., $per(A) = t!$).

In both cases there exist $t! = 6$ perfect matchings in the underlying graph, and the *system's anonymity level* computed as in [10] is maximal:

$$d(A) = \frac{log(per(A))}{log(t!)} = \frac{log(t!)}{log(t!)} = 1\,.$$

We express a perfect matching as a set of three correspondences $i_j o_l$. In Round 1, the perfect matchings are:
$\{i_1o_1, i_2o_2, i_3o_3\}$, $\{i_1o_1, i_2o_3, i_3o_2\}$, $\{i_1o_2, i_2o_3, i_3o_1\}$,
$\{i_1o_2, i_2o_1, i_3o_3\}$, $\{i_1o_3, i_2o_1, i_3o_2\}$, $\{i_1o_3, i_2o_2, i_3o_1\}$.

We note however, that the goal of the adversary is to infer the relationships between senders and recipients in the system, not to link specific inputs i_j to outputs o_l. If we replace inputs and outputs by senders and receivers, we obtain the following six perfect matchings:
$\{AD, BE, CF\}$, $\{AD, BF, CE\}$, $\{AE, BF, CD\}$,
$\{AE, BD, CF\}$, $\{AF, BD, CE\}$, $\{AF, BE, CD\}$.

Given that all perfect matchings are different and equally likely, the adversary does not obtain any additional information and the example shown in Round 1 achieves maximum anonymity as indicated by $d(A) = 1$.

In Round 2, the correspondences of inputs and outputs is the same as in Round 1. However, user-wise the relationships are different, namely:
$\{AD, AE, CF\}$, $\{AD, AF, CE\}$, $\{AE, AF, CD\}$,
$\{AE, AD, CF\}$, $\{AF, AD, CE\}$, $\{AF, AE, CD\}$.

The two matchings on the left indicate that user A sends two messages, one to user D and one to user E, and user C sends one message to user F. As the adversary is interested in learning which sender communicated with which receiver, there exist only three *distinct* assignments: $\{AD, AE, CF\}$,

$\{AD, AF, CE\}$, $\{AE, AF, CD\}$. Hence, the adversary's uncertainty is reduced to choosing amongst three options instead of six, which does not result in perfect anonymity as indicated by $d(A) = 1$.

4. INTRODUCING MULTIPLICITIES AT SENDER OR RECIPIENT SIDE

We have illustrated in the previous section how the fact that senders form multisets leads the *system's anonymity level* to over-estimate the anonymity provided by the system. This is because several perfect matchings in the underlying graph lead to equivalent sender-receiver relationships. In this section we redefine the *system's anonymity level* such that it takes the multiplicities of *either* senders *or* receivers into account.

Definition Let $\sim$ be the equivalence relation "leads to the same relationship between elements of S and R". Let $\mathcal{M}$ be the set of all possible perfect matchings on the graph G. The equivalence class $[M_p]$ of an element $M_p \in \mathcal{M}$ is the subset of all elements in $\mathcal{M}$ which are equivalent to M_p:

$$[M_p] = \{M_j \in \mathcal{M} | M_j \sim M_p\}\,.$$

Note that one may also see $\sim$ as inducing a partition of $\mathcal{M}$ where the blocks are the equivalence classes.

Given a multiset of senders S^*, where each distinct sender s_j appears with multiplicity n_j, and assuming that the receivers form a set and not a multiset (i.e., $R^* = R$), the size of each equivalence class is $\prod_{j=1}^{\sigma} n_j!$. The number of equivalence classes is:

$$\Xi = \frac{per(A)}{\prod_{j=1}^{\sigma} n_j!}\,.$$

A similar situation occurs if repetitions occur only on the receiver side. Given a multiset of receivers R^*, where each distinct receiver r_l appears with multiplicity k_l, and assuming that the senders form a set and not a multiset (i.e., $S^* = S$), the size of each equivalence class is $\prod_{l=1}^{\rho} k_l!$. The number of equivalence classes is:

$$\Psi = \frac{per(A)}{\prod_{l=1}^{\rho} k_l!}\,.$$

We therefore redefine the *system's anonymity level* as:

$$d^*(A) = \begin{cases} 0 & \text{if } t = 1 \\ \frac{\log(\Xi)}{\log(t!)} & \text{if } R^* = R \text{ and } t > 1 \\ \frac{\log(\Psi)}{\log(t!)} & \text{if } S^* = S \text{ and } t > 1 \\ \frac{\log(per(A))}{\log(t!)} & \text{if } R^* = R \text{ and } S^* = S \text{ and } t > 1\,. \end{cases}$$

For the example shown in Fig. 2 we obtain $d^* = 1$ for Round 1, and a lower $d^* = 0.61$ for Round 2 where one of the senders sends two messages.

5. GENERALIZING THE SYSTEM'S ANONYMITY METRIC

So far we have considered that either the senders or the receivers form a multiset. In this section we look at the case when *both* senders *and* receivers form multisets.

Let Θ denote the number of equivalence classes and let C_p denote the number of equivalent perfect matchings in class $[M_p]$ (i.e., its cardinality). In the new scenario the equivalence classes may have different sizes C_p, as illustrated in the following example.

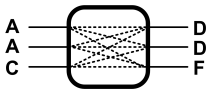

Figure 3: A round with repetitions on both sides

Consider the case shown in Fig. 3, where we can see repetitions on both sides. The number of possible perfect matchings is $per(A) = 3! = 6$. These matchings belong to one of $\Theta = 2$ equivalence classes. Class $[M_1]$ with cardinality $C_1 = 2$ is represented by $M_1 = \{AD, AD, CF\}$ and class $[M_2]$ with cardinality $C_2 = 4$ is represented by $M_2 = \{AD, AF, CD\}$.

The *system's anonymity level* aims at determining the amount of additional information needed to reveal all sender-receiver relationships; i.e., to find the equivalence class containing the correct perfect matching M_C between senders and receivers. In the previous section, we computed this amount of information as the logarithm of the number of equivalence classes. Note that this holds as long as M_C belongs to any class with equal probability (or in other words, as long as all equivalence classes have equal cardinality).

In the example of Fig. 3, however, the correct perfect matching M_C belongs to the equivalence class $[M_1]$ with probability $\Pr(M_C \in [M_1]) = \frac{C_1}{per(A)} = \frac{2}{6}$ and to the class $[M_2]$ with probability $\Pr(M_C \in [M_2]) = \frac{C_2}{per(A)} = \frac{4}{6}$. The amount of additional information required to identify the equivalence class that contains M_C is thus given by the Shannon entropy of the random variable with probability distribution $Pr(M_C \in [M_p]) = \frac{C_p}{per(A)}$. We generalize the previous definition of the *system's anonymity level* as:

$$d^*(A) = \begin{cases} 0 & \text{if } t = 1 \\ \frac{-\sum_{p=1}^{\Theta} \Pr(M_C \in [M_p]) \cdot \log(\Pr(M_C \in [M_p]))}{\log(t!)} & \text{if } t > 1\,. \end{cases}$$

In order to calculate $d^*(A)$ we need to obtain Θ and C_p. A naïve way of computing these values is to enumerate all possible perfect matchings, and to classify them into equivalence classes. We note that this process requires $\mathcal{O}(t!)$ operations and quickly becomes infeasible.

5.1 Computing the metric

We present a divide and conquer algorithm that obtains Θ and C_p more efficiently than exhaustive search. The idea is to divide the complex problem into smaller ones, and to use their solutions to solve the original problem. In the divide step (see Algorithm 1), the algorithm recursively constructs a tree $\mathcal{T}$ removing elements from S^* and R^* in each recursion. A node in the tree represents a certain intermediate situation and an edge between two nodes reflects a possible scenario. Each edge is assigned with a weight that describes the likelihood of that scenario. At a given node (recursion) the algorithm stops when the problem can either: (1) be solved with the approach explained in the previous section (i.e., the updated S^* and/or R^* become a set); or (2) we arrive to a situation where the updated S and/or R have only one element. When the divide step terminates, it has constructed the tree $\mathcal{T}$ and returns Θ.

Let $\mathcal{P}$ be a path in $\mathcal{T}$ from the root to a leaf. We note that there exist Θ paths $\mathcal{P}_p$ ($p = 1, \ldots, \Theta$) and that each

path $\mathcal{P}_p$ represents an equivalence class $[M_p]$.

The conquer step operates on the previously generated tree $\mathcal{T}$ to compute the cardinalities C_p of the equivalence classes. The cardinality C_p of an equivalence class is given by the product of the weights of the edges along the path $\mathcal{P}_p$ representing $[M_p]$: $C_p = \prod_{w \in \mathcal{P}_p} w$.

Algorithm 1 Divide step

Input: S, R, N, K, A (where (S, N) are sorted in decreasing order of n_i)
Output: Θ (the number of equivalence classes)

1: function exploreNode(S,R,N,K,A)
2: $\Theta = 0$
3: generate new node
4: **if** (|S| == 1) respectively (|R| == 1) **then**
5: $w = n_1!$ resp. $w = k_1!$
6: generate a new edge, associate the weight w to it, and generate a leaf node at the other end of that edge
7: return(1)
8: **else if** (all $n_j \in N == 1$) resp. (all $k_l \in K == 1$) **then**
9: $w = \prod k_l!$ resp. $w = \prod n_j!$
10: generate $per(A)/w$ new edges, associate the weight w to each of them, and generate a leaf node at the other end of each edge
11: return($per(A)/w$)
12: **else**
13: activeSender $\leftarrow s_1$
14: activeSenderMult $\leftarrow n_1$
15: $W := \{W_i \mid |W_i| = \text{activeSenderMult} \wedge W_i \subseteq R^* \wedge r_x \in W_i \Leftrightarrow a_{1,x} = 1\}$ //$a_{1,x}$ is an element of A
16: **for all** $W_i \in W$ **do**
17: $S' \leftarrow S \setminus$ activeSender
18: $N' \leftarrow$ update N //remove n_1 from the vector
19: $R' \leftarrow R$
20: $K' \leftarrow$ update K //decrease k_l by the multiplicity of r_l in W_i; if a multiplicity k_l becomes 0, k_l is removed from K' and r_l is removed from R'
21: $A' \leftarrow$ update A //A' is the adjacency matrix of the graph formed by the remaining nodes $s_j \in S'$, $r_l \in R'$ with their respective multiplicities N' and K', and edges between them
22: $w = \text{activeSenderMult}! \cdot \prod_{r_l \in W_i} \binom{k_l}{k_l - k'_l}$
23: generate a new edge, associate the weight w to it
24: $\Theta = \Theta + \text{exploreNode}(S', R', N', K', A')$ //this recursive call generates a node that is connected to the edge we just generated
25: **end for**
26: **end if**

5.2 Bounds

Although this algorithm has much lower complexity than $\mathcal{O}(t!)$, we note that for large values of t it may become rather expensive. Nevertheless, we can easily compute bounds for the *system's anonymity level* if the graph associated to the anonymous communication system is complete (e.g., the system is a threshold mix).

The upper bound is given by

$$d^*(A) \leq \min\left(\frac{log(\Psi)}{log(t!)}, \frac{log(\Xi)}{log(t!)}\right).$$

The intuition is the following: repetitions on both the sender and receiver sides can only reduce $d^*(A)$ with respect to the case where repetitions occur only on one side. Therefore, the *system's anonymity level* in this scenario cannot be larger than $d^*(A)$ considering that either $S = S^*$ or $R = R^*$.

To compute the lower bound, we take the minimum of σ and ρ, the cardinalities of the sets of senders S and receivers R, respectively, and we obtain

$$d^*(A) \geq \min\left(\frac{log(\sigma!)}{log(t!)}, \frac{log(\rho!)}{log(t!)}\right).$$

Let us assume $\sigma \leq \rho$. In this case there are at least $\sigma!$ distinct perfect matchings between the σ unique senders and any σ unique receivers. If $\sigma \geq \rho$ the reasoning is analogous.

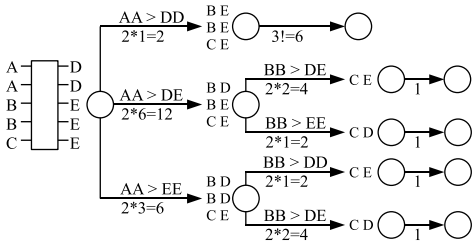

Figure 4: Example for the proposed algorithm

5.3 Example

We use an example to illustrate how our algorithm works. Consider a scenario where $S = \{A, B, C\}$, $N = \langle 2, 2, 1\rangle$, $R = \{D, E\}$, $K = \langle 2, 3\rangle$, and all elements of the adjacency matrix are equal to 1. Figure 4 shows the tree as it will be generated by the divide step. The first call of the algorithm generates the root node (the leftmost) which reflects the initial situation. Since none of the stop conditions is fulfilled, the algorithm selects the first sender in the set and analyzes her possible choices $W = \{\{D, D\}, \{D, E\}, \{E, E\}\}$. Let us look at the scenario "A sends two messages to $\{D, E\}$". To prepare the input to the next recursive call, A is removed from the set of senders (line 17) and the according multiplicity is removed from the vector N (line 18). We update the receiver side similarly: for each receiver, we decrease her multiplicity in the vector K according to how often she appears in the current scenario W_i. As k_2 becomes zero, the element is removed from the vector K' and the receiver D is removed from the set of receivers R' (lines 19 and 20). Next, we update the adjacency matrix such that it reflects the reduced multisets of senders and receivers. The last step before the recursive call is to insert a new edge, starting from the current node, into the tree and to associate the correct weight to it. For the scenario "A sends two messages to $\{D, E\}$" the weight is $w = 2! \cdot \binom{2}{1} \cdot \binom{3}{1} = 2 \cdot 2 \cdot 3 = 12$. The recursive call of the algorithm inserts a new node on the other end of that edge. It reflects the resulting situation $S = \{B, C\}, N = \langle 2, 1\rangle, R = \{D, E\}, K = \langle 1, 2\rangle$. This situation does still not fulfill any of the stop conditions. The first sender in the set is B and her possible choices are $W = \{\{D, E\}, \{E, E\}\}$. The possible scenarios are analyzed in the same way as explained above.

Once the last leaf has been generated, the initially called instance of the algorithm terminates and returns the computed number of equivalence classes $\Theta = 5$.

A path $\mathcal{P}$ from the root to a leaf represents an equivalence class. The conquer step determines the cardinality of an equivalence class by multiplying all edge weights along its corresponding path. The five classes contain 12, 48, 24, 12 and 24 perfect matchings respectively. Our proposed metric evaluates to $d^*(A) = 0.31$. The original $d(A) = 1$ indicates perfect anonymity which is much higher than our upper bound $d^*(A) \leq 0.48$. Our lower bound is $d^*(A) \geq 0.14$.

6. CONCLUSIONS

In this paper we revisit the combinatorial approach for quantifying the *system's anonymity level* proposed by Edman et al. in [10]. We argue that anonymity metrics should focus on the relations between senders and receivers rather than on the links between inputs and outputs. We show how the *system's anonymity level* as defined in [10] focuses only on individual messages and thus cannot reflect the reduction of anonymity in scenarios where senders and/or receivers form multisets. We generalize the metric in scenarios where user relations can be modeled as yes/no relations to capture the additional information provided by multiple messages from/to the same sender/recipient. We propose an algorithm to compute the redefined *system's anonymity level*. The algorithm may become rather expensive for large values of t and specifying a more efficient algorithm remains as an open problem. Nevertheless, we provide a simple and efficient way to obtain upper and lower bounds if the graph associated to the system is complete.

Acknowledgements: C. Troncoso is funded by a research grant of the Fundacion Barrie de la Maza (Spain). This work was supported in part by the Belgian State's (Belgian Science Policy) IAP Programme P6/26 BCRYPT, by the IWT SBO ADAPID project, by FWO projects G.0475.05 and G.0300.07, by the Flemish Government's Concerted Research Action (GOA) Ambiorics 2005/11, by the European Comission's IST Programme under the contract IST-2002-507932 ECRYPT NoE, and by the K.U. Leuven-BOF. The information in this document reflects only the authors' views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

7. REFERENCES

[1] Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: The disclosure attack. In *IEEE Security & Privacy*, 1(6):27–34, 2003.

[2] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. In *Communications of the ACM*, 24(2):84–88, 1981.

[3] Sebastian Clauß and Stefan Schiffner. Structuring anonymity metrics. In *ACM Workshop on Digital Identity Management*, pages 55–62, 2006.

[4] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a type III anonymous remailer protocol. In *IEEE Symposium on Security and Privacy*, pages 2–15, 2003.

[5] Yuxin Deng, Jun Pang, and Peng Wu. Measuring anonymity with relative entropy. In Theodosis Dimitrakos, Fabio Martinelli, Peter Y. A. Ryan, and Steve A. Schneider, editors, *Formal Aspects in Security and Trust*, pages 65–79. Springer, LNCS 4691, 2006.

[6] Claudia Diaz and Bart Preneel. Reasoning about the anonymity provided by pool mixes that generate dummy traffic. In *Information Hiding*, pages 309–325. Springer, LNCS 3200, 2004.

[7] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Designing Privacy Enhancing Technologies, PET'02*, pages 54–68. Springer-Verlag, LNCS 2482, 2002.

[8] Claudia Diaz, Carmela Troncoso, and Andrei Serjantov. On the impact of social network profiling on anonymity. In Nikita Borisov and Ian Goldberg, editors, *Privacy Enhancing Technologies*, page 19. Springer, LNCS 5134, 2008.

[9] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *13th USENIX Security Symposium*, pages 303–320. USENIX, 2004.

[10] Matthew Edman, Fikret Sivrikaya, and Bülent Yener. A combinatorial approach to measuring anonymity. In *Intelligence and Security Informatics, 2007 IEEE*, pages 356–363, 2007.

[11] Matthias Franz, Bernd Meyer, and Andreas Pashalidis. Attacking unlinkability: The importance of context. In Nikita Borisov and Philippe Golle, editors, *Privacy Enhancing Technologies*, pages 1–16. Springer, LNCS 4776, 2007.

[12] David Goldschlag, Michael Reed, and Paul Syverson. Hiding routing information. In *Information Hiding*, pages 137–150, 1996.

[13] Andreas Pfitzmann and Marit Hansen. Anonymity, unobservability and pseudonymity – a proposal for terminology. In *Designing Privacy Enhancing Technologies, PET'00*, pages 1–9. Springer-Verlag, LNCS 2009, 2001.

[14] Michael Reed, Paul Syverson, and David Goldschlag. Anonymous Connections and Onion Routing. In *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, 1998.

[15] Alfred Rényi. On measures of entropy and information. In *4th Berkeley Symposium Mathematical Statistics and Probability*, 1:547–561, 1961.

[16] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *Designing Privacy Enhancing Technologies, PET'02*, pages 41–53. Springer-Verlag, LNCS 2482, 2002.

[17] Claude Shannon. A mathematical theory of communication. In *The Bell System Technical Journal*, 27:379–423:623–656, 1948.

[18] Gergely Tóth, Zoltán Hornák, and Ferenc Vajda. Measuring anonymity revisited. In *9th Nordic Workshop on Secure IT Systems*, pages 85–90, 2004.

[19] Peter Palfrader Ulf Moller, Lance Cottrel and Len Sassaman. Mixmaster protocol - version 2. `http://www.abditum.com/mixmaster-spec.txt`, 2003.

[20] Ye Zhu and Riccardo Bettati. Anonymity vs. information leakage in anonymity systems. In *25th IEEE International Conference on Distributed Computing Systems (ICDCS'05)*, pages 514–524, 2005.